Review Psikiyatride Güncel Yaklaşımlar-Current Approaches in Psychiatry 2025; 17(4):643-660
doi: 10.18863/pgy.1538432

Artificial Intelligence and Psychotherapy

Yapay Zeka ve Psikoterapi

Fatih Bal[1]

[1]Sakarya University, Sakarya

ABSTRACT

With the development of artificial intelligence technologies, changes have also begun to be seen in psychotherapy. Although artificial intelligence does not currently have a major impact on the therapy field, it raises major questions about the nature of therapy and the value of the relationship future between people and therapists in understanding how artificial intelligence can be included in the therapy process, foreseeing the future, and being proactive are gaining importance. This article will examine current artificial intelligence applications used in psychotherapy fields with a literature review. Artificial intelligence can be used to increase the effectiveness of psychotherapy. However, it should not be forgotten that excessive reliance on artificial intelligence can overshadow the human aspect of psychotherapy and that the human factor is important. Although there are still uncertainties about how the profession will be affected by the use of artificial intelligence in psychotherapy and how it will be incorporated into therapy processes, it is envisaged that artificial intelligence can play a versatile role in psychotherapy.
Keywords: Artificial intelligence, therapy, psychotherapy

ÖZ

Gelişen yapay zeka teknolojileri ile psikoterapi alanında da değişimler görülmeye başlamıştır. Yapay zeka, terapi alanında şu anda büyük bir etkiye sahip olmasa da, gelecekte terapinin niteliği ve insanlar ile terapistler arasındaki ilişkinin değeri konusunda büyük sorular ortaya çıkartmaktadır. Yapay zekanın terapi sürecine nasıl dahil edilebileceğini anlamaya çalışmak geleceği öngörmek, proaktif olmak önem kazanmaktadır. Bu makalede psikoterapi alanlarında, kullanılan mevcut yapay zeka uygulamaları literatür taraması ile irdelenecektir. Psikoterapinin etkinliğini artırmak için yapay zekanın kullanılabilinmektedir. Ancak yapay zekaya aşırı güvenmenin psikoterapinin insan yönünü gölgeleyebileceği ve insan faktörünün önemli olduğunun unutulmaması gerekmektedir. Psikoterapide yapay zeka kullanımında mesleğin nasıl etkileneceği ve terapi süreçlerine nasıl dahil edileceği konusunda hala belirsizlikler olmakla birlikte yapay zekanın psikoterapide çok yönlü bir rol oynayabileceği öngörülmektedir.
Anahtar sözcükler: Yapay zeka, terapi, psikoterapi

Introductiom

Psychotherapy is a frequently utilised intervention in mental health services, and is a preferred option for clients (Andrews et al. 2001, Lambert 2004). Psychotherapy is an effective treatment for a range of psychiatric and psychological disorders in a variety of settings and populations (Nathan and Gorman 2015). Psychotherapy can be defined as an intervention involving the interaction of treatment-specific components, offered as part of a professional relationship, with the aim of bringing about clinical change in the symptoms and functions of clients (Richards 2024).

Psychiatric and psychological disorders are acknowledged as significant global issues with a notable economic impact (Kessler et al. 2009). In recent years, technological developments have been employed in the context of applied interventions in psychiatry. Digital technologies, including telehealth services and augmented/virtual reality (AR/VR), have been demonstrated to be effective in the diagnosis and treatment of mental health disorders (Sonmez and Hocaoglu 2024). As technology advances, it has become feasible for Artificial intelligence (AI) to fully emulate the capabilities of human imagination, emotions, intuition, and potential (Braga and Logan 2017). Furthermore, it is now feasible for AI to comprehend psychological processes, perceive and respond to a spectrum of human behaviours, including attention, motivation, emotion, creativity, planning and discussion (van den Bosch and Bronkhorst 2019, Dong et al. 2020).

The application of AI in therapy is not confined to the support of therapists and therapy chatbots. Such technology can assist therapists in the management of administrative tasks, thereby increasing efficiency and allowing more time to be devoted to direct patient care. Moreover, chatbots and virtual assistants are being

Address for Correspondence: Fatih Bal, Sakarya University Faculty of Humanities and Social Sciences, Department of Psychology, Sakarya, Türkiye **E-mail:** fatihbal@sakarya.edu.tr
Received: 25.08.2024 | **Accepted:** 29.12.2024

employed to furnish preliminary assistance and triage counsel. Such technology is regarded as a valuable adjunct to first aid by numerous individuals, including those employed in independent practices and larger organisations. The objective of this study is to provide a comprehensive overview of the intersections between AI and psychotherapy.

Areas of Use of Artificial Intelligence in Psychotherapy

AI is playing an increasingly prominent role in the field of mental healthcare, driving transformative changes across the entire spectrum of care, from initial diagnosis to subsequent treatment (Thakkar et al. 2024). The advent of AI is inaugurating a new epoch in the domain of mental health care, revolutionizing the entire spectrum of activities, from the precision of diagnostic procedures to the implementation of therapeutic interventions (D'Alfonso 2020). AI offers cost-effective assistance in response to growing demand and, in some cases, replaces human-led treatments. The field of AI has witnessed considerable advancement over time, from the utilisation of chatbots in therapy sessions to the development of sophisticated data analysis tools. In recent years, with the continued evolution of this field, there have been notable advances in the application of AI in therapeutic contexts. AI has become capable of analysing a range of data, including tools, speech patterns and facial expressions, thereby providing therapists with additional information about the client's emotional state.

The utilisation of AI in the domain of mental health can be traced back to the publication of Alan Turing's seminal 1950 paper, Computing Machinery and Intelligence (Turing 1950). The advent of Eliza, the inaugural chatbot to employ AI as a conversational methodology, marked the advent of AI utilisation in psychological contexts. In 1966, it facilitated a genuine sense of interaction between clients and human therapists (Weizenbaum 1976). Although Eliza's capacity for speech was constrained, it marked the advent of the discourse surrounding the potential for human-machine communication. During the 1990s and early 2000s, considerable progress was made in the utilisation of AI technologies for data analysis. In the field of healthcare, AI-enabled tools have been developed with the objective of discerning patterns in human behaviour and gaining insights through the analysis of data. These tools have made considerable headway in aiding therapeutic processes through the utilisation of quantitative data, thereby facilitating the assessment of treatment efficacy. This has facilitated the development of process-oriented methodologies, such as cognitive behavioural therapy. Furthermore, in the mid-2000s, AI-supported Socratic chatbots emerged. Such robots are capable of engaging in structured dialogues with their users, thereby providing a space for introspection regarding thoughts, feelings, and behaviour. Despite the limitations of such chatbots in mimicking the subtleties of human discourse, they have served as a non-judgemental conduit, facilitating a depth of exploration into the psyche. AI has demonstrated considerable potential for meeting client needs through the analysis of large data sets. The findings of studies conducted in this area demonstrate that the utilisation of AI can markedly enhance the efficacy of mental health services and more effectively address the needs of clients (D'Alfonso 2020, Holohan and Fiske 2021).

