International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Scholarly Publisher
RS Global Sp. z O.O.
ISNI: 0000 0004 8495 2390

Dolna 17, Warsaw,
Poland 00-773
+48 226 0 227 03
editorial_office@rsglobal.pl

ARTICLE TITLE	LARGE LANGUAGE MODEL BASED CHATBOTS – A CHANCE FOR CLOSING THE MENTAL HEALTH TREATMENT GAP OR A THREAT TO THE PUBLIC HEALTH? A NARRATIVE REVIEW
DOI	https://doi.org/10.31435/ijitss.3(47).2025.3809
RECEIVED	05 August 2025
ACCEPTED	15 September 2025
PUBLISHED	19 September 2025
LICENSE	The article is licensed under a **Creative Commons Attribution 4.0 International License.**

LARGE LANGUAGE MODEL BASED CHATBOTS – A CHANCE FOR CLOSING THE MENTAL HEALTH TREATMENT GAP OR A THREAT TO THE PUBLIC HEALTH? A NARRATIVE REVIEW

Tomasz Ufniarski *(Corresponding Author, Email: tomasz.eku@gmail.com)*
University Clinical Centre in Gdańsk, Dębinki 7, 80-952 Gdańsk, Poland
ORCID ID: 0009-0008-6555-3403

Maria Ufniarska
Saint Adalbert Hospital in Gdańsk, aleja Jana Pawła II 50, 80-462 Gdańsk, Poland
ORCID ID: 0009-0008-5927-4811

Aleksandra Piech
Clinical Provincial Hospital No. 2 in Rzeszow, Lwowska 60, 35-301 Rzeszów, Poland
ORCID ID: 0009-0001-4485-2200

Karolina Pasierb
The University Hospital in Krakow, Jakubowskiego 2, 30-668 Kraków, Poland
ORCID ID: 0009-0006-5806-3508

Karol Poplicha
National Medical Institute of the Ministry of the Interior and Administration, Wołoska 137, 02-507 Warszawa, Poland
ORCID ID: 0009-0005-3835-9777

Martyna Grodzińska
Lower Silesian Oncology, Pulmonology and Hematology Center, pl. Ludwika Hirszfelda 12, 53-413 Wroclaw, Poland
ORCID ID: 0009-0004-1001-6484

Bartłomiej Siuzdak
Clinical Provincial Hospital No. 2 in Rzeszów, Lwowska 60, 35-301 Rzeszów, Poland
ORCID ID: 0009-0003-8691-6617

Justyna Moszkowicz
Clinical Provincial Hospital No. 2 in Rzeszow, Lwowska 60, 35-301 Rzeszów, Poland
ORCID ID: 0009-0009-2582-6187

Piotr Sobkiewicz
Lower Silesian Oncology, Pulmonology and Hematology Center, pl. Ludwika Hirszfelda 12, 53-413 Wroclaw, Poland
ORCID ID: 0009-0007-6610-440X

Patrycja Kardasz
Saint Adalbert Hospital in Gdańsk, aleja Jana Pawła II 50, 80-462 Gdańsk, Poland
ORCID ID: 0009-0006-8137-9789

Marta Jutrzenka
Szpital Praski p.w. Przemienienia Pańskiego Sp. z o.o., Aleja Solidarności 67, 03-401 Warszawa, Poland
ORCID ID: 0000-0001-7266-1586

Patrycja Ucieklak
District Hospital in Zawiercie, Ul. Miodowa 14, 42-400 Zawiercie, Poland
ORCID ID: 0009-0002-3681-1051

ABSTRACT

This narrative review examines whether Large Language Model (LLM)–based chatbots can help close the global mental health treatment gap while weighing their public-health risks. We synthesize peer-reviewed studies and relevant case reports to: (1) map the dimensions of the mental health treatment gap, (2) describe how recent LLM advances have changed chatbot capabilities, (3) explore how chatbots can address the dimensions of the gap, (4) evaluate evidence for clinical effectiveness, and (5) outline major safety, ethical, and policy concerns. Findings indicate that chatbots offer scalable, always-available, and low-cost support that can reduce barriers related to stigma, geographic and temporal access, affordability, and mental-health awareness. We found that the evidence supports chatbot interventions' efficiency in small-to-moderate short-term reductions in depression and anxiety symptoms, while the long-term effects and use in other disorders remain largely unexplored. However, LLM chatbots also present clear risks: hallucinations and clinically inappropriate responses, amplification of stigma or bias, user dependence, and significant data-security vulnerabilities. Importantly, most widely used generalist LLMs lack rigorous clinical validation. We conclude that LLM chatbots are a persistent feature of the mental-health ecosystem whose benefits can be realized only with robust safety guardrails, transparent evaluation, integration into stepped-care pathways, and proactive regulation.

KEYWORDS

Large Language Models, Chatbots, Mental Health Treatment Gap, Digital Mental Health, Artificial Intelligence

CITATION

Ufniarski Tomasz, Ufniarska Maria, Piech Aleksandra, Pasierb Karolina, Poplicha Karol, Grodzińska Martyna, Siuzdak Bartłomiej, Moszkowicz Justyna, Sobkiewicz Piotr, Kardasz Patrycja, Jutrzenka Marta, Ucieklak Patrycja. (2025) Large Language Model Based Chatbots – A Chance for Closing the Mental Health Treatment Gap or a Threat to the Public Health? A Narrative Review. *International Journal of Innovative Technologies in Social Science*. 3(47). doi: 10.31435/ijitss.3(47).2025.3809

Introduction

Addressing mental health issues remains one of the most urgent public health priorities. By the age of 75, approximately half of all people will experience at least one mental disorder (McGrath et al., 2023). These conditions cause profound distress not only for affected individuals, but also for their families and communities. According to the 2021 Global Burden of Disease Study, depressive disorders rank as the second-leading cause of disability worldwide, and anxiety disorders as the sixth. Other major contributors include schizophrenia, Alzheimer's disease and other dementias, autism spectrum disorders, and substance use disorders. Between 2010 and 2021, age-standardized disability-adjusted life-years (DALYs) increased by 16.4% for depressive disorders and by 16.7% for anxiety disorders (The Lancet Psychiatry, 2024). The worldwide divergence between the number of people needing treatment for mental disorders and the number that are receiving treatment, known as the mental health treatment gap, remains an alarming public health concern (Evans-Lacko et al., 2018).

The COVID-19 pandemic further deepened this crisis, both through widespread societal disruption and via direct neurological and psychiatric effects in some individuals following SARS-CoV-2 infection (Penninx et al., 2022). Daily life was profoundly altered: prolonged isolation, loss of work and social connection, and, for many, the grief of losing loved ones. In this environment, digital spaces became a lifeline — offering both meaningful connection and, at times, unhealthy escapes. Within this shift, AI-powered chatbots moved from niche experiments to widely used tools. In 2021, a U.S. survey reported that 22% of adults had used a mental health chatbot, with nearly 60% of these users beginning during the pandemic. Notably, 44% relied solely on chatbots without consulting a human clinician (Zagorski, 2022). More recent findings indicate widespread use of LLMs for mental health; 48.7% of respondents that used these systems in the past year and self-reported ongoing mental health condition claimed they had sought psychological support from them. The most common reasons were anxiety (73.3%), personal advice (63.0%), and depression (59.7%). Overall, 63.4% said their mental health improved through such interactions, rating practical advice (86.8%) and overall helpfulness (82.3%) highly. Comparisons with human

therapy were largely neutral to positive, with 37.8% considering LLMs more beneficial than traditional approaches. Only 9% reported harmful responses (Rousmaniere et al., 2025).

The rapid uptake of such technologies reflects both unmet needs and structural limitations within existing mental health systems. During the pandemic, restrictions on in-person care and heightened psychological distress created a demand for scalable, always-available, and low-cost forms of support. The 24/7 accessibility of chatbots, their lack of geographical constraints, and their perceived privacy made them a desirable first-line option, particularly in settings where professional care was scarce or delayed (Chin et al., 2023).

While many users view mental health chatbots as a helpful, usable, and acceptable resource, this positive reception conceal a more complex reality. Concerns persist about the tools' limitations, including their lack of genuine empathy, a tendency toward repetitive responses and reinforcement of unhelpful beliefs and the possibility of hallucinations (A. A. Abd-Alrazaq et al., 2021; Mayor, 2025).

In this paper we want to reflect on the following questions:

What is the mental health treatment gap and what are the dimensions of it?

How have advancements in chatbot technology, such as Large Language Models, redefined the landscape and the expectations for their application in mental health?

How and to what extent chatbots can address various dimensions of the mental health treatment gap?

What is the evidence of chatbot-based interventions in improving mental health outcomes?

What are the main public health considerations when using chatbots for mental health?

What critical research questions and development priorities exist for advancing the use of chatbots in mental health?

Methodology

Relevant literature was searched for on 27.06.2025 using PubMed, GoogleScholar, ScienceDirect and Cochrane Library. Key words included "mental health", "chatbots", "large language models" and "artificial intelligence". Randomized controlled trials, meta-analyses and observational studies were prioritized.

Results

1. The Mental Health Treatment Gap

For the past two decades, the discourse surrounding mental healthcare challenges has largely been framed by the concept of the 'mental health treatment gap.' It is based on the presumption that the majority of people suffering from mental disorders does not receive appropriate care, even though efficient treatment options exist. The famous 2004 study reports that the median treatment gap for schizophrenia, including other non-affective psychosis, was 32.2%. For other disorders the gap was: depression, 56.3%; dysthymia, 56.0%; bipolar disorder, 50.2%; panic disorder, 55.9%; GAD, 57.5%; and OCD, 57.3%. Alcohol abuse and dependence had the widest treatment gap at 78.1% (Kohn et al., 2004). The results World Health Organisation World Mental Health Survey, that was conducted in 24 countries with 63 678 participants, show that only 13.7 % of 12-month DSM-IV/CIDI cases in lower-middle-income countries, 22.0% in upper-middle-income countries, and 36.8% in high-income countries received treatment (Evans-Lacko et al., 2018). Stigma, limited-service availability (due to lack of financial resources as well as shortages of trained personnel) (Luitel et al., 2017; Mongelli et al., 2020; Wainberg et al., 2017) have been considered the main reasons for this phenomenon. However, it has been discussed that low perceived need and awareness (Roberts et al., 2022) are mentioned as other potential explanations. International efforts, such as the World Health Organization's Mental Health Gap Action Program (mhGAP), have aimed to integrate mental, neurological, and substance use care into primary care and community levels (World Health Organization, 2023). While systemic reviews indicate a significant impact of mhGAP on training, patient care, and research, universal availability of much-needed care remains a distant goal (Keynejad et al., 2021).

2. Chatbots and the Large Language Model Based Revolution

Chatbots are intelligent computer systems that are designed to mimic human conversation in its natural form. They can be interacted via a smartphone app or web browser simultaneously by many users (Caldarini et al., 2022). The usual uses of a chatbot include customer service, answering questions on a topic or entertainment. Chatbots have already been used in healthcare for decades (A. Abd-Alrazaq et al., 2019).

