



Talking Mental Health: a Battle of Wits Between Humans and AI

Esha Lopes¹ · Gagan Jain¹ · Per Carlbring² · Samridhi Pareek¹

Received: 28 August 2023 / Revised: 12 October 2023 / Accepted: 18 October 2023 / Published online: 8 November 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract

This study aimed to investigate the differences between responses generated by ChatGPT and those produced by humans in terms of authenticity, professionalism, and practicality. It involved 140 participants of the age group ranging from 18 to 43 (101 females, 37 males, and 2 preferring not to disclose their gender). The participants were presented with the 10 solution statements against the 10 problem statements, generated by human participants and ChatGPT3.5 (gpt-3.5-turbo) and asked to rate the responses on a 5-point Likert scale, with higher scores indicating higher levels of authenticity, professionalism, and practicality. Paired sample *t*-test was conducted to compare the scores of the ChatGPT-generated responses and the human-generated responses. The results of the study indicated that there was significant difference between the two types of responses in the given dimensions. These findings suggest that ChatGPT-generated responses can be considered a reliable alternative to human-generated responses in certain applications. Additionally, the use of ChatGPT-generated responses can reduce the response time and workload of human responders, as well as the associated costs.

Keywords Chatbot · Artificial intelligence · Natural language processing · ChatGPT · Authenticity · Professionalism · Practicality · Human-generated responses · Mental health

Introduction

The rapid advancement of technology worldwide prompts a question that cannot be ignored: Can robots ever replace humans? Artificial intelligence has made notable progress in many facets of human existence, such as the development of self-driving cars and the ability to predict loan risks. It has emerged as a crucial tool in various domains, including healthcare, education, finance, and entertainment (Huang & Rust, 2018; Thompson & Bentzien, 2020). With the emergence of artificial intelligence, the potential for innovative approaches to addressing mental health issues has become apparent (Brown et al., 2020; Bohr & Memarzadeh, 2020; Harrington et al., 2021). Mental health is a multifaceted and intricate concern that demands a holistic and integrated approach (Ee et al., 2020; Smith et al., 2023). As per the World Health Organization (WHO), mental health disorders

are among the foremost contributors to global disability, with depression and anxiety ranking as the most widespread conditions (WHO, 2021). Despite the growing recognition of mental health as a public health priority, there are still significant challenges associated with addressing it. Many regions across the globe, especially low- and middle-income countries, are grappling with a notable scarcity of mental health practitioners (Pathare et al., 2018; Patel et al., 2018). To address this gap, researchers have proposed the use of artificial intelligence (AI) as a potential tool to support mental healthcare (Singh, 2023; Carlbring et al., 2023). Among the notable advancements in AI, natural language processing (NLP) technologies, such as chatbots and language models, have made significant strides in recent years (García-Peñalvo et al., 2023; Luxton et al., 2011; Tielman et al., 2018; Davenport et al., 2020). In the past couple of years, there has been greater interest in ChatGPT, an OpenAI language model which utilizes deep learning algorithms to construct text-based dialogues in response to human prompts (Radford et al., 2019).

ChatGPT has been used in various domains, including mental health, where it has been proposed as a tool to support mental healthcare by providing users with a virtual assistant that can respond to their mental health concerns in real time. In the past few years, there has been a growth of interest in harnessing the potential of AI-powered chatbots

✉ Gagan Jain
gaganjain1030@gmail.com

¹ Department of Psychology, Manipal University Jaipur, Jaipur, India

² Department of Psychology, Stockholm University, Stockholm, Sweden

as a tool to address mental health issues. These advances in technology have the potential to transform how we communicate with computers and with one another (Wu et al., 2020). But they also pose important questions related to their authenticity, professionalism, and practicality. In this exploratory research study, we explored the use of ChatGPT3.5 (gpt-3.5-turbo), one of the most advanced NLP technologies available today in dealing with some mental issues young people face in their day-to-day life, and compare its responses with the human responses in terms of authenticity, professionalism, and practicality.

The Collaboration of ChatGPT and Mental Health

The debate around machines' ability to track, analyze, and imitate human emotions has been ongoing in the AI community. Koko, a nonprofit organization and behavioral health platform, recently conducted an experiment that utilized ChatGPT to provide mental health support to individuals who were going through emotional distress. In a viral tweet, Koko's founder, Rob Morris, shared that ChatGPT assisted with over 30,000 requests for mental health support from more than 4000 users. Surprisingly, the messages generated by ChatGPT received higher ratings compared to those written by human peers, and the response times decreased significantly by 50%, taking less than a minute. However, some users found the AI-generated responses to be inauthentic, as they lacked the lived experiences of humans, resulting in ethical concerns about informed consent and the potential exploitation of sensitive information from a vulnerable population. During the experiment, users who were providing support had the option to use the ChatGPT3.5 feature to compose their responses. These responses were then filtered and supervised by humans in a "co-pilot" approach. Although the ethical implications of the experiment are still unresolved, it provides valuable insights into the potential of machines in mimicking human emotions.

Another AI-powered mental health bot, Woebot, has also published data on the effectiveness of support provided by robots after establishing a bond with users. However, concerns exist regarding over-reliance on machines for emotional support and the potential impact on human relationships. The experiment conducted by Koko highlights the importance of continued research in determining the right balance between empathetic machine and human connections. It raises important ethical considerations and underscores the need for careful exploration of the potential of machines in providing emotional support, while also acknowledging the value of human connections in the field of mental health (Agarwal & Sharma, 2023).

The present study thus aimed to compare the responses generated by ChatGPT to those of humans. By doing so, we hope to shed light on the potential of AI, specifically ChatGPT, as a tool for addressing mental health problems and understanding the perceptions and attitudes of individuals towards AI-generated responses in this context. We hope to gain a better understanding and provide insights into how we can best leverage the strengths of AI as a problem-solving tool while also addressing its limitations and potential risks. We aim to shed light on some of the challenges and opportunities presented by the growing use of AI in society further identifying the opportunities for humans and AI to collaborate in order to create solutions that are both authentic and feasible.