Use of Artificial Intelligence in Psychotherapy

Although AI-assisted virtual therapy and robot therapy have been employed in numerous medical fields for some time, the incorporation of AI through embodied agents remains a nascent phenomenon in the domain of mental health care. It represents the most recent addition to psychotherapy practices, offering support for a spectrum of emotional, cognitive, and social processes (Calderita et al. 2014).

The field of AI-enabled psychotherapeutic devices is undergoing rapid development. AI tools, including chatbots and virtual assistants, have the potential to emulate the work of psychologists and psychotherapists and even assist in meeting the fundamental therapeutic needs of individuals. For example, therapeutic applications such as Tess and chatbots such as Earkick, Mindspa, Youper, Sara, Wysa and Woebot, and Replikag, which operate through short message service, WhatsApp, internet platforms or mobile applications, are being investigated for their potential to address depression and anxiety. Such applications facilitate interaction. Woebot and other analogous programmes interact with the client in the capacity of a virtual psychotherapist, with the objective of assisting the client in recognising patterns of emotions and thoughts, and in developing skills such as resilience or anxiety reduction techniques. To illustrate, Tess employs natural language processing to discern expressions indicative of emotional distress. Chatbots are frequently asserted to be a digital instrument for reaching populations lacking mental health services on a global scale. Such applications are capable of elucidating complex clinical terminology, such as that pertaining to cognitive distortions, and can also furnish users with practical guidance on how to recognise and respond to challenging circumstances. Tess has been demonstrated to be an effective tool for alleviating depressive and anxious symptoms among users (Sachan 2018).

In recent years, avatar therapy has emerged as a powerful tool for the application of cognitive behavioural techniques for mood management and the exploration of alternative modes of being in the world, including those facilitated by virtual reality. The objective of avatar therapy is to equip individuals who hear voices with the tools to better cope with and challenge these voices. The therapy encourages individuals to challenge negative voices by creating and interacting with avatars. Guided by a therapist, these individuals learn to challenge, question, and test threats and negative statements (Ward et al. 2020). In addressing the issue of auditory hallucinations, avatar therapy provides individuals with the opportunity to develop self-confidence and self-kindness. The SPARX programme is an online therapeutic intervention designed for adolescents experiencing depressive symptoms. The programme is designed for young people with mild to moderate depression and is intended to serve as an adjunct to, rather than a replacement for, the input of a therapist. The SPARX programme is based on cognitive behavioural therapy (CBT) and aims to facilitate the development of more positive cognitive processes and emotional regulation in young people. The programme is comprised of a series of levels, which facilitate progression towards the attainment of the learning goals of the young people participating in the programme. As participants progress with their avatars, they are able to develop problem-solving abilities and identify negative thought patterns. A minimum of four sessions is required for completion of the programme.

The utilisation of avatars, such as avatar therapy, for the management of auditory hallucinations in psychotic patients necessitates the utilisation of facial images created on computer screens or tablets that interact with clients through intelligent algorithms (Craig et al. 2018). It is estimated that one in four individuals diagnosed with a psychotic disorder will continue to experience persistent auditory verbal hallucinations despite receiving treatment. Avatar therapy represents a novel approach in which individuals experiencing auditory hallucinations engage in a dialogue with an avatar whose voice is provided by the therapist. In a single-blind, randomised controlled trial conducted by Craig et al. (2018) to evaluate the efficacy of avatar therapy compared to supportive counselling 150 individuals diagnosed with schizophrenia spectrum or affective disorder and experiencing auditory hallucinations were randomly assigned to avatar therapy or supportive counselling. The results demonstrated that, after a 12-week period, the reduction in auditory verbal hallucinations was more pronounced in the avatar therapy group than in the supportive counselling group (Craig et al. 2018). This indicates that avatar therapy may be an appropriate method for the management of auditory verbal hallucinations. Furthermore, avatars are employed in the treatment of schizophrenia, particularly to enhance medication adherence (Bain et al. 2017).

Virtual reality-assisted therapy for schizophrenia employs the use of AI avatars to facilitate interaction with auditory hallucinations experienced by the patient. The available evidence indicates that the therapy can reduce the formulation of therapeutic goals in cases of schizophrenia that are particularly challenging (Dellazizzo et al. 2018). A study on patients diagnosed with treatment-resistant schizophrenia revealed that following therapy sessions, there were improvements in auditory and visual hallucinations, depression symptoms, and overall quality of life (du Sert et al. 2018). Furthermore, avatars have been employed in the treatment of acrophobia, providing medical students with realistic interview practice (Freeman et al. 2018) and in risk prevention training to identify individuals at risk of suicide (Rein et al. 2018).

The field of AI robot therapy represents an area of clinical practice in which clinicians and scientists collaborate to integrate innovations in AI and robotics. Robots that resemble furry animals, such as Paro, serve as assistants for patients with dementia. Paro is a healthcare assistant that engages in dynamic interaction with individuals, responding to verbal and non-verbal cues in a manner that can be described as a form of 'dialogue'. Furthermore, companion robots are being employed to assist elderly patients who are isolated or depressed in their own homes, providing them with a source of companionship. It has been demonstrated that such robots can play a role in reducing factors such as stress, loneliness, and agitation, as well as improving mood and social connections (Wada and Shibata 2007, Yu et al. 2015).

In 2022, the introduction of ChatGPT, a dialogic AI software developed by OpenAI, prompted interest in the potential of such technology to facilitate the creation of therapeutic and therapist tools, with particular reference to its possible applications in the field of mental health (Minerva and Giubilini 2023). A study demonstrated the efficacy of an AI assessment tool in identifying and categorising psychiatric and psychological disorders in clients with 89% accuracy, based on responses to only 28 questions (Tutun et al. 2023). A meta-analysis of research on mental health conversational agents indicates that chatbots can effectively alleviate psychological distress and may even play a role in the formation of therapeutic relationships (Li et al. 2023). In the treatment of psychiatric and psychological disorders, AI tools assist therapists in the identification and treatment of clients' conditions (Koutsouleris et al. 2022, Li et al. 2023). The use of AI-based chatbots in therapy has demonstrated efficacy in alleviating symptoms of depression and anxiety (Holohan and Fiske 2021). Such robots can facilitate diagnosis by offering virtual psychotherapy services, while also enabling consultations, providing psychoeducation, and

offering treatment options. The use of AI enables the provision of more personalised and adaptive responses through a variety of interaction modalities, including text and voice (Li et al. 2023). The deployment of AI tools for the treatment of psychiatric and psychological disorders is contingent upon the capacity of clients to report on their cognitive and emotional state, the progression of their symptoms, and information from their environment (Koutsouleris et al. 2022). Nevertheless, the input of the clinician remains indispensable for the formulation of diagnoses and treatment recommendations.