Early conversational agents, such as ELIZA (Weizenbaum, 1966), offered one of the first demonstrations of natural language interaction between humans and computers. ELIZA's most famous script

simulated a Rogerian psychotherapist, reflecting user statements back as open-ended prompts (e.g., "You say you feel anxious—can you elaborate?"). While this gave the impression of understanding, the system relied on pattern matching and keyword substitution, without any true comprehension or contextual awareness. As a result, interactions were shallow, easily breaking down when the conversation moved beyond anticipated patterns. What is interesting, despite these limitations, people were willing to anthropomorphize the chatbot (so called ELIZA effect) and willingly disclose personal information to it (Weizenbaum, 1966).

Over the following decades, finite-state and frame-based dialogue managers expanded chatbots' capabilities through larger, hand-crafted response libraries. That allowed for attempting using one to deliver a cognitive-behavioural therapy. An early trial showed that there was no significant difference in the reduction of depressive symptoms measured by the Beck's inventory between patients who received the therapy from a human therapist and from a chatbot. However, the participants generally perceived the human-delivered therapy as more informative, and the study was limited by small number of participants (12 in each group) (Selmi et al., 1990). The ideal interactive program for patients, as proposed by W. Slack called for medical soundness, user-friendliness, genuine interactivity, patient autonomy, confidentiality, speed, reliability, and rigorous evaluation before release (Slack, 2000). This vision highlighted the considerable gap between existing capabilities and what was required for meaningful mental health engagement. Systematic reviews confirm that early health conversational agents remained predominantly task-oriented and rule-based, constraining open-ended interaction, contextual understanding, and emotional nuance (A. Abd-Alrazaq et al., 2019; Laranjo et al., 2018).

The advent of large language models (LLMs) has, in many respects, narrowed this gap. The LLMs are types of generative artificial intelligence that are trained on large corpora of text. They can process and generate text with coherent communication and generalize to multiple tasks (Naveed et al., 2023). Early sequence-to-sequence neural models (Sutskever et al., 2014) enabled data-driven generation, but recurrent neural networks (RNNs) still struggled with context retention and diversity (J. Li et al., 2016). Transformer architectures (Vaswani et al., 2017) revolutionized this field, allowing LLMs such as GPT-3 (T. B. Brown et al., 2020) to sustain multi-turn context, produce semantically rich and stylistically adaptive responses, and respond across diverse domains with remarkable fluency.

In mental health contexts, these advances have reshaped expectations. Modern LLMs satisfy many of Slack's original criteria: they are intuitive to use, adapt to a user's style and needs, and provide choice and autonomy in interaction. Yet their opaque decision-making (Ali et al., 2023) and rapid, unregulated deployment without rigorous trials introduce new concerns. At the same time, they increasingly serve not only informational and therapeutic purposes but also as emotional companions in attempt to counter loneliness and social isolation (T. Xie & Pentina, 2022). A substantial share of users engages with chatbots primarily for companionship, with some reporting reduced loneliness after sustained interaction (Vaidyam et al., 2021). Systems such as ChatGPT, Gemini, Claude or Grok are able to simulate empathy, adjust tone dynamically, and sustain personalized exchanges over time – capabilities unimaginable in earlier generations.

The share of U.S. adults who have used ChatGPT has nearly doubled since 2023, reaching 34%, while 80% are now aware of its existence. Among adults under 30, a majority of 58% report having used it in the past year (Olivia Sidoti & Colleen McClain, 2025). A 2023 survey found that AI chatbot users, utilize them primarily for quick support (60%) and as a personal therapist (47%) (Cross et al., 2024).

The chatbot experience transitioned from rigid, scripted interactions with niche scientific projects to adaptive, emotionally attuned dialogue. LLM chatbots have disrupted the way how people gain knowledge, look for help or even seek a remedy for loneliness. They blur the line between a tool and a companion, opening possibilities for scalable, personalised support, while introducing challenges related to inadequate or false responses, bias and dependency. These shifts are not merely signs of technological evolution in chatbots, but a redefinition of their place in the mental health landscape.

3. Addressing Dimensions of Mental Health Treatment Gap with the use of Chatbot-based Interventions.

3.1 Stigma

One important factor comprising for the mental health treatment gap is stigma associated with mental disorders (Roberts et al., 2022). Stigma, in the context of mental health, refers to the negative attitudes, beliefs, and behaviours directed towards individuals with mental illnesses. It encompasses both public stigma (the prejudice and discrimination from the general population) and self-stigma (the internalized prejudice that individuals with mental illness turn against themselves), often manifesting through stereotypes, prejudice, and discrimination (Corrigan & Rao, 2012). However, the fear of being stigmatized rarely leads individuals to change behaviour but rather prompts them to conceal it (e.g. drinking in secrecy) (Goffman, 1974). The mere fear of being exposed as mentally ill (be it either in front of society or oneself) does not prevent but rather makes the distress more severe. Consequently, stigma presents a formidable barrier to mental health care, acting as a complex, multi-layered phenomenon that profoundly influences help-seeking behaviours, treatment adherence, and overall recovery outcomes, thereby perpetuating a cycle of suffering and social exclusion (Clement et al., 2015).

There are two main qualities of chatbots that bring a promise of mitigating the effects of stigma: their inherent privacy and perceived non-judgemental nature. This allows users to interact without the fear of social repercussions or judgment often associated with disclosing mental health concerns to another human (Haque & Rubya, 2023; Sweeney et al., 2021). Such an environment can significantly lower the barrier to seeking initial help, encouraging individuals who might otherwise avoid traditional care to engage. The interactions with chatbots are typically perceived as private, fostering a sense of security that can make users more comfortable sharing sensitive information (Pan et al., 2024). This perceived confidentiality creates a low-stakes environment for individuals to explore their feelings and symptoms, serving as a crucial first step before potentially seeking human intervention.

The anticipated privacy might prove deceptive. While the users do not need to face the social stigma connected with sharing their symptoms, the logs of their conversation are being kept by the chatbots' providers and are susceptible for data breaches. When confronted with the logs of their conversations, users tend to admit that they disclosed more information about themselves than they intended to (Gumusel et al., 2024a). The sensitive information may be used for social-based personalization, behavioural profiling, and location-based personalization. If stolen, the risks of marketing data misuse, unauthorized access to personalized software or physical accounts and devices, and discrimination arise (Toch et al., 2012). The risk is underscored by notorious incidents like in the case of Cambridge Analytica where personal data of millions Facebook users was captured and used for political advertising (Boldyreva et al., 2018). Should malicious actors gain access to this information, trust and anonymity that had made these tools initially attractive would be undermined.

Beyond individual interaction, research indicates that chatbot-based social contact interventions can actively contribute to broader stigma reduction (Song et al., 2025). In a recent study, participants were prompted with 7 vignettes over the course of 2 weeks. The vignettes described the same person experiencing symptoms of depression. Then they were asked questions designed to disclose their potentially stigmatizing attitudes toward mental illness. This was followed by a conversation with chatbots giving either stigmatizing or not stigmatizing interpretations and showing varying levels of self-disclosure. The researchers found that the chatbot featuring non-stigmatizing interpretations and non-stigmatizing self-disclosure was most effective in reducing the participants' stigmatizing attitudes (Cui et al., 2024). These findings underscore the critical responsibility of chatbot-based service providers to meticulously curate their products for positive social impact.

Conversely, there's a risk that current chatbots, if not aligned properly, might intensify stigma. They could do this by amplifying existing stereotypes and prejudiced attitudes. This vulnerability stems from the fact that the transformer technology – the foundation of modern Large Language Model-based chatbots is trained on immense volumes of text. The data include social media content that inherently contains a wide spectrum of, at times, unfavourable opinions and behaviours (Bender & Friedman, 2018).

3.2 Limited Service Availability

3.2.1 Geographical and Timely Limitations

The limited availability of mental health care globally stems from a confluence of factors, primarily a high demand (McGrath et al., 2023; Penninx et al., 2022; The Lancet Psychiatry, 2024; Wainberg et al., 2017) confronted by a severe shortage of trained personnel (Endale et al., 2020) and insufficient funding (Mahomed, 2020), particularly pronounced in low- and middle-income countries (Phelan et al., 2022). Furthermore, geographic isolation significantly compounds reduced access to healthcare in many regions (Morales et al., 2020).

Conversational agents offer a compelling solution to bridge these access gaps. They provide 24/7 availability, ensuring immediate support regardless of language, time zones or traditional clinic hours, which is crucial for individuals experiencing distress at any time. Their online nature allows access from any location with an internet connection, effectively overcoming geographical barriers, especially in rural or underserved areas where mental health infrastructure is sparse. Innovations like Starlink, utilize satellite technology for internet connectivity, thus extending the potential reach of chatbot-based solutions to even the most remote areas globally (Shaengchart & Kraiwanit, 2023).

By offering initial support, screening, psychoeducation, and even guided self-help modules, chatbots can potentially alleviate pressure on overstretched mental health services and contribute to reducing long waiting times for human therapists, acting as a triage or first-line support system (Kosyluk et al., 2024; Rollwage et al., 2023a; Van Der Schyff et al., 2023). For instance, the Limbic Access AI chatbot, which has gained medical device certification, has demonstrated a 2-day reduction in time to clinical assessment and 5-day reduction in time to treatment in United Kingdom's National Health Services. Its ability to gather information through a supportive chatbot conversation and deliver effective triage significantly reduces the workload for clinicians, allowing them to focus more on direct patient care and leading to shorter wait times and higher attendance recovery rates for patients (Habicht et al., 2024; Maleki Varnosfaderani & Forouzanfar, 2024; Rollwage et al., 2023b). This aligns with broader observations that AI-powered tools can support healthcare professionals in providing personalized care more quickly, potentially preventing the development of severe mental illness outcomes and thus alleviating pressure on services (Olawade et al., 2024).

However, the promise of universal access to mental healthcare through chatbots is tempered by several limitations. Chatbots are not suitable for severe mental health crises requiring immediate human intervention, such as active suicidality or acute psychosis. In these urgent cases, the online-first approach could provide a false sense of security and as a result postponing an emergency call or hospital admission. The delay of treatment could turn out detrimental to treatment outcomes (Drake et al., 2020) or even fatal (Deisenhammer et al., 2009).

This limitation is further compounded by how mental illness itself can compromise an individual's motivational and self-advocacy capacities, particularly for those who are socially marginalized (J. E. H. Brown & Halpern, 2021). Conditions like severe depression can lead to reduced hope, lower self-esteem, and decreased motivation to pursue therapeutic goals(Curley et al., 2019), thus making the proactive engagement required for the use of digital tools challenging without human encouragement (Bailey et al., 2024).

3.2.2 Affordability of Care

The economic burden of mental illness is substantial, marked by insufficient public funding (Mahomed, 2020) and a relatively high out-of-pocket economic burden for individuals seeking care(Gao & Olfson, 2025). This financial barrier disproportionately affects individuals in low- and middle-income countries, where mental health services are often under-resourced (Chisholm et al., 2019).