Role of AI-Based Chatbots in Mental Health

Mental health problems have been recognized as a major public health issue, impacting over 50% of the population in middle- and high-income countries with profound consequences for society. They cause significant economic and social burdens underscoring the importance of comprehensive strategies to foster mental well-being and provide accessible interventions (He et al., 2022; Trautmann et al., 2016). Despite the growing demand for mental health services, there are still many challenges in providing accessible, affordable, and effective care to those in need. Due to limited healthcare resources, unguided interventions are often implemented over more effective guided interventions, primarily due to their lower initial costs (Baumeister et al., 2019). Technology, including artificial intelligence (AI), has the potential to address some of the gaps in mental health services (Denecke et al., 2021). AI can support mental health services in various ways, such as by providing personalized, accessible, and scalable services. Given the challenges in meeting the current and future demand for mental healthcare, there is a growing suggestion to utilize technology, including artificial intelligence (AI) and conversational agents like chatbots and virtual assistants, as potential solutions to address these gaps and improve the efficiency and effectiveness of care delivery (Vaidyam et al., 2019; Abd-Alrazaq et al., 2020; Bucci et al., 2019).

One specific area of AI that has received increasing attention in mental health is conversational agents, also known as chatbots. Conversational agents replicate human-like conversations using natural language processing (D'Alfonso, 2020; Vaidyam et al., 2019; Diederich et al., 2022). Chatbots have gained popularity in mental health therapy in recent years, with positive outcomes documented in several research. Chatbots can help with mental health in several ways, including psychoeducation, screening for mental health issues, and giving cognitive behavioral therapy (CBT) (Ly et al., 2020; Baumeister et al., 2019; Maciejewski et al., 2023).

AI's potential impact on mental healthcare provision may differ depending on a country's resources and cultural attitudes towards mental illness. In countries with limited resources or high stigma, AI can help fill the gap by providing accessible and discreet mental health support. In more developed countries, AI can complement existing services but may not replace human expertise entirely. The successful implementation of AI in mental healthcare requires careful consideration of each country's specific context and needs (Reddy et al., 2021; Minerva & Giubilini, 2023; Adamopoulou & Moussiades, 2020).

Effectiveness of AI-Based Interventions

Several studies have examined the effectiveness of conversational agents for mental health and have compared their performance to that of human providers or gold standards. For example, a randomized controlled trial by Fitzpatrick and colleagues (2017) found that a chatbot-based CBT program was effective in reducing symptoms of depression and anxiety among young adults compared to a waitlist control group. In another study, Fitzpatrick et al. (2017) found that Woebot (a fully automated chatbot)-based intervention considerably decreased levels of depression, anxiety, and stress among the treatment groups in comparison with the control group. Additionally, participants reported a high level of engagement and satisfaction with Woebot. Chatbots have been employed in the field of student counselling. Kavakli et al. (2012) investigated the use of a chatbot to assist students in coping with exam stress, as university counselling services can often experience a surge in demand during exam periods. Utilizing chatbots for mental health counselling offers numerous advantages for users, including instant access to information and a more interactive experience when mental health screening tools are integrated into the chatbot interface. However, ethical considerations need to be carefully addressed in the implementation of chatbot-based mental health counselling (Cameron et al., 2017).

The evidence for the effectiveness of conversational agents in addressing mental health problems is still mixed and inconclusive, with some studies reporting positive outcomes and others showing no or negative effects. A systematic review of 13 studies on the use of chatbots found that the interventions were generally acceptable and feasible, but there was little evidence for their clinical efficacy or superiority over standard care. The use of conversational agents for mental health treatment is showing rapid growth, with five studies published in 2018 alone (Freeman et al., 2018; Bird et al., 2018; Fulmer et al., 2018; Inkster et al., 2018; Suganuma et al., 2018). The included interventions were typically brief, allowed participants to control the intensity of the intervention, and drew from various psychological approaches. All the studies reported reduced psychological

distress after using a conversational agent, with some controlled studies showing significant reductions compared to information or no treatment controls, indicating potential effectiveness for treating mild-to-moderate psychological distress in adults (Freeman et al., 2018; Fitzpatrick et al., 2017; Fulmer et al., 2018). However, their broader utility in promoting positive well-being in nonclinical populations remains uncertain. Controlled studies comparing conversational agents with other interventions (e.g., another conversational agent or human psychological therapy) did not consistently show superior effects (Bird et al., 2018; Gaffney et al., 2013), though it is essential to note that these studies assessed relative rather than absolute treatment efficacy, and more research is needed in this area (Gaffney et al., 2019).

A systematic review conducted by Rapp et al. (2021) primarily explores the human–computer interaction (HCI) aspect of chatbot usage. The review investigates the usability and user acceptance of chatbots designed to exhibit human-like characteristics and empathy in their interactions with users. Users can become frustrated when chatbot responses are irrelevant (Wang & Nakatsu, 2013), but chatbots that express emotions can improve the user's mood (Ghandeharioun et al., 2019). Additionally, users tend to anthropomorphize chatbots, affecting their interaction and behavior (Qin et al., 2023; Paul et al., 2021). Chatbots that mimic human behavior and emotions can establish better rapport, motivation, and engagement (Giger et al., 2019). As a result, researchers are actively exploring methods to enhance chatbot's empathetic and emotional capabilities (Rashkin et al., 2018; Asada, 2015). Currently, the primary focus lies in the development of conversational agents with the ability to discern users' emotions and provide appropriate emotional cues in response, thereby aiming to enhance user engagement. This growing emphasis on emotionally intelligent chatbots underscores the significance of creating such bots that can generate emotional responses to user requests, despite the inherent challenges, in order to ensure their wide acceptance and successful adoption (Bilquise et al., 2022).

Methodology

Concerns have been raised about the level of authenticity (degree to which a response is perceived to be genuine and reflective of the emotional and psychological states of the individual seeking support), professionalism (the degree to which a response is perceived to be appropriate, accurate, and consistent with established ethical and professional standards for mental health and psychological support services), and practicality (the degree to which a response is perceived to be helpful and relevant to the individual's specific needs and concerns) in the responses generated by artificial intelligence (chatbots). The study aimed to investigate how participants rate the responses generated by AI on

these three dimensions of authenticity, professionalism, and practicability as compared to those generated by humans.