Chatbots can be integrated into mobile or instant messaging applications to facilitate a range of useful functions, including aiding diagnosis, facilitating consultations, providing psychoeducation and offering treatment (Li et al. 2023). The utilisation of AI represents an efficacious methodology for the alleviation of psychiatric and psychological disorders, including depression and distress. The efficacy of this approach is contingent upon the quality of the human-AI therapeutic relationship (Li et al. 2023). Natural language processing (NLP) is a technique that enables the analysis of the client's language in a variety of contexts, including conversations, chats, emails, and social media posts. It is capable of detecting patterns associated with psychological disorders such as depression or anxiety, and constitutes a vital component of chatbots (Holohan and Fiske 2021, Li et al. 2023). Chatbots, which communicate with users in natural language through messaging applications, websites, and mobile apps, represent an alternative technique for supporting psychological problems. Woebot is an exemplar of a user-friendly chatbot based on cognitive behavioural therapy. Woebot is a chatbot designed to monitor and assess the emotional state of users. Woebot employs chat scenarios that are designed to assist users in comprehending and addressing their issues.

The chatbot initiates a dialogue with the user, soliciting information about their day. In the event of a negative response, the chatbot records the reasons provided and offers solutions. Woebot employs sophisticated techniques, including natural language processing, sentiment analysis, and the use of humour, to personalise the chat and enhance the conversational atmosphere. Furthermore, the application assists users in examining and questioning their assumptions through the utilisation of quizzes and videos. Woebot has been designed in such a way that it is able to evolve in accordance with the user's changing knowledge base, thereby becoming increasingly effective with continued use. Wysa is an online therapy application that employs an emotionally intelligent chatbot based on the cognitive behavioural therapy model. The app offers a variety of tools, including dialectical behaviour therapy, meditation, breathing exercises, yoga, and motivational videos. Additionally, the app furnishes users with self-improvement tools, including medication reminders, wellness exercises, and deep sleep training. The objective of the app is to address issues such as mild depression and anxiety disorders. On occasion, the vast quantity of queries can prove daunting for users. Sanvello is a chatbot based on CBT that places greater emphasis on cultivating mindfulness than on psychological relaxation. The app offers a range of features designed to facilitate the application of therapeutic techniques, including mood tracking and meditation. In contrast to other chatbots, Sanvello has been developed with audio and video recognition tools integrated with text input. This approach facilitates the analysis of images and sounds, thereby enhancing the user's mood and facilitating its comprehension. The chatbot does not pose questions that are irrelevant or distressing while monitoring the user's emotional state and providing suitable counsel.

The chatbot offers users a range of tools to assist them in managing stressful situations, monitoring their moods on a daily basis, and providing support in the event of an emergency. The most distinctive feature of Sanvello is that it provides users with the ability to engage in social activities, which differentiates it from other similar applications. The application serves to facilitate connections between users, thereby enabling them to exchange experiences and provide mutual support. This represents a significant departure from the concept of expertise and provides an extremely positive experience. The Joyable application comprises educational, behavioural and thought-based activities. The application was developed by experts in therapeutic practice and scientists and offers users a coaching service and cognitive behavioural therapy training based on real-life scenarios. The chatbot assists users in identifying a particular event that evokes negative emotions and cognitions, and provides strategies for their elimination. The Joyable application offers its users a coaching service via telephone and email, in addition to the application itself. Chatbots that provide online therapy are of great importance in terms of data collection. As the technology behind chatbots continues to develop and their data processing capabilities increase, the assistance they provide to users becomes increasingly effective. The popularity of chatbots offering ease of access and online therapy support is on the rise. Nevertheless, the question of whether chatbots can assume the role of human therapists remains a topic of contention. The advent of digital technology and AI applications has precipitated a transformation in the field of mental health.

The application of AI to mental health services has the potential to facilitate improvements in several areas, including prediction, detection, and treatment processes. AI can be employed in both research and practice to

predict the presence of psychiatric and psychological disorders. The models are trained using the responses of the participants to the assessment questions and other historical data (Tutun et al. 2023). Decision support systems can assist mental health professionals in making evidence-based treatment decisions, analysing data, and providing recommendations for the diagnosis and treatment of psychiatric and psychological disorders. The implementation of decision support systems has the potential to reduce overall healthcare costs associated with misdiagnosis, overdiagnosis and unnecessary treatment as a result of accurate diagnosis of psychiatric and psychological disorders (Tutun et al. 2023). Moreover, web and smartphone-based applications facilitate digital interventions, personalise mental health, and enhance the user experience. The data collected by smartphones and smartphone sensors can be employed to gain insights into behaviour and mental health, and to predict the onset of mental health conditions. The analysis of clinical texts and social media content represents a valuable tool for the detection of mental health conditions. Furthermore, it plays a pivotal role in the development of conversational agents designed to facilitate therapeutic interventions. Furthermore, chatbots and virtual agents provide accessible therapeutic options for a range of mental health issues. Furthermore, mobile devices have the potential to facilitate real-time psychological interventions and behavioural guidance. Consequently, precision medicine in mental health can mitigate issues pertaining to precise diagnoses, prognoses, and therapeutic choices (D'Alfonso 2020, Koutsouleris et al. 2022).

Advantages of Artificial Intelligence-Based Psychotherapy

One of the benefits of AI-based therapy is that it offers mental health professionals the opportunity to save time and streamline their organisational processes. Such assistance encompasses the automatic transcription of notes taken during video meetings, the review and synthesis of client notes, and the creation of bespoke exercises, activities, and interventions. The advantages of conducting therapy with AI include improved accessibility and utility 24/7 continuous availability, capacity to analyse extensive datasets, cost-effectiveness, reduced stigma, and greater efficiency in diagnosis and monitoring. Those who engage in discourse with an AI therapist may perceive a greater sense of psychological safety and reduced judgment. In this way, AI therapy offers significant advantages in the field of mental health (Holohan and Fiske 2021, Koutsouleris et al. 2022, Li et al. 2023, Tutun et al. 2023).