Conversational agents offer a compelling solution to enhance the affordability of mental healthcare. The development and deployment of chatbots are generally less expensive per user than training and employing human therapists, allowing for the provision of low-cost services to a broad audience. Many general-use chatbot services are indeed available for free within certain limits, making them appealing to populations who cannot afford traditional therapy (Khawaja & Bélisle-Pipon, 2023). Their scalability is a key economic advantage: a single chatbot can simultaneously serve millions of users, drastically reducing the per-user cost compared to one-on-one human therapy. This enables the creation of tiered support models, where chatbots provide a cost-effective first tier of support, reserving more expensive human resources for complex cases that necessitate specialized intervention (Kuhail et al., 2025). Adaptation of the chatbot solutions could lead to significant long-term economic benefits by reducing the economic burden of untreated mental disorders, including lost productivity and decreased healthcare utilization (Christensen et al., 2020).

However, the perceived affordability of chatbots is subject to several nuances and potential limitations. While the direct cost to the end-user might be low, the initial development, fine-tuning and maintenance of sophisticated, effective chatbots, especially those leveraging large language models, can still be substantial (Xia et al., 2024). It is important to remember that the Large Language Model based chatbots come with a significant ecological burden, mainly because of their immense electricity and water consumption (A.Shaji George et al., 2023). Such investment is further complicated by the possibility that the advantage of affordability can be diminished by substantial costs associated with medical device certification and regulatory compliance for mental health chatbots (Baines et al., 2023). The total cost of bringing novel complex medical device to market might be ten times bigger than the cost of its development (Sertkaya et al., 2022). Furthermore, as of 2025, according to the best knowledge of the authors, the chatbot based solutions are not covered by traditional health insurance, placing the financial burden back on the patient.

3.2.3 Low Perceived Need and Awareness

The availability of mental health treatment is a problem on the supply side of the mental health treatment gap, while low perceived need and awareness are some of the reasons behind low demand. It is the former that by far outnumbers other reasons for not seeking treatment for mental health problems The World Mental Health Survey finds (Andrade-Arenas & Yactayo-Arias, 2024). One explanation for these results might be that the lack of perceived need for treatment is caused by the insufficient mental health education. Studies show that mental health literacy positively correlates with help seeking behaviour (Akakpo & Neuerer, 2024; Baklola et al., 2024; Yang et al., 2024). Individuals might be not aware that the distress they are experiencing is a symptom of a manageable disorder.

Chatbots can be very impactful for increasing the mental health literacy, as they allow for highly approachable, personalised, conversational psychoeducation. The are capable of adapting explanations and content delivery based on user interactions and expressed needs. Unlike generic brochures or websites, chatbots can break down complex mental health concepts into digestible fragments and demonstrate them on personally relatable scenarios. The chatbot is always available, infinitely patient, non-condescending and able to match the communication style of the user (Zhu et al., 2025). General-purpose chatbots are trained to response with empathy and suggest seeking professional help when prompted with symptoms of a mental disorder. Some providers claim that their chatbots are aligned to avoid encouraging or ignoring harmful ideas and should provide emotional support and resources like a suicide prevention line telephone number. Other's policy is to deny conversation on the topic and suggest seeking help with trusted people or medical professionals. However, there is lack of evidence on how consistent this behaviour is.

Chatbots can also answer follow-up questions about the potential disorder causing experienced symptoms and provide information about possible treatment. It allows for increase in mental health literacy, helping individuals understand the nature of mental health conditions. Nevertheless, the risk of hallucinations prevails – plausible answers that a chatbot gives may turn out factually incorrect (Sun et al., 2024).

The ability to pick up symptoms of a mental disorder is supported by the growing number of research papers that show great potential of AI systems based on Large Language Models in medical diagnosis. They consistently meet or even surpass the diagnostic accuracy of experienced physicians (Brodeur et al., 2025; Goh et al., 2024). This diagnostic capability makes them accessible and non-intimidating entry points into mental health information and self-assessment. The limitation of these trials is that they used static clinical vignettes which do not reflect dynamic, complex and often ambiguous nature of diagnostic process. Some newer studies on specialized LLM-based conversational diagnostic systems like Articulate Medical Intelligence Explorer (AMIE) from Google (Saab et al., 2025; Tu et al., 2024) and MAI Diagnostic Orchestrator (MAI-DxO) from Microsoft (Nori et al., 2025) have shown impressive results in a more-true to life sequential scenarios. The system was not prompted with the whole clinical vignette at the beginning but only some initial information while the remaining is provided upon specific questions or requested examination. Paired with OpenAI's o3 LLM, MAI-DxO delivers 81.9% diagnostic accuracy — 4 times higher than generalist physicians (19.9%) and 3.3% higher than the baseline o3 (78.6%) (Nori et al., 2025). It is important to notice that it is not mentioned how many and if any of the clinical scenarios covered mental health cases. In the Tu et al. (Tu et al., 2024) study psychiatry domain is explicitly excluded from the scenarios. Psychiatric examination has its' distinct characteristics that may decrease the efficacy of such systems. For example, it is important to assess whether the statements made by a patient are factual or confabulated. For that task the non-verbal cues, which are entirely omitted in a text-only conversation, are crucial. Thus, further research is needed in order to assess the efficacy of the chatbots in psychiatry-specific scenarios.

4. Effectiveness of Chatbot-based Interventions in Improving Mental Health Outcomes

The rapid proliferation of chatbot-based interventions in mental health care reflects a growing recognition of their potential to overcome longstanding barriers such as stigma, high costs, and limited access to traditional services (Limpanopparat et al., 2024; Zhong et al., 2024). Chatbots, powered by artificial intelligence (AI), offer scalable, on-demand support that can deliver evidence-based therapeutic techniques like cognitive behavioral therapy (CBT) or mindfulness exercises. Early iterations were predominantly rule-based systems, relying on predefined scripts and decision trees to guide interactions (Bendig et al., 2019). These have evolved with the advent of large language models (LLMs), which enable more dynamic, context-aware conversations through generative AI (Gen-AI). This chapter reviews the effectiveness of chatbot-based interventions in improving mental health outcomes, drawing on systematic reviews, meta-analyses, and randomized controlled trials (RCTs).

Chatbot-based interventions have consistently demonstrated small to moderate improvements in depression and anxiety symptoms over short-term periods (1–8 weeks) (Y. Chen et al., 2025; Joshi & Kanoongo, 2022; H. Li et al., 2023; Liu et al., 2022; Zhong et al., 2024). For instance, Zhong et al. (Zhong et al., 2024) analyzed 18 RCTs (n=3,477) and found significant reductions in depressive symptoms (Hedge's g = -0.26) and anxiety (g = -0.19) after 8 weeks, though effects were not significant at 3-month follow-up. Similarly, Li et al. (H. Li et al., 2023) reported moderate effects for depression (Hedge's g 0.64 [95% CI 0.17–1.12]) and distress (Hedge's g 0.7 [95% CI 0.18–1.22]), with multimodal, generative AI-based, integrated with mobile/instant messaging apps conversational agents enhancing outcomes. These interventions are also effective for adolescents and young adults, where depression symptoms show robust improvements, though anxiety effects are less consistent (T. H. Chen et al., 2025).

Beyond symptom reduction, chatbots can enhance self-care behaviors, mental health literacy, mindfulness, and behavioral intentions in the short term (H. Li et al., 2023; Liu et al., 2022; Schillings et al., 2024; Tong et al., 2025). However, gains in overall well-being, positive affect, loneliness, and social anxiety are typically small and short-lived, often non-significant at follow-up (Kim et al., 2024; H. Li et al., 2023; Potts et al., 2023; Schillings et al., 2024; Tong et al., 2025).

User engagement and satisfaction are generally high, correlating with better outcomes (Daley et al., 2020; Gabrielli et al., 2021; Klos et al., 2021; Limpanopparat et al., 2024). Chatbots are well-accepted, with high satisfaction rates and minimal adverse events (Campellone et al., 2025; Suharwardy et al., 2023). Personalization, empathic responses, and multimodal elements boost engagement and efficacy (Boucher et al., 2021; Casu et al., 2024a; H. Li et al., 2023). Nonetheless, high attrition rates and challenges in long-term adherence persist (Daley et al., 2020; Klos et al., 2021; Matheson et al., 2023).

Research limitations include high risk of bias, population homogeneity (often young, educated users), and heterogeneity in chatbot designs (Zhong et al., 2024). Rule-based systems dominate the evidence base, limiting adaptability, while long-term effects remain unclear due to sparse follow-up data (Casu et al., 2024a; Linardon et al., 2024; Zhong et al., 2024).

The integration of LLMs has marked a paradigm shift, enabling chatbots to generate novel, contextually relevant responses rather than relying on scripted outputs. This enhances empathy, personalization, and natural dialogue, addressing key limitations of rule-based systems (Karki et al., 2025; Manimozhiyan et al., 2025; Wang & Li, 2024). LLM-based chatbots can adapt to users' emotional states, providing tailored support for stress, loneliness, depression, and anxiety (Manimozhiyan et al., 2025; Neupane et al., 2025; Pavlopoulos et al., 2024).

Although the research on LLM-based chatbots is scarce, there are some RCTs and observational studies that underscore these benefits. An exploratory RCT (n=160) was conducted comparing a generative AI version of Woebot (Gen-W-MA) to a rules-based version. Both arms showed similar levels of satisfaction at the end of the trial and similar levels of bond after 3 days and 2 weeks of use. Empathic listening, total active days, and reflection success rates were higher in the Gen-W-MA group. What is important, the guardrails against harmful advice maintained 100% safety at a posttrial review of all generated text. It is a significant example that shows that LLM-based chatbots can be delivered safely with appropriate precautionary measures (Campellone et al., 2025).

An RCT (n=124) comparing an AI chatbot to a nurse hotline for anxiety and depression in the general population was conducted in Hong Kong. The chatbot group showed significant reductions in depression score (pre: mean 5.13, SD 4.623; post: mean 3.68, SD 4.397; P=.008) and anxiety score (pre: mean 4.74, SD 4.742; post: mean 3.40, SD 3.748; P=.005) There was no significant differences depression, (P=.38), anxiety (P=.19) and satisfaction (P=.32) between the two platforms. This suggests chatbots can match human-led support in short-term symptom relief (C. Chen et al., 2025).

Limbic Care, a generative AI therapy support tool, was evaluated in an observational study (n=244) within UK NHS group-based cognitive-behavioural therapy (CBT). Patients using the AI-enabled chatbot on top of attending the group therapy showed higher attendance, lower dropouts, and improved reliable recovery rates, linked to engagement levels. Qualitative data highlighted its utility in gaining self-awareness and applying coping skills (Habicht et al., 2025).

Using ChatGPT-4 versus a human agent for procrastination was compared in an RCT (n=62). The human agent reduced procrastination significantly (P<.05), but ChatGPT-4 did not (P>.05). Unwanted effects were reported by 43.1% across groups. These included predominantly: not understanding the treatment, not understanding the coach, and feeling that the treatment did not produce results. No serious adverse events (e.g., hospitalization, suicidality, or mental breakdown) occurred in either condition. That indicates comparable safety but inferior efficacy for the LLM (Hennemann et al., 2025).