Sample

A combination of convenience and purposive sampling has been used in the current study. The study included participant ranging between 18 and 43 ($SD = 3.444$) years of age who could evaluate mental health solutions, were fluent in English for proper comprehension of the solutions, and had no prior experience with therapy or AI-based solutions. This allowed for diverse and broad spectrums of viewpoints. It also took into consideration educational diversity, to eliminate the bias of recruiting participants from just psychology or AI-based background like machine learning. A total of 140 participants (101 females, 37 males, and 2 opted not to provide information about their gender) took part and completed the study. Participants had various educational backgrounds, including undergraduate and graduate degrees in psychology, as well as degrees in architecture, business, law, education, microbiology, and computer science. By excluding participants with language problems, the study reduced the chance of misinterpretation of the responses. Participants who refused to participate or did not offer proper consent were excluded from the study. This was done to maintain ethical standards. The research also sought to identify and eliminate non-serious or insincere submissions. Convenience sampling was employed through the distribution of online forms and in-person visits to classrooms, which allowed the recruitment of potential participants who were available and willing to participate. Purposive sampling was also employed as a deliberate effort was made to collect data from individuals with different educational and occupational backgrounds.

Data Collection

The study was carried out in two phases: phase one included the collection of problem situations and solutions for the next phase of the study with an initial activity called the “listening circle.” It involved a total of 50 individuals, not involved in the study. They were invited to share questions/problem situations that currently caused distress in their lives. Participants filled out a paper–pencil survey. Provided with a sheet of paper, they detailed the problems they were currently facing in life. Upon collection, the papers were shuffled and redistributed randomly among the participants. This ensured that each participant received a problem statement different from their own to brainstorm solutions for. The study then included 10 open-ended questions selected from a data pool of 50 questions generated by these participants. These questions

described situations related to interpersonal problems, stress, anxiety, intrapersonal conflicts, etc., where they may seek mental health support. Phase two of the study included the presentation of the responses from two different sources. For each question, all 140 participants were provided with two types of responses: ChatGPT (version 3.5) and human-generated response. The following prompt was provided to ChatGPT for the solutions: “I am currently facing some challenges in my life and could really use some advice and perspective on how to handle it. Can you assist me in navigating through this situation? Please summarize the solutions without presenting them in a list format and respond in a supportive and conversational manner, similar to a human therapist.” It was a single-blind study where participants were unaware of the origin of AI-based responses. The responses were presented to the participants in a random order, ensuring that any biases that could result from the order of presentation were avoided. This was accomplished by utilizing the inherent functionality of Google Forms, enabling the permutation of question sequencing. Likewise, the responses were methodically arranged in a randomized manner to preclude the emergence of any discernible patterns. This methodology helps maintain the reliability of the results, making sure they are not impacted by the way the responses were shown. The responses provided by ChatGPT and human respondents for each question were observed to possess a relatively consistent length, differing by approximately 5 to 6 words on average for each question. No prompts were provided for the length of the solutions to ensure no limitations on the nature of solutions provided. However, it is important to note that they were not precisely uniform in their extent. Participants were asked to rate each response on a 5-point Likert scale for authenticity, professionalism, and practicality. These scenarios were designed to be realistic and engaging. Data were collected through a self-report survey form. Data collected from participants were stored in a secure database with restricted access. The data collected included the participant’s demographic details, the question number, and the ratings for authenticity, professionalism, and practicality.

Some of the sample problem statements and the responses received from human respondents and AI were as follows:

Sample 1

Lately, I’ve been feeling lonely. I’ve classmates, they are good to me, and I talk to them during college hours. But, I don’t have someone to eat pizza with or do just a little chit chat. I cannot focus on my academics as well because I overthink this a lot.

Human Response

Feeling lonely is okay and normal. Not every time you could be surrounded by ppl, start enjoying your own company. Be your own best friend. By doing so, you will build yourself confidence and people will get attracted to that. Stop chasing friends you might get some toxic ones which could lead to many bad things just start loving yourself and people will come to you. You deserve to be loved so you will get it, don't worry. Self-love is everything and it will get you everywhere. If you feel bad about not having friends, go and make some, try to find extrovert people cause it's easier to develop a bond with them, and they will push you into a big group.

AI Response

It seems like you're feeling lonely and that's affecting your studies. While you have friendly classmates to talk to during college, you still crave deeper connections outside of class. You could join a club or organization that aligns with your interests, volunteer in your community, or attend social events. This can be a great way to meet new people and potentially find someone you connect with on a deeper level. Alternatively, seeking help from a therapist or counselor could also provide you with strategies to cope with loneliness and help you develop new social skills and confidence. They can also help you work through any underlying issues that may be contributing to your feelings of loneliness. Remember, it's normal to feel lonely at times, but taking care of your mental health is important.

Sample 2

I just want to kick out old memories related to people who led me to trauma. Every day the incidents still haunt me. Tried everything, burning papers etc. but still the thoughts enter my mind how do I delete them forever?

Human Response

I feel you can't remove those memories forever; they will fade with time. It will take some time, but it will remain maybe 1% because it's your past. However, it's going to fade because new memories will take place and just try to invest yourself in something which keeps you busy, where you feel alive or happy. I feel you should share it with someone who understands you, if you keep it inside and won't talk about it you won't stop overthinking about it. Whatever happened was a

chapter of your life, learn from it because this is going to help you to understand yourself better only. That past is not going to decide your future you can make it better. I don't know you, but I feel you are a strong person, you have come so far that means you are a strong person, and you will get over it.

AI Response

It can be difficult to let go of traumatic memories from the past, but there are techniques to help you process them in a healthier way. Mindfulness meditation can help you develop a more accepting and compassionate attitude toward yourself and your experiences. Talking to a therapist or counselor can also provide a safe and supportive environment to process your trauma and learn strategies to manage intrusive thoughts and memories. It's important to take care of yourself physically and emotionally by getting enough sleep, eating a balanced diet, exercising regularly, and doing activities you enjoy. Remember that healing is a process, so be patient and kind to yourself. With time and the right support, you can learn to manage intrusive thoughts and memories and live a fulfilling life.

Ethical Considerations

Prior to their participation in the study, informed consent was obtained from each participant. Participants were given an information sheet as part of the survey form, which outlined the purpose and procedures of the study. They were explicitly informed that their participation was voluntary, and they had the right to withdraw from the study at any time without facing any penalty. All data collected was kept confidential and anonymous. Deception was used since the sources of responses were kept anonymous to fulfil the research objective.