The advantages and challenges of AI in therapy have been the subject of extensive research. The capacity of AI to adapt its communication, activities, feedback and counselling to the specific needs of the client has been demonstrated to be a significant advantage (Holohan and Fiske 2021, Tutun et al. 2023). Chekroud et al. (2024) observe that it can integrate disparate databases in psychotherapy, a practice currently prevalent in the field. Furthermore, the generation of synthetic data presents a distinct advantage. Despite the success of AI in drug discovery, its application in psychotherapy remains an emerging area of research. It is imperative that researchers adhere to the established psychotherapy theory and practice framework in order to analyse data with AI. Formulating evidence-based questions and hypotheses allows for the interpretation of results and the identification of any limitations in the quality of the data. Moreover, it is imperative to implement safeguards to prevent the inappropriate extrapolation of data and to ensure the dissemination of data in a responsible manner (Chen et al. 2020).

AI has the potential to facilitate the professional advancement of mental health professionals. This development can be achieved through the implementation of enhanced training opportunities, AI-supported supervision, and self-reflection. AI can be employed to simulate the role of the client in the development and practice of new skills, thereby assisting clinicians in enhancing their abilities by more accurately portraying a range of scenarios (Luxton 2014). AI-based tools, such as Clinicalnotes.ai, have the potential to provide clinicians with supervision-aligned interventions by analysing their notes and treatment progress. Moreover, these tools can assist supervisors in the planning of clinical practice.

Natural language processing (NLP) is a subfield of artificial intelligence and linguistics that focuses on the interaction between computers and human language. Text analysis, a significant application of NLP, seeks to derive meaningful insights from vast quantities of unstructured text data. The application of natural language processing (NLP) to verbal or text analysis enables clinicians to assess the emotional state of their clients and provide objective feedback. These developments can assist mental health professionals in developing their skills in a less resource-dependent manner, thereby enhancing the efficacy of their clinical practice.

The integration of AI capabilities with psychotherapists' expertise via a hybrid approach may prove to be a promising avenue for the future of mental health services. Among the advantages of AI are cost-effectiveness and greater accessibility for a larger number of people. AI automates routine tasks, thereby allowing therapists to devote a greater proportion of their time to client care. For therapists, AI-integrated software platforms

facilitate the streamlining of daily operations, thereby reducing the time and effort otherwise required. AI tools provide clients with support between sessions, offering mood monitoring, stress management techniques, and self-help guidance. It is compatible with cognitive behavioural therapy and other solution-focused therapies. AI is employed in pioneering interventions such as virtual reality therapy, which offers immersive experiences that are not feasible in conventional therapeutic environments. Virtual reality therapy is employed in the treatment of conditions such as post-traumatic stress disorder. Furthermore, it offers training and supervision for therapists through the provision of simulations that emulate real-life scenarios. The use of AI allows therapists to interact with their clients in a dynamic manner, thereby facilitating the provision of insights and perspectives between supervision sessions. Nevertheless, the utilisation of AI has the potential to erode fundamental aspects of healthcare delivery, such as empathy and trust. Nevertheless, the advent of AI in mental healthcare has the potential to yield favourable outcomes and benefits that could challenge the conventional wisdom based on human connection. It is anticipated that AI will facilitate further advances in the diagnosis of psychiatric and psychological disorders through the analysis of data drawn from a range of sources, including medical records, social media and wearable devices. Moreover, AI will be beneficial for individuals who encounter challenges in human interaction, such as those with depression or autism, or who are concerned about being stigmatised. Nevertheless, the extent to which AI can be employed to promote mental well-being remains uncertain.

Disadvantages of Artificial Intelligence-Based Psychotherapy

The utilisation of AI tools for therapeutic purposes and by therapists is not without its inherent challenges and risks. There are concerns regarding the reliability of these systems and their potential impact on mental health (Koutsouleris et al. 2022). One of the primary challenges and criticisms associated with the utilisation of AI in therapeutic settings pertains to its inability to empathise and establish authentic interpersonal connections with individuals (Minerva and Giubilini 2023). The deployment of AI in therapeutic contexts is constrained by a number of limitations and ethical considerations. The integration of AI into therapeutic practices is possible, yet the expertise and interpersonal skills of therapists remain indispensable (Minerva and Giubilini 2023). The intricate nature of human psychology renders it challenging for AI algorithms to address the distinctive requirements of each individual. An overreliance on AI in mental health services has the potential to diminish clients' access to emotional support and decision-making abilities, which could ultimately result in an increased sense of dependency. Moreover, the long-term effects of AI and the ethical and privacy concerns it may raise in the context of mental health services warrant consideration (Minerva and Giubilini 2023). The utilisation of AI in therapeutic contexts may give rise to ethical concerns. It is of the utmost importance to guarantee the confidentiality of client data, the therapist's autonomy in decision-making, and the protection of the human element. To illustrate, AI models amass personal data, thereby creating a potential risk of misuse or disclosure. Nevertheless, the dearth of empirical evidence is a source of contention. Some parties propose a complete prohibition on the collection of data, while others argue that it is possible to combine data from disparate sources using privacy-enhancing technologies in a secure manner. Moreover, it is possible for AI to learn biases that could potentially influence the treatment process or result in misdiagnosis.

This risk is compounded by the generation of increasingly complex data sets that encompass previously unmeasured dimensions of human experience, including physical, neurological, psychological and emotional aspects (Farahany 2023). Moreover, the deployment of AI to supplant the role of human physicians may diminish the human aspect of healthcare (Minerva and Giubilini 2023). It is of the utmost importance to rely on AI analysis when utilising client data, although it is acknowledged that the accuracy and interpretation of the data may vary. There is a risk of misinterpretation or simplification of complex human emotions and experiences. In the context of utilising an AI tool for the analysis of facial expressions and tone of voice, a decision must be made as to whether the output produced by the machine should be trusted or whether expertise and the context of the clients in question should be relied upon. While AI can provide valuable insights, ultimate decision-making authority must reside with the human operator. Some clients may prefer the convenience of 24/7 AI-based services, however, it is important to recognise that the foundation of therapy must be built on genuine empathy and connection. Moreover, it is essential that the client be an active participant in the therapeutic process. Providing them with uninterrupted access may foster an overconfident outlook.

It is imperative that measures be taken to address the inherent limitations of AI models. Firstly, models must be trained with high-quality, unbiased data (Koutsouleris et al. 2022). Moreover, data privacy and security are of paramount importance. It is therefore imperative that data is safeguarded against unauthorised access. It is imperative that users are informed about the intended use of their data, and that transparency is ensured.