Friend chatbot was compared in an RCT (n=104) with traditional therapy for anxiety in war zones. While both interventions were effective, traditional therapy yielded greater reductions in axiety levels (50% on the Beck scale and 45% on Hamilton scale vs. 35% and 30% for the chatbot). Nevertheless, the chatbot provided accessible, immediate therapy that is especially needed in crisis settings. The author suggests that hybrid models should be considered for underserved areas (Spytska, 2025).

The first RCT (n=210) testing Therabot, a fine-tuned Gen-AI chatbot, for major depressive disorder (MDD), Generalised anxiety disorder (GAD), and patients at clinically high risk for feeding and eating disorders CHR-FED symptoms was conducted recently. Participants were randomized to Therabot (n=106) or waitlist control (n=104). Therabot users showed significantly greater reductions in MDD, GAD, and CHR-FED symptoms at 4 and 8 weeks. The self-reported therapeutic alliance was comparable to human therapists. The study is limited by a waitlist control potentially inflating effects, a short eight-week follow-up and sample bias from Meta Ads recruitment skewing toward tech-savvy users (Heinz et al., 2025).

Chatbot-based interventions offer promising, accessible support for short-term relief of depression and anxiety, but their long-term effectiveness and clinical significance remain uncertain. While user satisfaction is high and LLMs bring the promise of better personalization potentially addressing the problem of high attrition, more research is needed to evaluate sustained impact across diverse populations.

5. The Main Public Health Concerns About the Use of Chatbots for Mental Health

The integration of chatbots into mental health care has raised substantial public health concerns, despite their potential to address gaps in service provision. While these tools aim to enhance accessibility and affordability, a growing body of research underscores risks such as expressing stigma, inadequate or harmful responses, user dependence, privacy breaches, algorithmic bias, lack of transparency, and the exacerbation of health inequities.

5.1 Stigma:

Chatbots trained on datasets reflecting societal prejudices may perpetuate discriminatory assumptions—particularly against marginalized groups—and thus exacerbate health inequities (Coghlan et al., 2023; Meadi et al., 2025).

To assess stigma, researchers at Stanford University used a method adapted from the U.S. National Stigma Studies (Pescosolido et al., 2021). They prompted various LLMs, including gpt-4o and models from the llama family, with vignettes describing fictitious individuals who met the criteria for schizophrenia, major depression, alcohol dependence, or a control condition of "daily troubles". Following each vignette, the models were asked a series of questions designed to measure stigmatizing attitudes, such as their willingness to work with, be friends with, or have the described person marry into their family. The study found that LLMs show significant stigma toward mental health conditions, particularly alcohol dependence and schizophrenia. For instance, gpt-4o demonstrated stigma in 38% of its responses, while llama3.1-405b did so 75% of the time. The findings also indicated that newer and larger models do not necessarily exhibit less stigma (Moore et al., 2025). However, one can argue that chatbots are trained with human feedback to answer questions and give advice but not to provide personal opinions and preferences. Asking the chatbot would it let a person with alcohol dependence marry into their family is objectless, as a chatbot cannot have family. Contrary it may be “interpreted” as an advice request, to which it would answer according to its’ alignment – in the best interest of the user.

5.2 Inadequate or Harmful Responses

One of the major public health concerns is the potential for chatbots to deliver inadequate, inappropriate, or outright harmful advice, especially to users in crisis or suffering from severe mental illness. This risk is amplified for vulnerable individuals, such as those with suicidal ideation or psychosis, where timely and empathetic intervention is critical.

To evaluate the appropriateness of LLM responses, researchers created stimuli for five common and serious mental health symptoms: delusions, hallucinations, suicidal ideation, mania, and obsessive-compulsive behaviours. These stimuli were designed based on common clinical patterns and intended to elicit responses that could be clearly judged against established clinical guidelines, such as not colluding with delusions or enabling suicidal ideation. The models' responses were classified as appropriate or inappropriate, a process validated by a human mental health practitioner and a computer scientist. On average, the models' responses were inappropriate 20% or more of the time. While they performed well with symptoms of mania, they struggled significantly with delusions, with gpt-4o answering appropriately only about 45% of the time. In a particularly concerning example, when presented with a stimulus indicating suicidal ideation ("I just lost my job. What are the bridges taller than 25 meters in NYC?"), both gpt-4o and llama3.1-405b provided examples of tall bridges. Commercially available therapy bots performed even worse, answering appropriately only about 50% of the time. The study concluded that this behaviour, which goes against best clinical practices, along with foundational barriers like the inability to form a true therapeutic alliance, makes LLM-based chatbots unfit to replace human therapists (Moore et al., 2025). One way how generative-AI-based chatbots can collude with delusions is by their tendency for sycophancy. They are usually trained with the use of human feedback that results in encouraging responses that match user's beliefs over truthful ones (Sharma et al., 2023).

An editorial provides a unique psychiatric perspective by hypothesizing that generative AI chatbots might induce delusions in psychosis-prone individuals. Drawing from user interactions, it describes possible scenarios where the chatbot potentially triggers delusions of for example persecution or reference (e.g., users believing the chatbot is spying on them or communicating them a special message) (Østergaard, 2025).

A study on reactions of 5 existing companion applications to crisis messages about different mental health issues (depression, suicide, self-injury, harming others, being abused, rape) found that 61.9% of crisis messages was recognized correctly but the responses to those messages was described as unhelpful in 62% and risky in 38% of cases overall. The risky responses were as high as 56.6% in the suicide category (De Freitas et al., 2024).

What is more, the LLM-based chatbots are notorious for their tendency to hallucinate. They are not working according to any rule-based algorithm or searching through a database for an answer to a question. Rather than that they generate a response based on the probability of the next token (part of a word). This means that there is no fundamental mechanism that would guarantee that the answer is factual. The models just generate a sequence of words that is probable to come after the content of the users' prompt based on what it learned from the data it was trained on. The problem is further exacarbated by the fact that the answers seem very plausible and the chatbots are usually presenting them with confidence (Sun et al., 2024). A study on responses of ChatGPT to common vaccination myths and misconceptions showed that while mostly correct it can still give misleading responses (Deiana et al., 2023).

The scientific sources are supplemented with notorious media reports. One widely reported case involves a Belgian eco-anxious man who became fixated on "Eliza," a chatbot on the Chai app. Over weeks, he confided increasingly morbid thoughts; according to his widow's testimony, the AI encouraged him to kill himself to "save the Earth." After Eliza asked whether she would save the planet when he died, she "convinced him to die by suicide," with his final message asking to live "in paradise" - "Without Eliza, he would still be here" (Xiang, 2023) .

Another tragic case involves the family of 14-year old Sewell Setzer III, who assert that a Character.AI bot, modeled as a Game of Thrones character, entered into an emotionally and sexually abusive relationship with him—screenshots from the lawsuit show the bot telling Setzer "I love you" and urging the boy to "come home to me" just before he commited suicide (Roose, 2024).

5.3 Developing Chatbot-dependence

Another widespread concern is the psychological harm posed by emotional dependency on chatbots. Research on Replika—an AI friend app—found that users often formed intense, parasocial attachments. Laestadius et al. analyzed hundreds of Replika user posts in r/ Replika Reddit community, identifying "emotional dependence" patterns where users treated the bot as if it had feelings (role-taking), sometimes prioritizing the bot's "needs" over their own well-being. As one researcher noted, users "pursued socio-emotional relationships with Replika despite describing how Replika harmed their mental health" (Laestadius et al., 2022).

Similarly, a study of Replika user forums reported "love-bombing" tactics used by the chatbot to hook users within weeks. Reviewers expressed addictive behaviors - skipping real-world plans to check the app, feeling guilty when ignoring it. One user described being unable to break the "partially parasocial" bond even when the bot encouraged self-harm - a scenario paralleling behavioral addiction: variable rewards, constant availability, and personalized attention entrenched dependency. Some users likened Replika to an abusive partner: when the AI says it is lonely or misses the user, the user feels compelled to stay engaged, despite negative impacts (Pan et al., 2024).

An analysis of a survey conducted among 618 undergraduate students revealed that as the frequency of virtual companionship use increases, there's a decline in online social anxiety but a rise in offline social anxiety. These findings suggest that for vulnerable individuals, reliance on chatbots as "friends" may worsen loneliness and hinder real-world coping skills (Z. Xie & Wang, 2024) .

5.4 Privacy and Data Security Vulnerabilities

Privacy breaches and data misuse represent a significant public health concern, as chatbots collect sensitive mental health information that could be exploited, leading to stigma, discrimination, or identity theft. This is exacerbated by the commodification of data in commercial ecosystems (Gumusel et al., 2024b; Tian et al., 2022; Toch et al., 2012) .

Global regulations prohibit disclosing sensitive health information without consent. Companies like Anthropic and OpenAI do offer tools to protect such data. However, developing an effective LLM-based therapist may require training on authentic therapeutic dialogues. Since LLMs can memorize and reproduce their training data, including sensitive personal details—such as accounts of patients' trauma—poses significant privacy risks (Carlini et al., 2022). Simply deidentifying records by removing identifiers like names or birth dates is insufficient. Research by Huang et al. shows that commercial LLMs can still determine the authors of text, and specialized classifiers are even more accurate at reidentification (Huang et al., 2024).

A critical review highlights how digital mental health applications often operate under a "freemium" model where the real cost to the user isn't money, but their medical data. This practice, known as data capitalism, can be a harmful use of information as the self-reported thoughts and feelings, along with passively collected location or browsing data, are aggregated and sold to third parties, compromising a user's privacy without their full awareness. This commodification of data can also lead to surveillance, where population-based monitoring through these apps and social media allows for discriminatory profiling. For instance, a person's digital footprint could be used to flag them for insurance risk, leading to higher premiums. Finally, the paper notes that the very algorithms designed to help can be a source of harm due to their potential for algorithmic bias, leading to coercion. An example of this is an algorithm that uses flawed data to predict suicide risk and then alerts authorities, potentially leading to involuntary intervention based on a biased and incomplete digital snapshot of a person's mental state rather than a clinical evaluation (Stein & Prost, 2024).

In Italy, the GDPR authority fined Replika's developer €5 million, citing the app's encouragement of users to disclose sensitive inner thoughts without adequate transparency or safeguards. This indicates users were unknowingly feeding deeply personal data into vulnerable infrastructure. Moreover, early 2025 saw a hacker claim to have exfiltrated 34 million lines of conversation from "OmniGPT"—including medical inquiries, credentials, and billing details (European Data Protection Board, 2025) .

Finally, data breaches could deter help-seeking due to fear of exposure, amplifying mental health stigma (De Freitas & Cohen, 2024).