Statistical Analysis

The data collected from 140 participants were analyzed using SPSS version 26. The data underwent an examination for outliers and normality, and it was determined that there were no outliers present in the data, and the assumptions of normality were met. Additionally, there were no missing data as all questionnaires were completed in full, which means that no participants were removed from the analysis. Descriptive statistics were conducted to provide a quantitative summary of the data and to identify any patterns or trends. The mean ratings of authenticity, professionalism, and practicality were calculated separately for the ChatGPT and human-generated responses for all the 10 questions. Mean ratings for overall 10 questions for all three matrices of both the response types were also calculated. Further inferential statistics were applied, and the mean ratings for

authenticity, professionalism, and practicality of ChatGPT-generated responses and human-generated responses were compared using a paired sample *t*-test.

Result

Table 1 shows that the ChatGPT responses demonstrated higher means for authenticity (36.33), professionalism (36.18), and practicality (36.73) as compared to the means of human responses in authenticity (33.23), professionalism (32.26), and practicality (34.19). The results were significantly higher in all the 3 areas with authenticity ($t=5.33$), professionalism ($t=6.86$), and practicality ($t=4.34$) significant at 0.001 level. Overall, these results suggest that participants had a positive experience with ChatGPT and perceived it as authentic, professional, and practical. However, it is important to note that these ratings are based on the participants' subjective perceptions and may not reflect objective measures of effectiveness or competency in delivering therapy. In the context of a paired sample *t*-test, the provided medium effect size values offer insights into the magnitude of differences observed between two sets of responses, specifically evaluating authenticity ($d=0.40$), professionalism ($d=0.51$), and practicality (0.31) of AI-generated and human-generated responses.

Discussion

With the ongoing research on the effectiveness of chatbots, it is important to note that their success may be influenced by how users perceive their authenticity and usability. As Table 1 suggests, the mean ratings for authenticity, professionalism, and practicality were higher for ChatGPT as compared to the human responses. The result suggests a strong influence of response source on these aspects, with AI responses consistently exhibiting higher levels of these attributes compared to human responses. A *t*-test was conducted to see if the mean differences were statistically significant, which further suggested a significant difference in terms of authenticity, professionalism, and practicality metrics at 0.001 level of significance (Table 1).

In the realm of authenticity, the significant difference indicates that participants perceived the conversations with ChatGPT as more genuine and sincere compared to the

conversations with human therapists. The table indicates that there were also statistically significant differences for professionalism, suggesting that participants perceived ChatGPT as more competent and skilled in delivering therapy compared to human therapists. Finally, the *t* scores for practicality were also significantly higher for ChatGPT responses compared to human responses, indicating that participants found ChatGPT methods of therapy to be convenient and useful in their daily lives. This suggests that participants perceived AI-generated responses as significantly more authentic, professional, and practical as compared to human responses.

Authenticity is very much humane when it comes to its characteristics. In the realm of mental health support, user perceptions regarding the authenticity of AI-generated responses, compared to human-generated ones, reveal intriguing patterns. However, this study indicates that when the response source is masked, the users in fact have a stronger inclination towards the AI responses as compared to the human ones. If the AI-generated responses are able to provide relevant and helpful information, convey empathy, and engage in a meaningful conversation, users may perceive them as authentic and embrace the genuineness of the AI-generated response. Concurrently, as like with authenticity realm, users perceived the professionalism of AI-generated responses better as compared to human-generated ones. They felt more comfortable and confident in the expertise of AI-generated responses which could provide personalized and tailored responses based on their individual needs. AI's accessibility, impartiality, and consistency may have had an influence on this perception. In terms of practicality also, we notice a bend towards AI responses. ChatGPT, as an AI language model, can provide basic information and counseling on various topics, including mental health, offering general advice and suggestions based on its training data. People also find these responses practical and convenient. However, it is crucial to understand that these perceptions could be influenced by various factors, such as the specific context of the problem, the characteristics of the AI system and human responders, and the individual preferences.

Nonetheless, research has shown that individuals' perceptions of AI-generated responses can be influenced by their knowledge or awareness of the source of the responses (Hsu, 2023; Ingram, 2023). The "AI effect" suggests that individuals may rate AI-generated responses lower in quality or effectiveness simply because they are aware that they

Table 1 ChatGPT and human responses for authenticity, professionalism, and practicality

	ChatGPT			Human			<i>t</i>	<i>p</i>	Cohen's <i>d</i>
	M	SD	SEM	M	SD	SEM			
Authenticity	36.33	7.43	0.628	33.27	7.58	0.639	5.33	0.000	0.40
Professionalism	36.18	7.70	0.651	32.26	7.67	0.648	6.86	0.000	0.51
Practicality	36.73	7.70	0.651	34.19	8.20	0.694	4.34	0.000	0.31

are interacting with an AI system, even if the responses are objectively of high quality (Santoro & Monin, 2023). It shall however be noted that without knowledge of the source, the scoring of responses that were received relied solely on the quality, the effectiveness, the conversational flow, the relevance of the content provided, and the overall user experience, rather than attributing their perceptions to the origin of the response. This masking was therefore necessary to reduce the “AI effect,” and thus, we can safely appreciate the differences in ratings between AI-generated and human-generated responses.

It is worth noting that even when individuals are unaware of the source, they may try to identify which source is the response coming from, using subtle cues or differences in the language, style, or tone of the responses that could potentially reveal whether the response was generated by an AI or a human. For example, individuals may believe that AI-generated responses may exhibit patterns or language indicative of automated content, while human-generated responses may have nuances or personal touches that reflect human communication styles. It can be fairly accepted that despite this guessing, if present, the AI responses outperformed the human response at least in the domains measured. AI-generated responses thus exhibit the potential to rival human-generated response. This could be suggestive of AI-generated responses having the capability to be better than human-generated responses in certain contexts, at least in terms of these metrics.

It shall however be noted that AI-generated responses in mental health support are perceived differently based on user factors and the AI system’s capabilities. Understanding these perceptions requires more research and user feedback to ensure AI’s acceptance and effectiveness.

How Can We Use ChatGPT Effectively in Providing Mental Health Support to Individuals in Emotional Distress?

ChatGPT and other AI conversational agents offer effective mental health support, including psychoeducation, self-help interventions, and mobile mental health apps. In addition to this, it can provide potential assistance in therapy sessions, emotional support through verbal channels, and hand out referrals to professional services. Their 24/7 availability enhances accessibility, especially for those facing barriers like cost and time constraints through services like online mental health platforms and crisis hotlines. Integration with various technologies allows ChatGPT to provide mental health services across different settings (Aminah et al., 2023).