Moreover, it is imperative to implement procedures and controls to ensure the fairness and impartiality of AI tools. The data on AI psychotherapy may be biased. For example, cognitive behavioural therapy was initially developed and subsequently applied predominantly to white, well-educated heterosexuals. The use of historical data in the construction of algorithms may result in the exclusion of significant demographic groups, including those with neurodiversity, racial and ethnic minorities, cultural diversity, and individuals from disparate socio-economic backgrounds. Moreover, historical context, background, experiences, and context are crucial elements in psychotherapy. For instance, the data sources employed in the schizophrenia study were selected for "comparability and consistency," which are essential for developing inferences. However, numerous environmental factors that are pivotal for the treatment outcomes of psychiatric and psychological disorders have not been taken into account.

The application of AI automates homogeneity and, as a consequence, places the individuals whose lives the data represents in a position of relative uncertainty (Crawford 2021). It is possible that AI may not fully comprehend the subtleties of human experience and psychology, which could result in misguided interventions. It is of the utmost importance to take measures to prevent potential biases from influencing the outcome. It is of the utmost importance to establish transparent and unambiguous regulations pertaining to liability in the event of any adverse consequences. It is imperative to emphasise that the use of AI recommendations should be limited to a supplementary role in human decision-making processes. It is incumbent upon therapists utilising AI tools to monitor the functionality of said tools and the recommendations they generate. The aforementioned measures facilitate the overcoming of the limitations of AI models, thereby ensuring their utilisation in a beneficial and appropriate manner.

Research on the Effectiveness of Artificial Intelligence-Assisted Therapies

The utilisation of AI has the potential to markedly enhance psychotherapy and mitigate the expression of clinical psychiatric and psychological disorders, resulting in noteworthy levels of satisfaction, engagement and retention (Gual et al. 2022). The role of AI in mental health is becoming increasingly significant, particularly in the context of therapeutic interventions within psychotherapy. The range of innovations is diverse, encompassing virtual psychotherapists, social robots in dementia care and autism disorder, and robots for sexual disorders (Torjesen 2017, Martin and Kreitmair 2018, Góngora et al. 2019). Virtual and robotic agents with AI are becoming increasingly prevalent in mental health support. They are capable of providing not only basic comfort and social interaction but also high-level therapeutic interventions that were previously the domain of highly trained and skilled health professionals such as psychotherapists (Inkster et al. 2018). It is noteworthy that such virtual or robotic therapists comprise an artificially intelligent algorithmic entity that responds to the client or patient through a virtually embodied interface, such as a facial icon, or a physically embodied entity, such as a robotic interface, independent of any expert clinician guidance. These new applications differ from many forms of web-based therapy, which often require human, albeit remote, therapy or the client working independently with manuals, questionnaires, or other self-help materials (Mehrotra et al. 2017). The utilisation of AI-based therapist tools and software has the potential to enhance the quality of care provided to clients, whilst simultaneously increasing the accessibility and standard of mental health services (Minerva and Giubilini 2023).

The utilisation of AI applications in the domain of mental health services has the potential to enhance the quality of care provided and exert a beneficial influence on expenditure control. Moreover, these applications have the potential to enhance access to mental health services for currently underserved population groups and vulnerable individuals. Nevertheless, despite these rapid developments, there remains a gap in the successful adaptation of technological tools by therapists and clients in clinical settings. Moreover, the majority of studies assess the efficacy or ethical utilisation of a single application, yet there is a paucity of research examining the broader impact of AI integration in mental health (Fiske et al. 2019). Moreover, some therapists employ AI to address ethical concerns or enhance their comprehension of theoretical frameworks. Some therapists utilise AI-assisted therapy to provide support to their clients outside of scheduled sessions. It is, however, important to note that this use carries certain risks.

The role of AI in therapy is currently regarded as supplementary, providing additional resources and support while maintaining the primacy of the therapeutic relationship and decision-making authority with professionals (Koutsouleris et al. 2022). The advent of novel technologies, including virtual psychotherapists, social robots for dementia care and autism disorder, and robots for sexual disorders, has resulted in the emergence of artificially intelligent virtual and robotic agents capable of performing advanced therapeutic interventions that were previously the domain of highly trained healthcare professionals. It is proposed by researchers that a hybrid

approach combining the input of therapists and AI will shape the future of mental healthcare (Minerva and Giubilini 2023).

Use of Artificial Intelligence in Diagnosis and Classification of Psychiatric Disorders

AI has the potential to play an important role in the diagnostic process, utilising large datasets to provide guidance and objectivity. In a study conducted by Nemesure et al. (2021), AI methods were employed to detect common disorders based on electronic health record data and biomedical and demographic information. These techniques were employed to predict generalized anxiety disorder and major depressive disorder, and advanced methods were utilized to identify the most salient risk factors for each disorder. This approach facilitates early diagnosis and intervention for clients who are challenging to diagnose with readily available information. Rosellini et al. (2020) employed AI to forecast the incidence of internalising disorders using both survey data and prospective survey data, thereby developing algorithms that predict the occurrence of internalising disorders (generalised anxiety, panic, social phobia, depression, and mania) in adults. The efficacy of these algorithms has been demonstrated in the prediction of depression and social phobia, as well as the provision of preventive intervention. Reese and Danforth (2017) employed AI techniques to analyse Instagram photos (content, colour, metadata) posted by individuals diagnosed with depression. The resulting models demonstrated an accuracy of 70% in predicting depression, exceeding the success rate of clinicians' unaided diagnoses. This study illustrates the potential of novel approaches to early screening utilising social media data.

The process of diagnosing and classifying psychiatric and psychological disorders is inherently time-consuming, subjective, and limited in scope. The application of AI-based tools has the potential to enhance the accuracy and efficiency of diagnosis. The utilisation of neuroimaging data, genetics, neuropsychology and measures such as EEG is becoming increasingly prevalent in academic research (Shatte et al. 2019). The results of these studies demonstrate that AI tools are capable of accurately diagnosing psychiatric and psychological disorders (Abd-Alzaraq et al. 2023). Nevertheless, some researchers have highlighted the potential for prediction models to impact decision-making autonomy, citing the expansive clinical definitions and subjective judgments that underpin them (Su et al. 2020). It is therefore recommended that clinical decisions be informed by a combination of expertise and AI models. The most valuable application of AI is the development of decision-support tools for differential diagnosis (Dwyer et al. 2018). A diagnostic support system for differentiating depression and anxiety disorders employs a cognitive test battery to assess these disorders using AI, achieving a classification accuracy of 66-80%. It is anticipated that this objective tool will be employed in conjunction with clinical interviews, thereby assisting clinicians in formulating more precise diagnoses (Richter et al. 2021). Moreover, an AI-based signature has been developed for differentiating between depression and schizophrenia. This signature may assist in the evaluation of samples with ambiguous diagnoses, such as psychosis and high-risk conditions (Koutsouleris 2018). A digital application, designated EarlyDetect, has been developed for the purpose of screening for risk factors associated with psychiatric and psychological disorders. This application can be utilised in mental health screenings within tertiary healthcare facilities, thereby facilitating the early detection of conditions such as bipolar disorder (Liu et al. 2021a, Liu et al. 2021b).