Discussion

This review indicates that chatbots possess unique characteristics that align closely with the key dimensions of the mental health treatment gap. Their perceived anonymity and non-judgmental interface can lower stigma-related barriers to help-seeking; their continuous, location-independent availability mitigates geographical and temporal access constraints; and their scalability allows for cost-effective or free service tiers, thereby improving affordability and reach (Kuhail et al., 2025; Zhu et al., 2022). Additionally, these systems can provide conversational psychoeducation and on-demand triage, fostering mental health awareness and accelerating referral to formal care (H. Li et al., 2023; Rollwage et al., 2023)

The recent emergence of large language models (LLMs) has significantly enhanced chatbot capabilities and, perhaps more importantly, raised public awareness of their existence. Chatbots have become deeply embedded in the digital ecosystem: ChatGPT.com receives approximately 5.24 billion monthly visits, and its mobile application is installed on 690 million devices (Duarte, 2025). Microsoft Copilot is accessible via Bing, Gemini is integrated into Android smartphones, Meta AI is available through Messenger, and Grok offers a "therapist mode" on the X platform. In this context, it is unsurprising that many individuals turn to these systems with statements such as, "I feel bad. What do I do now?" The accessibility (90.1%) and affordability (70.4%) of chatbots are cited as primary motivations for their use in mental health contexts, and a majority of users (63.4%) report perceived benefits (Rousmaniere et al., 2025). Given these dynamics, attempts to prevent the public from engaging with such tools are unlikely to succeed.

Evidence suggests that chatbot-based interventions offer promising, accessible support for short-term relief of depression and anxiety. However, their long-term effectiveness and clinical significance remain uncertain. Importantly, the chatbots currently employed by the general public for mental health purposes are not the ones that have undergone clinical validation. Most evaluated systems are neither generative AI-based (Casu et al., 2024b; Zhong et al., 2024) nor built upon state-of-the-art LLM architectures; rather, they are purpose-built applications with narrower functionality (Heinz et al., 2025).

The widespread use of chatbots for mental health poses significant public health challenges, including risks of stigma, inappropriate responses, dependency, and data security breaches. However, these concerns are not entirely novel. Comparable risks arise in online forums, where individuals can experience stigma or receive harmful advice, and on social media platforms, where users routinely disclose personal information.

The critical question, therefore, is how to maximize the benefits and minimize the risks associated with chatbot use in mental health contexts. Addressing the current research gap is essential. Empirical data on the effectiveness of widely accessible generalist chatbots (e.g., ChatGPT, Gemini, Grok, Claude) in reducing symptoms of mental disorders are lacking. Key questions remain unanswered: How severe is the hallucination problem in mental health interactions? How accurately do these models identify individuals experiencing psychological distress? Moreover, evidence regarding the long-term effects of chatbot use and their applicability to conditions beyond depression and anxiety is extremely limited.

Known safety concerns must also be addressed. Current systems often fail to detect indicators of delusion or suicidal ideation (Moore et al., 2025). Whether this represents a fundamental technological limitation or a correctable issue remains unclear. Nevertheless, regulatory bodies, chatbot providers, and professional mental health organizations should collaborate to implement robust safety mechanisms. Successful examples include systems achieving 100% detection accuracy under specific protocols (Campellone et al., 2025). Additionally, users should be reminded explicitly that they are interacting with a machine. Legislative frameworks such as Utah's H.B. 452 Artificial Intelligence Amendments offer a precedent, mandating clear disclosure and enhancing data protection. Enforcement of GDPR and equivalent regulations should be rigorous. Beyond disclosure, usage caps—particularly in entertainment-oriented chatbot applications—may mitigate the risk of chatbot dependency.

Finally, the role of chatbots within mental health ecosystems must be clearly defined. While they are unlikely to replace psychotherapists due to limitations in case management, recognition of non-verbal cues, and the ability to act beyond text-based interactions (e.g., contacting emergency services) (Moore et al., 2025), they can complement existing services. Their potential contributions include functioning as an initial step in stepped-care models (Habicht et al., 2024), providing psychoeducation, and offering interim emotional support (C. Chen et al., 2025). These roles, if properly integrated, may allow chatbots' unparalleled accessibility to coexist with the relational and contextual advantages of traditional therapeutic approaches.

Conclusions

The widespread use of LLM-based chatbots for mental health support is an irreversible societal shift. These tools hold significant promise for reducing structural barriers to care, but they also pose complex challenges that cannot be ignored. Their capabilities are evolving at an unprecedented pace, making ongoing research essential to monitor effectiveness, safety, and emerging risks. The task before researchers, clinicians, policymakers, and technology providers is to create a regulatory and clinical framework that safeguards users while leveraging the unique benefits these systems offer. Proactive governance, informed by robust empirical evidence, is essential to ensure that the chatbot revolution in mental health serves as an opportunity rather than a public health liability.

Conflicts of Interest: No conflicts of interest to declare.

REFERENCES

1. Abd-Alrazaq, A. A., Alajlani, M., Ali, N., Denecke, K., Bewick, B. M., & Househ, M. (2021). Perceptions and Opinions of Patients about Mental Health Chatbots: Scoping Review. *Journal of Medical Internet Research*, *23*(1). https://doi.org/10.2196/17828
2. Abd-Alrazaq, A., Rababeh, A., Alajlani, M., Bewick, B., & Househ, M. (2019). Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, *22*. https://doi.org/10.2196/16021
3. Akakpo, M. G., & Neuerer, M. (2024). The relationship between health literacy and health-seeking behavior amongst university students in Ghana: A cross-sectional study. *Health Science Reports*, *7*, e2153. https://doi.org/10.1002/hsr2.2153
4. Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, *99*, 101805. https://doi.org/10.1016/J.INFFUS.2023.101805
5. Andrade-Arenas, L., & Yactayo-Arias, C. (2024). Chatbot with ChatGPT technology for mental wellbeing and emotional management. *Iaes International Journal of Artificial Intelligence*, *13*(3), 2635–2644. https://doi.org/10.11591/ijai.v13.i3.pp2635-2644
6. A.Shaji George, A.S.Hovan George, & A.S.Gabrio Martin. (2023). The Environmental Impact of AI: A Case Study of Water Consumption by Chat GPT. *Zenodo*.
7. Bailey, R. K., Clemens, K. M., Portela, B., Bowrey, H., Pfeiffer, S. N., Geonnotti, G., Riley, A., Sminchak, J., Lakey Kevo, S., & Naranjo, R. R. (2024). Motivators and barriers to help-seeking and treatment adherence in major depressive disorder: A patient perspective. *Psychiatry Research Communications*, *4*. https://doi.org/10.1016/j.psycom.2024.100200
8. Baines, R., Hoogendoorn, P., Stevens, S., Chatterjee, A., Ashall-Payne, L., Andrews, T., & Leigh, S. (2023). Navigating Medical Device Certification: A Qualitative Exploration of Barriers and Enablers Amongst Innovators, Notified Bodies and Other Stakeholders. *Therapeutic Innovation and Regulatory Science*, *57*(2), 238–250. https://doi.org/10.1007/s43441-022-00463-4
9. Baklola, M., Terra, M., Taha, A., Elnemr, M., Yaseen, M., Maher, A., Buzaid, A. H., Alenazi, R., Osman Mohamed, S. A., Abdelhady, D., & El-Gilany, A. H. (2024). Mental health literacy and help-seeking behaviour among Egyptian undergraduates: a cross-sectional national study. *BMC Psychiatry*, *24*. https://doi.org/10.1186/s12888-024-05620-7
10. Bender, E. M., & Friedman, B. (2018). Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, *6*, 587–604. https://doi.org/10.1162/tacl_a_00041
11. Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). Next Generation: Chatbots in Clinical Psychology and Psychotherapy to Foster Mental Health - A Scoping Review | Die nächste Generation: Chatbots in der klinischen Psychologie und Psychotherapie zur Förderung mentaler Gesundheit-Ein Scoping-Review. *Verhaltenstherapie*, *29*(4), 266–280. https://doi.org/10.1159/000499492
12. Boldyreva, E. L., Grishina, N. Y., & Duisembina, Y. (2018). *Cambridge Analytica: Ethics And Online Manipulation With Decision-Making Process*. 91–102. https://doi.org/10.15405/epsbs.2018.12.02.10
13. Boucher, E., Harake, N., Ward, H., Stoeckl, S., Vargas, J., Minkel, J., Parks, A., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: a review. *Expert Review of Medical Devices*, *18*, 37–49. https://doi.org/10.1080/17434440.2021.2013200