ChatGPT can offer valuable insight into mental health conditions, providing information on symptoms, causes, risk factors, and treatment options like therapy, medication,

lifestyle changes, and self-care strategies. This psychoeducation can empower individuals to understand their mental health better, leading to reduced anxiety, improved self-esteem, better treatment adherence, and overall quality of life (Christensen et al., 2009). Additionally, it can also aid in recognizing warning signs of relapse and taking proactive steps for effective condition management. Integrating ChatGPT into online mental health platforms offers accessible on-demand support. Through text-based interactions, users can receive psychoeducation, personalized self-care recommendations, and a safe space for sharing concerns. This benefits those hesitant to seek in-person therapy due to stigma or access issues (Lindqvist et al., 2022; Andersson & Titov, 2014; Koelen et al., 2022; Hedman-Lagerlöf et al., 2023). Incorporating ChatGPT into mobile mental health apps provides convenient support for those with busy lifestyles or limited access to mental health services in rural areas (Ernsting et al., 2020; Smith et al., 2023). These apps provide a safe and private space for discussing mental health concerns without fear of judgment, enhancing assistance with personalized, effective guidance using machine learning algorithms (Suta et al., 2020). Crisis hotlines can leverage ChatGPT to deliver immediate assistance to individuals in distress, offering on-demand support without waiting for a human operator. Using natural language processing, it can detect signs of distress and provide appropriate crisis intervention and referrals to emergency services or mental health providers (Walsh et al., 2017). This integration not only saves lives but also eases the burden on human operators, allowing them to focus on more complex cases. Utilizing ChatGPT in therapy sessions, mental health professionals can offer personalized support by providing psychoeducation, coping strategies, and exploring thoughts in a non-judgmental setting (Erbe et al., 2017; Balaji et al., 2022). Mental health professionals can assign interactive exercises or journal prompts generated by ChatGPT for clients to complete outside of sessions. Further, it can assist in tracking clients’ progress by recording session summaries, mood fluctuations, and insights. This data can aid therapists in making informed decisions about treatment plans. Clients can role-play or practice communication skills, conflict resolution, or assertiveness techniques with ChatGPT, which can then be transferred to real-life situations. In the realm of self-help interventions, ChatGPT serves as a valuable resource by offering personalized strategies like relaxation techniques, mindfulness exercises, and CBT methods to manage specific mental health concerns. Tailored recommendations based on NLP enhance the effectiveness of support provided. Although it is not a substitute for human therapists, ChatGPT can provide users with emotional support, offering a non-judgmental and empathetic ear to listen and help users process emotion (Seifert et al., 2023). While creating a confidential space for discussing mental health

concerns, it can utilize natural language processing to detect tone and emotions and the user's language and respond in a way that is empathetic and supportive thus validating individuals' experiences (Sabour et al., 2023; Bubeck et al., 2023; Sharma et al., 2023). ChatGPT can refer users to professional mental health services if their concerns surpass the AI's capabilities. By directing them to local mental health professionals or online resources, it can ensure that individuals receive specialized care. This can include providing the user with a list of mental health professionals in their area or directing them to online resources for finding mental health services (Vis et al., 2023). While ChatGPT is valuable in mental health support, it should not replace professional care and is most effective when used in conjunction with other mental health services to provide comprehensive support.

Potential Limitations of Incorporating ChatGPT into Existing Mental Healthcare Models

Incorporating ChatGPT into existing mental healthcare models has promising benefits, but it also comes with potential limitations. While the results show that ChatGPT-generated responses can be perceived positively, they may lack the personalized and empathetic support that human therapists can offer, especially for individuals with severe or complex mental health issues. AI-powered chatbots are limited by their reliance on data and algorithms, which may not fully comprehend the depth of human emotions and experiences. Moreover, they may not have the capacity to handle crisis situations or provide tailored interventions for specific individuals. Therefore, while ChatGPT can supplement mental healthcare, it should not replace the essential role of trained mental health professionals in providing comprehensive and individualized support to those in need (Morris et al., 2018).

Ethical Issues

As the implementation of chatbots and AI in psychiatrist and counselling services holds promise, it also raises noteworthy ethical concerns that require careful consideration (Rana, 2023). One significant issue is the risk of bias in AI and conversational agents, which could perpetuate stereotypes and discrimination against certain populations (Kaddari et al., 2020; Laranjo et al., 2018). Conversational bias in chatbots, stemming from flawed training data, can lead to offensive content. Yet, bias can be mitigated through thorough analysis of input data and modifying training algorithms. Detecting and eliminating bias via data examination allows chatbots to provide balanced and unbiased responses (Bradley & Alhajjar, n.d.; Coghlan et al., 2023). Diversifying training data, involving stakeholders in the design process, and using explainable AI techniques can also help mitigate bias. The usability and

acceptability of conversational agents for different user groups is also a crucial issue to consider. Factors such as perceived usefulness, ease of use, trust, and empathy can influence users' attitudes and behaviors towards these tools. For example, a study by Fitzpatrick and colleagues (2017) found that users preferred conversational agents that used natural language, expressed empathy, and provided personalized feedback. User involvement in the design process and tailoring these tools to individual needs and preferences can enhance their effectiveness. Conversational agents raise ethical concerns regarding their effect on the therapeutic alliance and patient-provider relationship. Due to their non-human nature, patients' expectations and perceptions of interactions may differ, impacting engagement and treatment adherence (Miloff et al., 2020; Bond et al., 2019; Wampold & Flückiger, 2023). Furthermore, there may generate concerns related to informed consent (Hsu, 2023), privacy, trust (Carlbring et al., 2023), and confidentiality, as conversational agents may collect and store sensitive data about users. Established guidelines and frameworks, like the American Psychological Association's ethical principles and code of conduct for chatbots, can provide guidance in addressing these ethical issues (APA, 2020).