AI offers analytical tools that can facilitate the comprehension of psychopathology. The sophisticated classification systems inherent to these approaches have significant implications for the development and implementation of targeted interventions (Cuthbert 2019). The derivation of client subgroups from data according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) and the International Classification of Diseases (ICD-11) suggests that treatment outcomes can be more accurately predicted (Dwyer et al. 2018). The RDoC (Research Domain Criteria) project offers a complementary approach to the understanding of psychiatric and psychological disorders, providing behavioural and neurobiological measures consistent with those employed in AI approaches. AI clustering methods were employed to identify patient subgroups with early-stage schizophrenia. It is asserted that these subgroups are based on cognitive performance and neuroanatomical patterns. An early diagnosis allows for the implementation of appropriate interventions for these subgroups (Wenzel et al. 2021). AI methods can be applied to data obtained from wearable devices to assess the severity of depression and to inform decisions regarding screening and treatment selection (Ahmed et al. 2022). The Mood Assessment Framework (Moodable) represents a significant advance in population-level mental health screening, offering immediate depression and suicide risk assessment based on retrospectively collected smartphone data (Dogrucu et al. 2020).

The advent of digital phenotyping has facilitated the acquisition of comprehensive, real-time, continuous, and non-invasive data about clients' daily lives. Information can be gathered from the subject using smartphones and wearable devices, encompassing self-report data and behavioural and physiological data. Such data may be

monitored remotely and supplemented with real-time feedback. AI-driven tools can inform diagnostic and treatment decisions based on the aforementioned data. The utilisation of smartphones and wearable sensors can facilitate access to a plethora of real-time data that is typically beyond the reach of clinicians. The data may be collected continuously and non-invasively and may include a range of indicators, such as sleep patterns, exercise routines, stress levels, geographic location, cell phone use, social interaction, fatigue, and mood. AI-driven diagnostic tools have the potential to integrate this data with traditional sources, thus facilitating more accurate and comprehensive diagnoses (Roberts et al. 2018). The application of digital phenotyping methodologies provides novel opportunities for the surveillance of patients within the mental health field. Transformations in movement or mobile usage have the potential to reflect mood changes that may serve as pre-symptoms of an impending crisis (Meyerhoff et al. 2021). Moreover, digital phenotyping can facilitate the monitoring of medication adherence and the identification of efficacious pharmacological agents. The presentation of behavioural data can assist clinicians in guiding clients and providing feedback that increases motivation, thus enabling the setting of goals and encouraging engagement in the treatment process. Despite the growing importance of digital phenotyping in mental health, there remains a paucity of validated and applicable e-tools (Roberts et al. 2018). Wearable AI technologies have the potential to detect anxiety and depression, however, they are not yet sufficiently developed for clinical use (Abd-Alzaraq et al. 2023).

The assertion that digital data can be objective and unbiased is challenged by some experts, who emphasise the necessity of addressing the ethical and practical challenges associated with digital phenotyping (Birk and Samuel 2022, Moura et al. 2023). The passive collection of smartphone data has the potential to facilitate the monitoring of schizophrenia patients, identify warning signs of relapse in real time, provide support for early intervention, and improve client outcomes (Barnett et al. 2018). It is feasible to evaluate the severity of depression through the use of wearable devices. The use of sensors in smartphones and wearable devices for mobile sensing can infer behavioural signs of depression and provide a potential objective method for assessing depression severity (Lamichhane et al. 2024).

Artificial Intelligence and Prognosis

The advancement of AI presents a substantial opportunity for anticipating client outcomes and administering tailored preventive treatments. This progress offers the potential for the development of AI-driven prognostic tools capable of anticipating a range of individual outcomes, including symptom severity, changes in daily functioning and quality of life, the occurrence of relapses or remissions, and transitions (for example, from psychosis to remission). These developments facilitate the creation of prognostic tools and treatments designed to predict individual outcomes. AI has been demonstrated to be capable of predicting cognitive impairment and the development of Alzheimer's disease (Arbabshirani et al. 2017). The PRONIA project is a multidisciplinary European study that applies AI to clinical, cognitive, and neuroimaging data on patients with new-onset psychosis. The project's findings demonstrate that the utilisation of multimodal data and AI facilitates the prediction of treatment outcomes and the optimisation of individualised treatment decisions (Koutsouleris et al. 2021). Moreover, an AI-based tool for forecasting the functional consequences of psychotic episodes was devised and validated using routine client data (Koutsouleris et al. 2018). Recent studies have extended this model to include environmental factors. Moreover, the incorporation of AI models that predict the transition to psychosis by integrating clinical and biological data has notably reduced the false negative rate of clinical predictions, which has been shown to exceed 15% (Koutsouleris et al. 2021).

Additionally, AI has the potential to serve as a valuable tool in suicide prevention initiatives. The preliminary findings of the research indicate that AI-based tools demonstrate superior performance compared to existing models. These tools have been demonstrated to effectively identify individuals who are at high risk of suicide (Roy et al. 2020). Notably, neural network models have demonstrated efficacy in detecting suicide risk through the analysis of publicly accessible social media data. Additionally, AI can assist clinicians in making treatment decisions for individuals with suicidal tendencies. The capacity for continuous risk assessment and real-time monitoring enables these tools to facilitate rapid intervention for individuals at risk.

Artificial Intelligence and the Psychotherapy Process

The implementation of individualized treatment plans has the potential to enhance the selection of appropriate treatment modalities based on the unique characteristics and needs of the individual, as opposed to the use of group-specific treatment approaches. This can optimise the treatment process and ensure the most efficient use of resources. The application of AI can enhance the process of selecting appropriate treatments by facilitating a deeper comprehension of the intricate interrelationships between biological, genetic, behavioural, personality,

social and environmental factors (Bzdok and Meyer 2018). A clinical decision tool should incorporate factors such as risks, side effects, treatment resistance, time, and cost (Dwyer et al. 2018). The application of AI to the field of medicine has demonstrated the potential of big data sources to inform treatment decisions. Nevertheless, further research is required. In their study, Sajjardian et al. (2021) investigated the reproducibility of results produced by AI. A further study assessing the potential of AI applications in the treatment of depression revealed that studies of superior quality exhibited reduced accuracy, and the findings were not independently replicated. Furthermore, it has been demonstrated that the utilisation of AI, based on an array of genetic, clinical and demographic factors, can facilitate the attainment of more accurate results in the selection of antidepressant medications (Taliaz et al. 2021). In order to provide therapists with additional support in their clinical decision-making, a combination of AI and traditional techniques was employed in order to develop a treatment selection algorithm. This algorithm recommends either CBT or psychodynamic therapy, based on the pre-treatment characteristics of the patient (Schwartz et al. 2021). It has been demonstrated that the efficacy of cognitive behavioural therapy for social anxiety disorder can be predicted (Ashar et al. 2021). These studies illustrate the potential applications of AI in treatment decision-making, yet further research and evaluation are required.