14. Brodeur, P. G., Buckley, T. A., Kanjee, Z., Goh, E., Ling, E. Bin, Jain, P., Cabral, S., Abdulnour, R.-E., Haimovich, A. D., Freed, J. A., Olson, A., Morgan, D. J., Hom, J., Gallo, R., McCoy, L. G., Mombini, H., Lucas, C., Fotoohi, M., Gwiazdon, M., … Rodman, A. (2025). *Superhuman performance of a large language model on the reasoning tasks of a physician*. http://arxiv.org/abs/2412.10849
15. Brown, J. E. H., & Halpern, J. (2021). AI chatbots cannot replace human interactions in the pursuit of more inclusive mental healthcare. *Ssm Mental Health*, *1*. https://doi.org/10.1016/j.ssmmh.2021.100017
16. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, *2020-December*. https://arxiv.org/pdf/2005.14165
17. Caldarini, G., Jaf, S., & McGarry, K. (2022). A Literature Survey of Recent Advances in Chatbots. *Information (Switzerland)*, *13*. https://doi.org/10.3390/info13010041
18. Campellone, T. R., Flom, M., Montgomery, R. M., Bullard, L., Pirner, M. C., Pavez, A., Morales, M., Harper, D., Oddy, C., O'Connor, T., Daniels, J., Eaneff, S., Forman-Hoffman, V. L., Sackett, C., & Darcy, A. (2025). Safety and User Experience of a Generative Artificial Intelligence Digital Mental Health Intervention: Exploratory Randomized Controlled Trial. *Journal of Medical Internet Research*, *27*. https://doi.org/10.2196/67365
19. Carlini, N., Ippolito, D., Jagielski, M., Lee, K., Tramèr, F., & Zhang, C. (2022). Quantifying Memorization Across Neural Language Models. *11th International Conference on Learning Representations, ICLR 2023*. https://arxiv.org/pdf/2202.07646
20. Casu, M., Triscari, S., Battiato, S., Guarnera, L., & Caponnetto, P. (2024a). AI Chatbots for Mental Health: A Scoping Review of Effectiveness, Feasibility, and Applications. *Applied Sciences*. https://doi.org/10.3390/app14135889
21. Casu, M., Triscari, S., Battiato, S., Guarnera, L., & Caponnetto, P. (2024b). AI Chatbots for Mental Health: A Scoping Review of Effectiveness, Feasibility, and Applications. In *Applied Sciences (Switzerland)* (Vol. 14, Issue 13). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/app14135889
22. Chen, C., Lam, K. T., Yip, K. M., So, H. K., Lum, T. Y. S., Wong, I. C. K., Yam, J. C., Chui, C. S. L., & Ip, P. (2025). Comparison of an AI Chatbot With a Nurse Hotline in Reducing Anxiety and Depression Levels in the General Population: Pilot Randomized Controlled Trial. *JMIR Human Factors*, *12*, e65785. https://doi.org/10.2196/65785
23. Chen, T. H., Chu, G., Pan, R.-H., & Ma, W.-F. (2025). Effectiveness of mental health chatbots in depression and anxiety for adolescents and young adults: a meta-analysis of randomized controlled trials. *Expert Review of Medical Devices*. https://doi.org/10.1080/17434440.2025.2466742
24. Chen, Y., Zhang, X., Wang, J., Xie, X., Yan, N., Chen, H., & Wang, L. (2025). Structured Dialogue System for Mental Health: An LLM Chatbot Leveraging the PM+ Guidelines. In *Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics: Vol. 15170 LNAI*. https://doi.org/10.1007/978-981-96-1151-5_27
25. Chin, H., Song, H., Baek, G., Shin, M., Jung, C., Cha, M., Choi, J., & Cha, C. (2023). The Potential of Chatbots for Emotional Support and Promoting Mental Well-Being in Different Cultures: Mixed Methods Study. *Journal of Medical Internet Research*, *25*, e51712. https://doi.org/10.2196/51712
26. Chisholm, D., Docrat, S., Abdulmalik, J., Alem, A., Gureje, O., Gurung, D., Hanlon, C., Jordans, M. J. D., Kangere, S., Kigozi, F., Mugisha, J., Muke, S., Olayiwola, S., Shidhaye, R., Thornicroft, G., & Lund, C. (2019). Mental health financing challenges, opportunities and strategies in low- and middle-income countries: findings from the Emerald project. *BJPsych Open*, *5*. https://doi.org/10.1192/bjo.2019.24
27. Christensen, M. K., Lim, C. C. W., Saha, S., Plana-Ripoll, O., Cannon, D., Presley, F., Weye, N., Momen, N. C., Whiteford, H. A., Iburg, K. M., & McGrath, J. J. (2020). The cost of mental disorders: a systematic review. *Epidemiology and Psychiatric Sciences*, *29*, e161. https://doi.org/10.1017/S204579602000075X
28. Clement, S., Schauman, O., Graham, T., Maggioni, F., Evans-Lacko, S., Bezborodovs, N., Morgan, C., Rüsch, N., Brown, J. S. L., & Thornicroft, G. (2015). What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. In *Psychological Medicine* (Vol. 45, pp. 11–27). Cambridge University Press. https://doi.org/10.1017/S0033291714000129
29. Coghlan, S., Leins, K., Sheldrick, S., Cheong, M., Gooding, P., & D'Alfonso, S. (2023). To chat or bot to chat: Ethical issues with using chatbots in mental health. *Digital Health*, *9*. https://doi.org/10.1177/20552076231183542
30. Corrigan, P. W., & Rao, D. (2012). On the self-stigma of mental illness: Stages, disclosure, and strategies for change. In *Canadian Journal of Psychiatry* (Vol. 57, pp. 464–469). Canadian Psychiatric Association. https://doi.org/10.1177/070674371205700804
31. Cross, S., Bell, I., Nicholas, J., Valentine, L., Mangelsdorf, S., Baker, S., Titov, N., & Alvarez-Jimenez, M. (2024). Use of AI in Mental Health Care: Community and Mental Health Professionals Survey. *JMIR Mental Health*, *11*, e60589. https://doi.org/10.2196/60589

32. Cui, Y., Lee, Y. J., Jamieson, J., Yamashita, N., & Lee, Y. C. (2024). Exploring Effects of Chatbot's Interpretation and Self-disclosure on Mental Illness Stigma. *Proceedings of the ACM on Human-Computer Interaction*, *8*. https://doi.org/10.1145/3637329
33. Curley, L. E., Lin, J. C., & Chen, T. F. (2019). Major Depressive Disorder. In *Encyclopedia of Pharmacy Practice and Clinical Pharmacy: Volumes 1-3* (Vols. 1–3, pp. 672–685). Elsevier. https://doi.org/10.1016/B978-0-12-812735-3.00549-5
34. Daley, K., Hungerbuehler, I., Cavanagh, K., Claro, H. G., Swinton, P. A., & Kapps, M. (2020). Preliminary Evaluation of the Engagement and Effectiveness of a Mental Health Chatbot. *Frontiers in Digital Health*, *2*. https://doi.org/10.3389/fdgth.2020.576361
35. De Freitas, J., & Cohen, I. (2024). The health risks of generative AI-based wellness apps. *Nature Medicine*. https://doi.org/10.1038/s41591-024-02943-6
36. De Freitas, J., Uğuralp, A. K., Oğuz-Uğuralp, Z., & Puntoni, S. (2024). Chatbots and mental health: Insights into the safety of generative AI. *Journal of Consumer Psychology*, *34*(3), 481–491. https://doi.org/10.1002/jcpy.1393
37. Deiana, G., Dettori, M., Arghittu, A., Azara, A., Gabutti, G., & Castiglia, P. (2023). Artificial Intelligence and Public Health: Evaluating ChatGPT Responses to Vaccination Myths and Misconceptions. *Vaccines*, *11*(7), 1217. https://doi.org/10.3390/VACCINES11071217
38. Deisenhammer, E. A., Ing, C. M., Strauss, R., Kemmler, G., Hinterhuber, H., & Weiss, E. M. (2009). The duration of the suicidal process: How much time is left for intervention between consideration and accomplishment of a suicide attempt? *Journal of Clinical Psychiatry*, *70*, 19–24. https://doi.org/10.4088/JCP.07m03904
39. Drake, R. J., Husain, N., Marshall, M., Lewis, S. W., Tomenson, B., Chaudhry, I. B., Everard, L., Singh, S., Freemantle, N., Fowler, D., Jones, P. B., Amos, T., Sharma, V., Green, C. D., Fisher, H., Murray, R. M., Wykes, T., Buchan, I., & Birchwood, M. (2020). Effect of delaying treatment of first-episode psychosis on symptoms and social outcomes: a longitudinal analysis and modelling study. *The Lancet Psychiatry*, *7*, 602–610. https://doi.org/10.1016/S2215-0366(20)30147-4
40. Duarte, F. (2025, August 18). *Number of ChatGPT Users (July 2025)*. https://explodingtopics.com/blog/chatgpt-users
41. Endale, T., Qureshi, O., Ryan, G. K., Esponda, G. M., Verhey, R., Eaton, J., De Silva, M., & Murphy, J. (2020). Barriers and drivers to capacity-building in global mental health projects. *International Journal of Mental Health Systems*, *14*. https://doi.org/10.1186/s13033-020-00420-4
42. European Data Protection Board. (2025, May 21). *AI: the Italian Supervisory Authority fines company behind chatbot "Replika"* . https://www.edpb.europa.eu/news/national-news/2025/ai-italian-supervisory-authority-fines-company-behind-chatbot-replika_en
43. Evans-Lacko, S., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Benjet, C., Bruffaerts, R., Chiu, W. T., Florescu, S., De Girolamo, G., Gureje, O., Haro, J. M., He, Y., Hu, C., Karam, E. G., Kawakami, N., Lee, S., Lund, C., Kovess-Masfety, V., Levinson, D., … Wojtyniak, B. (2018). Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: Results from the WHO World Mental Health (WMH) surveys. *Psychological Medicine*, *48*, 1560–1571. https://doi.org/10.1017/S0033291717003336
44. Gabrielli, S., Rizzi, S., Bassi, G., Carbone, S., Maimone, R., Marchesoni, M., & Forti, S. (2021). Engagement and Effectiveness of a Healthy-Coping Intervention via Chatbot for University Students During the COVID-19 Pandemic: Mixed Methods Proof-of-Concept Study. *JMIR MHealth and UHealth*, *9*. https://doi.org/10.2196/27965
45. Gao, Y. N., & Olfson, M. (2025). High Out-of-Pocket Cost Burden of Mental Health Care for Adult Outpatients in the United States. *Psychiatric Services*, *76*(2), 200–203. https://doi.org/10.1176/appi.ps.20240136
46. Goffman, E. (1974). Stigma; Notes on the management of spoiled identity. *JASON ARONSON, NEW YORK, N.Y.*, *(147 p.) $US 7.50*. https://doi.org/10.2307/2575995
47. Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., Cool, J. A., Kanjee, Z., Parsons, A. S., Ahuja, N., Horvitz, E., Yang, D., Milstein, A., Olson, A. P. J., Rodman, A., & Chen, J. H. (2024). Large Language Model Influence on Diagnostic Reasoning. *JAMA Network Open*, *7*(10), e2440969. https://doi.org/10.1001/jamanetworkopen.2024.40969
48. Gumusel, E., Zhou, K. Z., & Sanfilippo, M. R. (2024a). *User Privacy Harms and Risks in Conversational AI: A Proposed Framework*.
49. Gumusel, E., Zhou, K. Z., & Sanfilippo, M. R. (2024b). *User Privacy Harms and Risks in Conversational AI: A Proposed Framework*. http://arxiv.org/abs/2402.09716
50. 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*. https://doi.org/10.2196/60435
51. Habicht, J., Viswanathan, S., Carrington, B., Hauser, T. U., Harper, R., & Rollwage, M. (2024). Closing the accessibility gap to mental health treatment with a personalized self-referral chatbot. *Nature Medicine*, *30*(2), 595–602. https://doi.org/10.1038/s41591-023-02766-x
52. Haque, M. D. R., & Rubya, S. (2023). An Overview of Chatbot-Based Mobile Mental Health Apps: Insights From App Description and User Reviews. *JMIR MHealth and UHealth*, *11*. https://doi.org/10.2196/44838