Limitations

The study has provided the basis to explore and construct theory on conversational bots and their role in mental health services, but it has its limitations. The study had a limited sample size and does not contain a broad sample of participants with respect to race, ethnic background, socioeconomic status, or other demographic characteristics, limiting the findings' generalizability to a wider and more diverse population. It is crucial to highlight that the findings of this research are based on subjective ratings provided by study participants, and the study did not evaluate the effectiveness or clinical outcomes of therapy offered by ChatGPT or human therapists. It is important to interpret the results with caution, as subjective ratings may be subject to bias and may not necessarily reflect actual clinical outcomes. Conducting further research utilizing rigorous methodologies and objective measures is needed to thoroughly evaluate the efficacy and clinical impact of therapy delivered by ChatGPT or other conversational agents in comparison to human therapists. Further, the problem situations presented in the study were generalized whereas real-life situations can be more complex. Thus, the limitations of chatbots regarding the complexity of real-life situations should be taken into account. The developer's bias is a concern as chatbots are developed by humans and work on an algorithm of available data. The current chatbot used may be surpassed by a more advanced version and may produce different responses that can change the results of the current study. The ethical issues in the current study are taken into consideration. The study used deception particularly because

the participants were not given the information about the source of the responses as a requirement of the study. Nonetheless, this approach may pose ethical concerns and potentially compromise the standards of research integrity.

Implications and Future Research

While there are concerns about bias, ethical implications, and usability, studies have shown that conversational agents can be effective, cost-saving, and acceptable for diverse populations (Kelly et al., 2023). To fully realize their potential, it is necessary to engage in interdisciplinary collaboration, involve stakeholders in the design and implementation process, and address the technical, ethical, and regulatory challenges that arise from their use. By doing so, we can develop innovative and scalable solutions aimed at enhancing the accessibility of mental healthcare services and fostering mental health and well-being.

Further, the research can take an experimental form and aim to understand people's acceptance for AI, particularly chatbots to seek mental health advice. It can be interesting to explore if individuals make an active choice between human-generated and AI-generated responses when they know the source of the response and additionally to investigate the reasons behind their choices. This can help us understand where AI and chatbots stand in terms of acceptance and voluntary usage in mental health services and if they can be used in conjunction with traditional therapies. Studies can be done on a diverse population who may present complex problems and challenge the AI with generation of an individually tailored solution. With a larger and diverse sample, the research findings can further be generalized.

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 22(7), e16021. <https://doi.org/10.2196/16021>
- Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. In IFIP international conference on artificial intelligence applications and innovations (pp. 373–383). Springer, Cham.
- Agarwal, J., & Sharma, S. (2023). Artificial Intelligence Enabled Cognitive Computer-Centered Digital Analysis Model for Examination of the Children's Mental Health.
- American Psychological Association. (2020). Ethical Principles of Psychologists and Code of Conduct. Including 2010 and 2016 Amendments. Washington, DC: American Psychological Association 2020. <https://www.apa.org/ethics/code>
- Aminah, S., Hidayah, N., & Ramli, M. (2023). Considering ChatGPT to be the first aid for young adults on mental health issues. *Journal of Public Health*. <https://doi.org/10.1093/pubmed/fdad065>
- Andersson, G., & Titov, N. (2014). Advantages and limitations of internet-based interventions for common mental disorders. *World Psychiatry*, 13(1), 4–11.
- Asada, M. (2015). Development of artificial empathy. *Neuroscience Research*, 90, 41–50.
- Balaji, D., He, L., Giani, S., Bosse, T., Wiers, R., & de Bruijn, G. J. (2022). Effectiveness and acceptability of conversational agents for sexual health promotion: A systematic review and meta-analysis. *Sexual Health*, 19(5), 391–405. <https://doi.org/10.1071/sh22016>
- Baumeister, H., Reichler, L., Munzinger, M., & Lin, J. (2019). The impact of guidance on internet-based mental health interventions—A systematic review. *Internet Interventions*, 14, 1–10.
- Bilquise, G., Ibrahim, S., & Shaalan, K. (2022). September 26). Emotionally intelligent chatbots: A systematic literature review. *Human Behavior and Emerging Technologies*, 2022, 1–23. <https://doi.org/10.1155/2022/9601630>
- Bird, T., Mansell, W., Wright, J., Gaffney, H., & Tai, S. (2018). Manage your life online: A web-based randomized controlled trial evaluating the effectiveness of a problem-solving intervention in a student sample. *Behavioural and Cognitive Psychotherapy*, 46(5), 570–582.
- Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. *Artificial Intelligence in Healthcare*, 25–60. <https://doi.org/10.1016/b978-0-12-818438-7.00002-2>
- Bond, R. R., O'Neill, S., Ennis, E., & Mulvenna, M. (2019). Chatbots for mental health and suicide prevention: Is this even ethical?. In *European Conference on Mental Health*.
- Bradley, & Alhajjar. (n.d.). AI ethics: Assessing and correcting conversational bias in machine learning based chatbots. icwsm.org. https://workshop-proceedings.icwsm.org/pdf/2022_67.pdf
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv preprint arxiv.org/abs/2303.12712
- Bucci, S., Schwannauer, M., & Berry, N. (2019). The digital revolution and its impact on mental health care. *Psychology and Psychotherapy: Theory, Research and Practice*, 92(2), 277–297.
- Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., & McTear, M. (2017). Towards a chatbot for digital counselling. In Proceedings of the 31st International BCS Human Computer Interaction Conference (HCI 2017). BCS Learning & Development.
- Carlbring, P., Hadjistavropoulos, H., Kleiboer, A., & Andersson, G. (2023). A new era in internet interventions: The advent of ChatGPT and AI-assisted therapist guidance. *Internet Interventions*, 32, 100621. <https://doi.org/10.1016/j.invent.2023.100621>
- Christensen, H., Griffiths, K. M., & Farrer, L. (2009). Adherence in internet interventions for anxiety and depression: Systematic review. *Journal of Medical Internet Research*, 11(2), e13. <https://doi.org/10.2196/jmir.1194>
- 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, 20552076231183544.
- D'Alfonso, S. (2020). AI in mental health. *Current Opinion in Psychology*, 36, 112–117.
- Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48, 24–42.