Chatbots are AI-powered tools that can encourage individuals to seek assistance for psychiatric and psychological disorders by offering anonymous and impartial support. Additionally, they can facilitate self-care and help address the increasing demand for mental health services. This could be of particular benefit to younger people and those who are difficult to reach. Chatbots are tools that employ machine learning and AI to simulate human communication. Such applications can be accessed via voice or text and are compatible with mobile devices, computers, and smart speakers. Examples of such chatbots include Tess (a mental health chatbot that provides on-demand, affordable and quality emotional support to improve mental health and resilience), and Wysa (a chatbot that responds to expressed emotions and uses evidence-based CBT, dialectical behaviour therapy (DBT), meditation, breathing, and yoga). Many chatbots are currently being used in the delivery of mental health services, with the objective of reducing psychological symptoms or improving wellbeing by increasing self-awareness (Bendig et al.). This is achieved through the use of AI-based services such as Yoga, which employs motivational interviewing and micro-actions to help users improve their mental resilience skills and feel better. (Bendig et al. 2022). Furthermore, they have the potential to provide additional therapeutic interventions (Bendig et al. 2022).

Chatbots offer a number of advantages in the field of mental health. Chatbots offer significant benefits, including 24/7 accessibility, the provision of support beyond the scope of traditional face-to-face sessions, and the ability to respond to the growing demand for accessible mental health services. Moreover, chatbots have the potential to mitigate feelings of stigma and enhance motivation to seek assistance by providing anonymous and non-judgemental support. Research has indicated that some individuals may be more at ease conversing with a computer than with a clinician (Lucas et al. 2017). Chatbots may prove to be a particularly valuable source of support for groups such as young people and those who are difficult to reach, given that technology offers greater acceptance and convenience.

Moreover, chatbots have the potential to assist operators of mental health services. In a hybrid model, however, a chatbot service is combined with a real-life service, with the aim of improving quality and reducing service time while also minimising some of the risks associated with an automated system. The extant evidence on chatbots is inconclusive. A meta-analytic review revealed that an insufficient number of studies, risk of bias and conflicting findings represent some of the main limitations (Abd-Alzaraq et al. 2020). Nevertheless, some favourable outcomes have been observed with regard to personalisation, empathic responses and the efficacy of long-term interaction (He et al. 2023). The user experience of chatbots has been found to be largely associated with the quality of the personal therapy relationship, content engagement, and effective communication (Li et al. 2023). However, technical issues include the misinterpretation of emotional expressions, while ethical issues include concerns pertaining to privacy and data security (Coghlan et al. 2023).

Additionally, there is a potential for an overreliance on chatbots, which may ultimately diminish the significance of human interaction. The use of chatbots has the potential to create a blurring of the boundaries between reality and fiction. This is of particular concern for groups susceptible to emotional transference or other negative outcomes, as evidenced by Fiske et al. (2019). It is crucial to implement human oversight and regulation of chatbot interactions as a strategy to enhance the efficacy of chatbots. In the context of mental health, there are several practical examples that illustrate this point. In a study conducted by Shah et al. (2022), a chatbot was developed and evaluated in conjunction with online eating disorder treatment, with the aim of enhancing treatment motivation. The chatbot prototype, Alex, was enhanced through the incorporation of user feedback,

and participants provided favourable responses. The study conducted by Shah et al. (2022) offers preliminary evidence of the potential of chatbots to facilitate the utilisation of services among individuals with eating disorders. In a study conducted by Fitzpatrick et al. (2017), a self-help program called Woebot, a text-based conversational agent, was employed to reduce symptoms of depression and anxiety among college students, with successful results. The objective of this study was to ascertain the viability, acceptability and efficacy of utilising a conversational agent to facilitate the delivery of a self-help programme for university students. The results indicate that conversational agents are an effective method of delivering self-help programmes (Fitzpatrick et al. 2017). Giguère et al. (2023) demonstrated the short-term efficacy of an avatar intervention for the treatment of cannabis use disorder. The intervention is delivered via virtual reality techniques and motivational interviewing, thereby enabling participants to practise these techniques in real time. The aforementioned examples illustrate the efficacy of chatbots in the context of mental health.

Electronic health records (EHRs) can be employed to construct predictive models for suicide risk (Su et al. 2020). The utilisation of EHRs in the development of clinical risk prediction models is becoming increasingly prevalent in clinical practice. In light of the recent surge in suicide rates, it has become imperative to gain insight into, anticipate, and avert suicide risk. Bayramlı et al. (2022) conducted a comparative analysis of the predictive efficacy of structured and unstructured EHR data in predicting suicide risk with the aid of electronic health records and clinical risk prediction models. Their findings revealed that models trained on structured data, including the Nai and Bayes Classifier and Random Forest models, demonstrated superior performance compared to their unstructured counterparts. The identification of feature pairs and utilisation of this framework has the potential to enhance clinical modelling endeavours.

The application of machine learning (ML) models facilitates the identification of high-risk individuals within vulnerable communities (Rozek et al. 2020). Mental health is a state of emotional, psychological, and social well-being, influencing our thoughts, feelings, and behaviour. The advent of the SARS-CoV-2 pandemic has precipitated a surge in demand for mental health services, with the emergence of a plethora of novel psychological disorders associated with the pandemic. Nevertheless, a significant proportion of the population is unable to access timely assistance due to financial constraints and a shortage of available therapists. Attempts are being made to enhance the prediction, diagnosis and treatment of psychological disorders through the utilisation of machine learning. This approach has the potential to facilitate the delivery of more precise and personalised mental health services. Furthermore, machine learning is being employed in other domains of psychology. For instance, it is employed in the context of recruitment and job interviews within organisational settings, facilitating the identification of personality types. Furthermore, machine learning is employed in the field of cognitive science with the objective of enhancing comprehension of mental processes (Fitzpatrick et al. 2017).