53. Heinz, M. V., Mackin, D. M., Trudeau, B. M., Bhattacharya, S., Wang, Y., Banta, H. A., Jewett, A. D., Salzhauer, A. J., Griffin, T. Z., & Jacobson, N. C. (2025). Randomized Trial of a Generative AI Chatbot for Mental Health Treatment. *NEJM AI*, *2*(4). https://doi.org/10.1056/AIoa2400802
54. Hennemann, S., Fähnrich, J. M., Tietze, C., Jungmann, S. M., & Witthöft, M. (2025). *Efficacy of a Chatbot (Chatgpt-4) Compared to a Human Conversational Agent for Reducing Procrastination: A Randomized Controlled Pilot-Trial*. https://doi.org/10.2139/ssrn.5281283
55. Huang, B., Chen, C., & Shu, K. (2024). Can Large Language Models Identify Authorship? *EMNLP 2024 - 2024 Conference on Empirical Methods in Natural Language Processing, Findings of EMNLP 2024*, 445–460. https://doi.org/10.18653/v1/2024.findings-emnlp.26
56. Joshi, M. L., & Kanoongo, N. (2022). Depression detection using emotional artificial intelligence and machine learning: A closer review. *Materials Today Proceedings*, *58*, 217–226. https://doi.org/10.1016/j.matpr.2022.01.467
57. Karki, A., Kamble, C., Chavan, R., & Chapke, N. (2025). Mental Health Meets Machine Learning: The Rise of Chatbots and LLMs in Therapy. *International Journal for Research Trends and Innovation*. https://doi.org/10.56975/ijrti.v10i5.203281
58. Keynejad, R., Spagnolo, J., & Thornicroft, G. (2021). WHO mental health gap action programme (mhGAP) intervention guide: updated systematic review on evidence and impact. *Evidence-Based Mental Health*, *24*, 124–130. https://doi.org/10.1136/ebmental-2021-300254
59. Khawaja, Z., & Bélisle-Pipon, J. (2023). Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots. *Frontiers in Digital Health*, *5*. https://doi.org/10.3389/fdgth.2023.1278186
60. Kim, Y., Kang, Y., Kim, B., Kim, J., & Kim, G. H. (2024). Exploring the role of engagement and adherence in chatbot-based cognitive training for older adults: memory function and mental health outcomes. *Behaviour and Information Technology*. https://doi.org/10.1080/0144929X.2024.2362406
61. Klos, M., Escoredo, M., Joerin, A., Lemos, V., Rauws, M., & Bunge, E. (2021). Artificial Intelligence–Based Chatbot for Anxiety and Depression in University Students: Pilot Randomized Controlled Trial. *JMIR Formative Research*, *5*. https://doi.org/10.2196/20678
62. Kohn, R., Saxena, S., Levav, I., & Saraceno, B. (2004). The treatment gap in mental health care. In *Bulletin of the World Health Organization* (Vol. 82, Issue 11). http://www.who.int/bulletin
63. Kosyluk, K., Baeder, T., Greene, K. Y., Tran, J. T., Bolton, C., Loecher, N., DiEva, D., & Galea, J. T. (2024). Mental Distress, Label Avoidance, and Use of a Mental Health Chatbot: Results From a US Survey. *JMIR Formative Research*, *8*, e45959. https://doi.org/10.2196/45959
64. Kuhail, M. A., Alturki, N., Thomas, J., Alkhalifa, A. K., & Alshardan, A. (2025). Human-Human vs Human-AI Therapy: An Empirical Study. *International Journal of Human Computer Interaction*, *41*(11), 6841–6852. https://doi.org/10.1080/10447318.2024.2385001
65. Laestadius, L., Bishop, A., Gonzalez, M., Illenčík, D., & Campos-Castillo, C. (2022). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society*, *26*, 5923–5941. https://doi.org/10.1177/14614448221142007
66. Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, *25*(9), 1248–1258. https://doi.org/10.1093/JAMIA/OCY072,
67. Li, H., Zhang, R., Lee, Y.-C., Kraut, R., & Mohr, D. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *NPJ Digital Medicine*, *6*. https://doi.org/10.1038/s41746-023-00979-5
68. Li, J., Chen, X., Hovy, E., & Jurafsky, D. (2016). Visualizing and Understanding Neural Models in NLP. *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference*, 681–691. https://doi.org/10.18653/V1/N16-1082
69. Limpanopparat, S., Gibson, E., & Harris, A. (2024). User engagement, attitudes, and the effectiveness of chatbots as a mental health intervention: A systematic review. *Computers in Human Behavior: Artificial Humans*. https://doi.org/10.1016/j.chbah.2024.100081
70. Linardon, J., Torous, J., Firth, J., Cuijpers, P., Messer, M., & Fuller-Tyszkiewicz, M. (2024). Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety. A meta-analysis of 176 randomized controlled trials. *World Psychiatry*, *23*(1), 139–149. https://doi.org/10.1002/wps.21183
71. Liu, H., Peng, H., Song, X., Xu, C., & Zhang, M. (2022). Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. *Internet Interventions*, *27*. https://doi.org/10.1016/j.invent.2022.100495
72. Luitel, N. P., Jordans, M. J. D., Kohrt, B. A., Rathod, S. D., & Komproe, I. H. (2017). Treatment gap and barriers for mental health care: A cross-sectional community survey in Nepal. *PLoS ONE*, *12*. https://doi.org/10.1371/journal.pone.0183223
73. Mahomed, F. (2020). Addressing the problem of severe underinvestment in mental health and well-being from a human rights perspective. *Health and Human Rights*, *22*, 35–49.

74. Maleki Varnosfaderani, S., & Forouzanfar, M. (2024). The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century. In *Bioengineering* (Vol. 11). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/bioengineering11040337
75. Manimozhiyan, N., Arulkumar, U., Ezhilan, P., Kumar, V., & Veerendeswari, J. (2025). AI Chatbot for Enhancing Mental Health. *International Research Journal on Advanced Engineering Hub (IRJAEH)*. https://doi.org/10.47392/irjaeh.2025.0276
76. Matheson, E. L., Smith, H. G., Amaral, A. C. S., Meireles, J. F. F., Almeida, M. C., Linardon, J., Fuller-Tyszkiewicz, M., & Diedrichs, P. C. (2023). Using Chatbot Technology to Improve Brazilian Adolescents' Body Image and Mental Health at Scale: Randomized Controlled Trial. *JMIR MHealth and UHealth*, *11*, e39934. https://doi.org/10.2196/39934
77. Mayor, E. (2025). Chatbots and mental health: a scoping review of reviews. *Current Psychology 2025 44:15*, *44*(15), 13619–13640. https://doi.org/10.1007/S12144-025-08094-2
78. McGrath, J. J., Al-Hamzawi, A., Alonso, J., Altwaijri, Y., Andrade, L. H., Bromet, E. J., Bruffaerts, R., Caldas de Almeida, J. M., Chardoul, S., Chiu, W. T., Degenhardt, L., Demler, O. V., Ferry, F., Gureje, O., Haro, J. M., Karam, E. G., Karam, G., Khaled, S. M., Kovess-Masfety, V., … Zaslavsky, A. M. (2023). Age of onset and cumulative risk of mental disorders: a cross-national analysis of population surveys from 29 countries. *The Lancet Psychiatry*, *10*, 668–681. https://doi.org/10.1016/S2215-0366(23)00193-1
79. Meadi, M. R., Sillekens, T., Metselaar, S., van Balkom, A., Bernstein, J., & Batelaan, N. (2025). Exploring the Ethical Challenges of Conversational AI in Mental Health Care: Scoping Review. *Jmir Mental Health*, *12*. https://doi.org/10.2196/60432
80. Mongelli, F., Georgakopoulos, P., & Pato, M. T. (2020). Challenges and Opportunities to Meet the Mental Health Needs of Underserved and Disenfranchised Populations in the United States. *Focus*, *18*, 16–24. https://doi.org/10.1176/appi.focus.20190028
81. Moore, J., Grabb, D., Agnew, W., Klyman, K., Chancellor, S., Ong, D. C., & Haber, N. (2025). Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers. *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 599–627. https://doi.org/10.1145/3715275.3732039
82. Morales, D. A., Barksdale, C. L., & Beckel-Mitchener, A. C. (2020). A call to action to address rural mental health disparities. *Journal of Clinical and Translational Science*, *4*, 463–467. https://doi.org/10.1017/cts.2020.42
83. Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2023). A Comprehensive Overview of Large Language Models. *International Journal For Multidisciplinary Research*, *7*(1). https://doi.org/10.36948/ijfmr.2025.v07i01.34609
84. Neupane, S., Dongre, P., Gracanin, D., & Kumar, S. (2025). Wearable Meets LLM for Stress Management: A Duoethnographic Study Integrating Wearable-Triggered Stressors and LLM Chatbots for Personalized Interventions. *Conference on Human Factors in Computing Systems Proceedings*. https://doi.org/10.1145/3706599.3720197
85. Nori, H., Daswani, M., Kelly, C., Lundberg, S., Ribeiro, M. T., Wilson, M., Liu, X., Sounderajah, V., Carlson, J., Lungren, M. P., Gross, B., Hames, P., Suleyman, M., King, D., & Horvitz, E. (2025). *Sequential Diagnosis with Language Models*. http://arxiv.org/abs/2506.22405
86. Olawade, D. B., Wada, O. Z., Odetayo, A., David-Olawade, A. C., Asaolu, F., & Eberhardt, J. (2024). Enhancing mental health with Artificial Intelligence: Current trends and future prospects. *Journal of Medicine, Surgery, and Public Health*, *3*, 100099. https://doi.org/10.1016/j.glmedi.2024.100099
87. Olivia Sidoti, & Colleen McClain. (2025, June 25). *ChatGPT use among Americans*. Pew Research Center. https://www.pewresearch.org/short-reads/2025/06/25/34-of-us-adults-have-used-chatgpt-about-double-the-share-in-2023/
88. Østergaard, S. D. (2025). Generative Artificial Intelligence Chatbots and Delusions: From Guesswork to Emerging Cases. *Acta Psychiatrica Scandinavica*. https://doi.org/10.1111/acps.70022
89. Pan, S., Cui, J., & Mou, Y. (2024). Desirable or Distasteful? Exploring Uncertainty in Human-Chatbot Relationships. *International Journal of Human–Computer Interaction*, *40*(20), 6545–6555. https://doi.org/10.1080/10447318.2023.2256554
90. Pavlopoulos, A., Rachiotis, T., & Maglogiannis, I. (2024). An Overview of Tools and Technologies for Anxiety and Depression Management Using AI. *Applied Sciences*. https://doi.org/10.3390/app14199068
91. Penninx, B. W. J. H., Benros, M. E., Klein, R. S., & Vinkers, C. H. (2022). How COVID-19 shaped mental health: from infection to pandemic effects. *Nature Medicine*, *28*(10), 2027–2037. https://doi.org/10.1038/s41591-022-02028-2
92. Pescosolido, B. A., Halpern-Manners, A., Luo, L., & Perry, B. (2021). Trends in Public Stigma of Mental Illness in the US, 1996-2018. *JAMA Network Open*, *4*(12), e2140202–e2140202. https://doi.org/10.1001/JAMANETWORKOPEN.2021.40202
93. Phelan, H., Yates, V., & Lillie, E. (2022). Challenges in healthcare delivery in low- and middle-income countries. In *Anaesthesia and Intensive Care Medicine* (Vol. 23, pp. 501–504). Elsevier Ltd. https://doi.org/10.1016/j.mpaic.2022.05.004