- Denecke, K., Abd-Alrazaq, A., & Househ, M. (2021). Artificial intelligence for chatbots in mental health: Opportunities and challenges. Multiple perspectives on artificial intelligence in healthcare: Opportunities and challenges, 115–128.
- Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems*, 23(1), 96–138. <https://doi.org/10.17705/1jais.00724>
- Ee, C., Lake, J., Firth, J., Hargraves, F., de Manincor, M., Meade, T., Marx, W., & Sarris, J. (2020). An integrative collaborative care model for people with mental illness and physical comorbidities. *International Journal of Mental Health Systems*, 14(1), 83. <https://doi.org/10.1186/s13033-020-00410-6>
- Erbe, D., Eichert, H. C., Riper, H., & Ebert, D. D. (2017). Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: Systematic review. *Journal of Medical Internet Research*, 19(9), e306.
- Ernsting, C., Dombrowski, S. U., Oedekoven, M., Sullivan, J. L., Kanzler, M., & Kuhlmeier, A. (2020). Using smartphones and health apps to change and manage health behaviors: A population-based survey. *Journal of Medical Internet Research*, 22(2), e16956.
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, 4(2), e19. <https://doi.org/10.2196/mental.7785>
- Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., & Nickless, A. (2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: A single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry*, 5(8), 625–632.
- Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. *JMIR Mental Health*, 5(4), e9782.
- Gaffney, H., Mansell, W., Edwards, R., & Wright, J. (2013). Manage your life online (MYLO): A pilot trial of a conversational computer-based intervention for problem solving in a student sample. *Behavioral and Cognitive Psychotherapy*, 42(6), 731–746. <https://doi.org/10.1017/s135246581300060x>
- Gaffney, H., Mansell, W., & Tai, S. (2019). Conversational agents in the treatment of mental health problems: Mixed-method systematic review. *JMIR Mental Health*, 6(10), e14166. <https://doi.org/10.2196/14166>
- García-Peñalvo, F. J., Llorens Largo, F., & Vidal, J. (2023). The new reality of education in the face of advances in generative artificial intelligence.
- Ghandeharioun, A., McDuff, D., Czerwinski, M., & Rowan, K. (2019). Towards understanding emotional intelligence for behavior change chatbots. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 8–14). IEEE.
- Giger, J. C., Piçarra, N., Alves-Oliveira, P., Oliveira, R., & Arriaga, P. (2019). Humanization of robots: Is it really such a good idea? *Human Behavior and Emerging Technologies*, 1(2), 111–123.
- Harrington, L., et al. (2021). AI-powered chatbots in mental health: Benefits, limitations, and ethical considerations. *Ethics in Health Informatics and Health Information Management*, 6(2), 67–72.
- He, Y., Yang, L., Zhu, X., Wu, B., Zhang, S., Qian, C., & Tian, T. (2022). Mental health chatbot for young adults with depressive symptoms during the COVID-19 pandemic: Single-blind, three-arm randomized controlled trial. *Journal of Medical Internet Research*, 24(11), e40719. <https://doi.org/10.2196/40719>
- Hedman-Lagerlöf, E., Carlbring, P., Svärdman, F., Riper, H., Cuijpers, P., & Andersson, G. (2023). Therapist-supported Internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: An updated systematic review and meta-analysis. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, 22(2), 305–314. <https://doi.org/10.1002/wps.21088>
- Hsu, J. (2023). Mental health service criticised for experiment with AI chatbot. *New Scientist*. <https://www.newscientist.com/article/2354077-mental-health-service-criticised-for-experiment-with-ai-chatbot/>
- Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155–172.
- Ingram, D. (2023). A mental health tech company ran an AI experiment on real users. Nothing's stopping apps from conducting more. <https://www.nbcnews.com/tech/internet/chatgpt-ai-experiment-mental-health-tech-app-koko-rcna65110>
- Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, 6(11), e12106.
- Kaddari, Z., Mellah, Y., Berrich, J., Belkasm, M. G., & Bouchentouf, T. (2020, March). Natural language processing: Challenges and future directions. In *International Conference on Artificial Intelligence & Industrial Applications*, 236–246. Cham: Springer International Publishing.
- Kavakli, M., Li, M., & Rudra, T. (2012). Towards the development of a virtual counselor to tackle students' exam stress. *Journal of Integrated Design and Process Science*, 16(1), 5–26.
- Kelly, S., Kaye, S. A., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, 77, 101925. <https://doi.org/10.1016/j.tele.2022.101925>
- Koelen, J. A., Vonk, A., Klein, A., de Koning, L., Vonk, P., de Vet, S., & Wiers, R. (2022). Man vs. Machine: A meta-analysis on the added value of human support in text-based internet treatments (“e-therapy”) for mental disorders. *Clinical Psychology Review*, 96, 102179.
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., & Bashir, R. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, 25(9), 1248–1258.
- Lindqvist, K., Mechler, J., Midgley, N., Carlbring, P., Carstorp, K., Neikter, H. K., & Philips, B. (2022). “I didn't have to look her in the eyes”—Participants' experiences of the therapeutic relationship in internet-based psychodynamic therapy for adolescent depression. *Psychotherapy Research*, 1–15.
- Luxton, D. D., McCann, R. A., Bush, N. E., Mishkind, M. C., & Reger, G. M. (2011). mHealth for mental health: Integrating smartphone technology in behavioral healthcare. *Professional Psychology: Research and Practice*, 42(6), 505–512. <https://doi.org/10.1037/a0024485>
- Ly, K. H., Ly, A. M., Andersson, G., & Cuijpers, P. (2020). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions*, 20, 100321. <https://doi.org/10.1016/j.invent.2020.100321>
- Maciejewski, J., & Smoktunowicz, E. (2023). Low-effort internet intervention to reduce students' stress delivered with Meta's Messenger chatbot (Stressbot): A randomized controlled trial. *Internet Interventions*, 33, 100653. <https://doi.org/10.1016/j.invent.2023.100653>
- Miloff, A., Carlbring, P., Hamilton, W., Andersson, G., Reuterskiöld, L., & Lindner, P. (2020). Measuring alliance toward embodied virtual therapists in the era of automated treatments with the virtual therapist alliance scale (VTAS): Development and psychometric evaluation. *Journal of Medical Internet Research*, 22(3), e16660.
- Minerva, F., & Giubilini, A. (2023). Is AI the future of mental healthcare?. *Topoi*, 42(3), 809–817. <https://doi.org/10.1007/s11245-023-09932-3>
- Morris, R. R., Kouddous, K., Kshirsagar, R., & Schueller, S. M. (2018). Towards an artificially empathic conversational agent for mental