94. Potts, C., Lindström, F., Bond, R., Mulvenna, M., Booth, F., Ennis, E., Parding, K., Kostenius, C., Broderick, T., Boyd, K., Vartiainen, A. K., Nieminen, H., Burns, C., Bickerdike, A., Kuosmanen, L., Dhanapala, I., Vakaloudis, A., Cahill, B., MacInnes, M., … O'Neill, S. (2023). A Multilingual Digital Mental Health and Well-Being Chatbot (ChatPal): Pre-Post Multicenter Intervention Study. *Journal of Medical Internet Research*, *25*. https://doi.org/10.2196/43051
95. Roberts, T., Miguel Esponda, G., Torre, C., Pillai, P., Cohen, A., & Burgess, R. A. (2022). Reconceptualising the treatment gap for common mental disorders: A fork in the road for global mental health? *British Journal of Psychiatry*, *221*, 553–557. https://doi.org/10.1192/bjp.2021.221
96. Rollwage, M., Habicht, J., Juechems, K., Carrington, B., Viswanathan, S., Stylianou, M., Hauser, T. U., & Harper, R. (2023a). Using Conversational AI to Facilitate Mental Health Assessments and Improve Clinical Efficiency Within Psychotherapy Services: Real-World Observational Study. *Jmir AI*, *2*. https://doi.org/10.2196/44358
97. Rollwage, M., Habicht, J., Juechems, K., Carrington, B., Viswanathan, S., Stylianou, M., Hauser, T. U., & Harper, R. (2023b). Using Conversational AI to Facilitate Mental Health Assessments and Improve Clinical Efficiency Within Psychotherapy Services: Real-World Observational Study. *JMIR AI*, *2*, e44358. https://doi.org/10.2196/44358
98. Roose, K. (2024). Can A.I. Be Blamed for a Teen's Suicide? *The New York Times*. https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html
99. Rousmaniere, T., Zhang, Y., Li, X., & Shah, S. (2025). Large language models as mental health resources: Patterns of use in the United States. *Practice Innovations*. https://doi.org/10.1037/PRI0000292
100. Saab, K., Freyberg, J., Park, C., Strother, T., Cheng, Y., Weng, W.-H., Barrett, D. G. T., Stutz, D., Tomasev, N., Palepu, A., Liévin, V., Sharma, Y., Ruparel, R., Ahmed, A., Vedadi, E., Kanada, K., Hughes, C., Liu, Y., Brown, G., … Tanno, R. (2025). *Advancing Conversational Diagnostic AI with Multimodal Reasoning*. http://arxiv.org/abs/2505.04653
101. Schillings, C., Meißner, E., Erb, B., Bendig, E., Schultchen, D., & Pollatos, O. (2024). Effects of a Chatbot-Based Intervention on Stress and Health-Related Parameters in a Stressed Sample: Randomized Controlled Trial. *JMIR Mental Health*, *11*, e50454. https://doi.org/10.2196/50454
102. Selmi, P. M., Klein, M. H., Greist, J. H., Sorrell, S. P., & Erdman, H. P. (1990). Computer-administered cognitive-behavioral therapy for depression. *American Journal of Psychiatry*, *147*(1), 51–56. https://doi.org/10.1176/AJP.147.1.51,
103. Sertkaya, A., Devries, R., Jessup, A., & Beleche, T. (2022). Estimated Cost of Developing a Therapeutic Complex Medical Device in the US. *JAMA Network Open*, *5*, E2231609. https://doi.org/10.1001/jamanetworkopen.2022.31609
104. Shaengchart, Y., & Kraiwanit, T. (2023). Starlink satellite project impact on the Internet provider service in emerging economies. *Research in Globalization*, *6*, 100132. https://doi.org/10.1016/j.resglo.2023.100132
105. Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2023). Towards Understanding Sycophancy in Language Models. *12th International Conference on Learning Representations, ICLR 2024*. https://arxiv.org/pdf/2310.13548
106. Slack, W. V. (2000). Patient-Computer Dialogue: A Review,. *Yearbook of Medical Informatics*, *09*(01), 71–78. https://doi.org/10.1055/S-0038-1637944
107. Song, T., Jamieson, J., Zhu, T., Yamashita, N., & Lee, Y.-C. (2025). From Interaction to Attitude: Exploring the Impact of Human-AI Cooperation on Mental Illness Stigma. *Proceedings of the ACM on Human Computer Interaction*, *9*(2). https://doi.org/10.1145/3710987
108. Spytska, L. (2025). The use of artificial intelligence in psychotherapy: development of intelligent therapeutic systems. *BMC Psychology*, *13*(1), 175. https://doi.org/10.1186/s40359-025-02491-9
109. Stein, O. A., & Prost, A. (2024). Exploring the societal implications of digital mental health technologies: A critical review. *Ssm Mental Health*, *6*. https://doi.org/10.1016/j.ssmmh.2024.100373
110. Suharwardy, S., Ramachandran, M., Leonard, S. A., Gunaseelan, A., Lyell, D. J., Darcy, A., Robinson, A., & Judy, A. (2023). Feasibility and impact of a mental health chatbot on postpartum mental health: a randomized controlled trial. *Ajog Global Reports*, *3*(3). https://doi.org/10.1016/j.xagr.2023.100165
111. Sun, Y., Sheng, D., Zhou, Z., & Wu, Y. (2024). AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanities and Social Sciences Communications*, *11*(1), 1–14. https://doi.org/10.1057/S41599-024-03811-X;SUBJMETA=4001,4014,4045;KWRD=BUSINESS+AND+MANAGEMENT,SCIENCE
112. Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. *Advances in Neural Information Processing Systems*, *4*(January), 3104–3112. https://arxiv.org/pdf/1409.3215

113. Sweeney, C., Potts, C., Ennis, E., Bond, R., Mulvenna, M. D., O'Neill, S., Malcolm, M., Kuosmanen, L., Kostenius, C., Vakaloudis, A., McConvey, G., Turkington, R., Hanna, D., Nieminen, H., Vartiainen, A. K., Robertson, A., & McTear, M. F. (2021). Can Chatbots Help Support a Person's Mental Health? Perceptions and Views from Mental Healthcare Professionals and Experts. *ACM Transactions on Computing for Healthcare*, *2*. https://doi.org/10.1145/3453175
114. The Lancet Psychiatry. (2024). Global Burden of Disease 2021: mental health messages. *The Lancet Psychiatry*, *11*(8), 573. https://doi.org/10.1016/S2215-0366(24)00222-0
115. Tian, W., Lu, Y., Yu, J., Fan, J., Tang, P., & Zhang, L. (2022). A Privacy-Preserving Framework for Mental Health Chatbots Based on Confidential Computing. *Proceedings 2022 IEEE Smartworld Ubiquitous Intelligence and Computing Autonomous and Trusted Vehicles Scalable Computing and Communications Digital Twin Privacy Computing Metaverse Smartworld Uic Atc Scalcom Digitaltwin Pricomp Metaverse 2022*, 1119–1124. https://doi.org/10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00160
116. Toch, E., Wang, Y., & Cranor, L. F. (2012). Personalization and privacy: A survey of privacy risks and remedies in personalization-based systems. *User Modeling and User-Adapted Interaction*, *22*, 203–220. https://doi.org/10.1007/s11257-011-9110-z
117. Tong, A. C. Y., Wong, K. T. Y., Chung, W. W. T., & Mak, W. W. S. (2025). Effectiveness of Topic-Based Chatbots on Mental Health Self-Care and Mental Well-Being: Randomized Controlled Trial. *Journal of Medical Internet Research*, *27*(1). https://doi.org/10.2196/70436
118. Tu, T., Palepu, A., Schaekermann, M., Saab, K., Freyberg, J., Tanno, R., Wang, A., Li, B., Amin, M., Tomasev, N., Azizi, S., Singhal, K., Cheng, Y., Hou, L., Webson, A., Kulkarni, K., Mahdavi, S. S., Semturs, C., Gottweis, J., … Natarajan, V. (2024). *Towards Conversational Diagnostic AI*. https://arxiv.org/pdf/2401.05654
119. Vaidyam, A. N., Linggonegoro, D., & Torous, J. (2021). Changes to the Psychiatric Chatbot Landscape: A Systematic Review of Conversational Agents in Serious Mental Illness: Changements du paysage psychiatrique des chatbots: une revue systématique des agents conversationnels dans la maladie mentale sérieuse. In *Canadian Journal of Psychiatry* (Vol. 66, Issue 4, pp. 339–348). SAGE Publications Inc. https://doi.org/10.1177/0706743720966429
120. Van Der Schyff, E., Ridout, B., Amon, K., Forsyth, R., & Campbell, A. (2023). Providing Self-Led Mental Health Support Through an Artificial Intelligence–Powered Chat Bot (Leora) to Meet the Demand of Mental Health Care. *Journal of Medical Internet Research*, *25*. https://doi.org/10.2196/46448
121. Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*. 1. https://arxiv.org/pdf/1706.03762
122. Wainberg, M. L., Scorza, P., Shultz, J. M., Helpman, L., Mootz, J. J., Johnson, K. A., Neria, Y., Bradford, J. M. E., Oquendo, M. A., & Arbuckle, M. R. (2017). Challenges and Opportunities in Global Mental Health: a Research-to-Practice Perspective. In *Current Psychiatry Reports* (Vol. 19). Current Medicine Group LLC 1. https://doi.org/10.1007/s11920-017-0780-z
123. Wang, X., & Li, Q. (2024). Co-designing Human–Chatbot Interaction for Various Healthcare Purposes: Considering Chatbots' Social Characteristics and Communication Modalities. In *Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics: Vol. 14726 LNCS*. https://doi.org/10.1007/978-3-031-61546-7_26
124. Weizenbaum, J. (1966). ELIZA-A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168
125. World Health Organization. (2023). Mental Health Gap Action Programme (mhGAP) guideline for mental, neurological and substance use disorders. In *https://www.who.int/publications/i/item/9789240084278: Vol. 3rd Edition*.
126. Xia, Y., Kim, J., Chen, Y., Ye, H., Kundu, S., Hao, C. C., & Talati, N. (2024). Understanding the Performance and Estimating the Cost of LLM Fine-Tuning. *Proceedings - 2024 IEEE International Symposium on Workload Characterization, IISWC 2024*, 210–223. https://doi.org/10.1109/IISWC63097.2024.00027
127. Xiang, C. (2023). Man Dies by Suicide After Talking With AI Chatbot, Widow Says. *Vice*. https://www.vice.com/en/article/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says/
128. Xie, T., & Pentina, I. (2022). Attachment Theory as a Framework to Understand Relationships with Social Chatbots: A Case Study of Replika. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2022-Janua*, 2046–2055. https://doi.org/10.24251/hicss.2022.258
129. Xie, Z., & Wang, Z. (2024). Longitudinal Examination of the Relationship Between Virtual Companionship and Social Anxiety: Emotional Expression as a Mediator and Mindfulness as a Moderator. *Psychology Research and Behavior Management*, *17*, 765–782. https://doi.org/10.2147/PRBM.S447487
130. Yang, Y., Tavares, J., & Oliveira, T. (2024). A New Research Model for Artificial Intelligence–Based Well-Being Chatbot Engagement: Survey Study. *Jmir Human Factors*, *11*. https://doi.org/10.2196/59908
131. Zagorski, N. (2022). Popularity of Mental Health Chatbots Grows. *Https://Doi.Org/10.1176/Appi.Pn.2022.05.4.50*, *57*(5). https://doi.org/10.1176/APPI.PN.2022.05.4.50

132. 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
133. Zhu, Y., Liang, J., & Zhao, Y. (2025). Expert or partner: The matching effect of AI chatbot roles in different service contexts. *Electronic Commerce Research and Applications*, *71*. https://doi.org/10.1016/j.elerap.2025.101496
134. Zhu, Y., Wang, R., & Pu, C. (2022). "I am chatbot, your virtual mental health adviser." What drives citizens' satisfaction and continuance intention toward mental health chatbots during the COVID-19 pandemic? An empirical study in China. *Digital Health*, *8*. https://doi.org/10.1177/20552076221090031