- health applications: System design and user perceptions. *Journal of Medical Internet Research*, 20(6), e10148.
- Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., & Unutzer, J. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, 392(10157), 1553–1598. [https://doi.org/10.1016/s0140-6736\(18\)31612-x](https://doi.org/10.1016/s0140-6736(18)31612-x)
- Pathare, S., Brazinova, A., & Levav, I. (2018). Care gap: A comprehensive measure to quantify unmet needs in mental health. *Epidemiology and Psychiatric Sciences*, 27(5), 463–467. <https://doi.org/10.1017/S2045796018000100>
- Paul, S. C., Bartmann, N., & Clark, J. L. (2021). Customizability in conversational agents and their impact on health engagement. *Human Behavior and Emerging Technologies*, 3(5), 1141–1152.
- Qin, C., Zhang, A., Zhang, Z., Chen, J., Yasunaga, M., & Yang, D. (2023). Is ChatGPT a general-purpose natural language processing task solver?. arXiv preprint [arXiv:2302.06476](https://arxiv.org/abs/2302.06476).
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), 9.
- Rana, J. (2023). The pros and cons of healthcare chatbots | REVE Chat. REVE Chat. <https://www.revechat.com/blog/healthcare-chatbots/>
- Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, 151, 102630.
- Rashkin, H., Smith, E. M., Li, M., & Boureau, Y. L. (2018). Towards empathetic open-domain conversation models: A new benchmark and dataset. arXiv preprint [arXiv:1811.00207](https://arxiv.org/abs/1811.00207).
- Reddy, M., Reddy, T., & Kumar, N. (2021). AI chatbot for mental health : PTSD assessment . *International Journal of Creative Research Thoughts (IJCRT)*, 9(5), f145–f150. <https://ijcrt.org/papers/IJCRT2105554.pdf>
- Sabour, S., Zhang, W., Xiao, X., Zhang, Y., Zheng, Y., Wen, J., Zhao, J., & Huang, M. (2023). A chatbot for mental health support: exploring the impact of Emohaa on reducing mental distress in China. *Frontiers in Digital Health*, 5. <https://doi.org/10.3389/fdgth.2023.1133987>
- Santoro, E., & Monin, B. (2023). The AI effect: People rate distinctively human attributes as more essential to being human after learning about artificial intelligence advances. *Journal of Experimental Social Psychology*, 107, 104464. <https://doi.org/10.1016/j.jesp.2023.104464>
- Seifert, C., Vogel, L., Aas, B., Brandhorst, I., Carlbring, P., Conzelmann, A., Esfandiari, N., Finkbeiner, M., Hollmann, K., Lautenbacher, H., Meininger, E., Newbold, A., Opitz, A., Renner, T. J., Sander, L. B., Santangelo, P. S., Schoedel, R., Schuller, B., Stachl, C., & Löchner, J. (2023). How to e-mental health: A guideline for researchers and practitioners using digital technology in the context of mental health. *Nature Mental Health*, 1, 542–554. <https://doi.org/10.1038/s44220-023-00085-1>
- Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2023). Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence*, 5(1), 46–57.
- Singh, O. P. (2023). Artificial intelligence in the era of ChatGPT- Opportunities and challenges in mental health care. *Indian Journal of Psychiatry*, 65(3), 297.
- Smith, K. A., Blease, C., Faurholt-Jepsen, M., Firth, J., Van Daele, T., Moreno, C., Carlbring, P., Ebner-Priemer, U. W., Koutsouleris, N., Riper, H., Mouchabac, S., Torous, J., & Cipriani, A. (2023). Digital mental health: Challenges and next steps. *BMJ Mental Health*, 26(1), e300670. <https://doi.org/10.1136/bmjment-2023-300670>
- Suganuma, S., Sakamoto, D., & Shimoyama, H. (2018). An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: Feasibility and acceptability pilot trial. *JMIR Mental Health*, 5(3), e10454.
- Suta, P., Lan, X., Wu, B., Mongkolnam, P., & Chan, J. H. (2020). An overview of machine learning in chatbots. *International Journal of Mechanical Engineering and Robotics*.
- Thompson, D. C., & Bentzien, J. (2020). Crowdsourcing and open innovation in drug discovery: Recent contributions and future directions. *Drug Discovery Today*, 25(12), 2284–2293.
- Tielman, M. L., Jonker, C. M., & van Riemsdijk, M. B. (2018, July). What Should I Do? Deriving Norms from Actions, Values and Context. In *MRC@ IJCAI* (pp. 35–40).
- Trautmann, S., Rehm, J., & Wittchen, H. U. (2016). The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? *EMBO Reports*, 17(9), 1245–1249.
- Vaidyam, A., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *Canadian Journal of Psychiatry*, 64(7), 456–464. <https://doi.org/10.1177/0706743719839389>
- Vis, C., Schuurmans, J., Aouizerate, B., Atipei Craggs, M., Batterham, P., Bührmann, L., Calcar, A., Cerga Pashoja, A., Christensen, H., Dozeman, E., Duedal Pedersen, C., Ebert, D. D., Etzelmueller, A., Fanaj, N., Finch, T. L., Hanssen, D., Hegerl, U., Hoogendoorn, A., Mathiasen, K., Riper, H. (2023). Effectiveness of self-guided tailored implementation strategies in integrating and embedding internet-based cognitive behavioral therapy in routine mental health care: Results of a multicenter stepped-wedge cluster randomized trial. *Journal of Medical Internet Research*, 25, e41532. <https://doi.org/10.2196/41532>
- Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, 5(3), 457–469.
- Wampold, B. E., & Flückiger, C. (2023). The alliance in mental health care: Conceptualization, evidence, and clinical applications. *World Psychiatry*, 22(1), 25–41.
- Wang, X., & Nakatsu, R. (2013). How do people talk with a virtual philosopher: Log analysis of a real-world application. In *Entertainment Computing–ICEC 2013: 12th International Conference, ICEC 2013, São Paulo, Brazil, October 16–18. (2013). Proceedings 12*, 132–137. Springer.
- World Health Organization. (2021). Mental disorders. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/mental-disorders>
- Wu, F., Lu, C., Zhu, M., Chen, H., Zhu, J., Yu, K., ... & Pan, Y. (2020). Towards a new generation of artificial intelligence in China. *Nature Machine Intelligence*, 2(6), 312–316.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.