The subsequent stages of this initiative will entail the formulation of clinical directives and the undertaking of external verification. In a recent study, an NLP-based system was trained and deployed with the objective of improving response times to callers in crisis accessing a national crisis helpline (Swaminathan et al. 2023). Suicidal Ideation Detection Visual Interactive Systems (SIDVis) is an interactive visual dashboard designed to detect suicidal ideation from social media data and provide proactive interventions and support. Suicide is a global phenomenon that necessitates an examination of the underlying factors, including mental illness, substance abuse, financial stress, and trauma. To address this problem, it is essential to identify individuals at risk and implement preventative measures. Nevertheless, the detection of suicidal thoughts can prove challenging. It would appear that research in this field has not devoted sufficient attention to the development of visually interpretable and interactive systems. Islam et al. (2023) enhance the detection of suicidal thoughts through the utilisation of deep learning and machine learning techniques, thereby contributing to the development of responsible AI. To this end, an interactive visualisation system, designated SIDVis, is employed. The comprehensive assessment demonstrates that SIDVis exhibits superior performance compared to existing methodologies and holds promise for enhancing suicidal ideation and intervention strategies (Islam et al. 2023). It is hoped that these findings will contribute to improvements in suicide prevention and assistance services.

The implementation of AI-based practice management tools has the potential to enhance the quality of service provided, boost productivity, and curtail the financial outlay associated with clinic management. Otter.ai is an automated tool designed to facilitate the generation of meeting notes and transcription of real-time communications. Heidi is a computerised medical printer. PatientNotes is an automated clinical note-taking tool. Such tools facilitate the streamlining of administrative processes, including those pertaining to appointment scheduling, billing, and electronic health records. This enables clinicians to devote more attention to direct patient care and complex tasks (CoieraveLiu 2022).

Moreover, large language models (LLMs) can be utilised on digital printers to generate summaries of health records by listening to clinicians' speech. AI tools have the capacity to incorporate the objectives of clients and pertinent elements from preceding sessions into future treatment plans, thereby facilitating the continuity of care. Such tools have the potential to benefit both clients and clinicians alike, facilitating enhanced decision-making processes. The field of machine learning and AI has seen rapid evolution, with the development of numerous AI-based practice management tools that are now widely available in clinical practice. AI tools have the capacity to import pertinent elements from preceding sessions, including assignments, assessments, and clients' goals, into future treatment plans and subsequent documentation. This can facilitate the maintenance of treatment continuity, ensure the accuracy of treatment fidelity, and enhance the organisation of sessions and agenda setting (Biswas and Talukdar 2024). There are now accessible and user-friendly tools, such as Dovetail, which facilitate the recording of clients' engagement in interventions and their reactions to clinical presentations and communication. Such tools are capable of generating and analysing session feedback summaries based on verbal data obtained from sessions, thereby facilitating self-reflection and the potential for clinical practice adjustments among clinicians.

Notable examples of such tools include Clinicalnotes.ai, which has been developed as a platform with the capacity to comprehend the clinical context and to personalise automated documentation. Dovetail is a tool designed to facilitate the conversion of data into actionable insights, thereby streamlining the process of turning data into work. Additionally, WhyHive is a user-friendly natural language processing (NLP) data analysis tool. AI-based tools are capable of analysing intricate patterns within client data, facilitating the dissemination of novel and innovative clinical research on a range of topics, including interventions, therapeutic modalities, group programmes and more.

Artificial Intelligence and Empathy

Following an analysis of 43 definitions, Cuff et al. (2016) defined empathy as "an emotional response resulting from the interaction between an individual's trait capacities and the situational context." Although empathic processes occur spontaneously, they are also subject to control processes. Empathy manifests in three distinct forms: emotional, cognitive, and compassionate empathy. Emotional empathy refers to the capacity to reflect the emotional states of others, whereas cognitive empathy pertains to the ability to comprehend these emotional states. These various forms of empathy have been the subject of investigation with a view to enhancing empathy between individuals (Raamkumar and Yang 2022). Computational modelling of empathy represents a valuable tool for enhancing our comprehension of human relationships. Empathy modelling studies have employed computational and theoretical approaches to investigate emotional communication competence, emotion regulation, and cognitive mechanisms. The specific manifestations of empathic behaviours are contingent upon a number of factors, including projection, emotional matching, empathic concern, consolation, altruistic assistance, and perspective-taking (Teding van Berkhout and Malouff 2016).

There is a growing interest in the field of AI in the development of empathetic AI systems (Raamkumar and Yang 2022). However, empathy is frequently inadequately represented through proxy data such as facial expressions, voice signals, and gestures, without sufficient consideration of the multifaceted nature of empathy and its subjective conceptualisation (Stark and Hoey 2021). Such systems may result in the formation of biased judgments due to the exclusion of the broader social context surrounding the issue in question. This limitation can be attributed to a lack of comprehension of the interdependence between empathy and accountability. Despite the substantial body of research on empathy and responsibility in AI, these concepts have frequently been examined in isolation, without fully acknowledging their intrinsic interdependence. A growing body of research in the social sciences, organisation studies and psychology suggests that greater accountability can be achieved by regulating empathy (Blader and Rothman 2014).

The regulation of empathy acknowledges the existence of individual differences in emotional responses and perspectives, as well as the influence of broader social contexts on any given problem. This indicates that data derived from non-specific sources, such as facial expressions and gestures, are inadequate indicators of empathy. It is therefore necessary to elucidate the function of empathy in the context of accountability in AI. The incorporation of empathy facilitates the realisation that accountability is inherently a social-technical challenge, rather than a purely technical one. It is hypothesised that the integration of empathy into the design of AI systems could enhance accountability and address some of the challenges associated with AI accountability (Srinivasan and González 2022).

Ethical and Legal Issues

A plethora of mental health applications are currently available for a variety of purposes. These practices are designed to facilitate a range of outcomes, including emotional regulation, enhancement of well-being, cultivation of mindfulness, and monitoring of mood. However, the level of regulation of these practices is low, and the efficacy of not all mental health practices is established. Such practices include monitoring, mindfulness and meditation exercises, CBT-based treatments, and peer support. Mental health apps can facilitate access to mental health services, particularly for individuals lacking the requisite resources or ability to engage with a therapist. The majority of users have indicated that they find these apps to be useful. The factors influencing user satisfaction include usability, personalised content, privacy and security, perceived effectiveness, and client support (Oyebode et al. 2020). In the context of a traditional therapeutic relationship, there are clear ethical obligations to protect the client. However, in the case of online mental health applications, these obligations are less clear. This gives rise to concerns regarding safety and privacy (Martinez and Kreitmair 2018). Furthermore, there is a paucity of evidence regarding the efficacy of mental health practices (Eisenstadt et al. 2021). In a study of 28 popular mental health apps, only five were found to have empirical support (Wang et al. 2020). Refresh is an application-based, non-directive sleep intervention. The WeClick app is designed to enhance well-being and encourage help-seeking behaviours among Australian adolescents by fostering positive relationships.

The objective of the Made4Me Program is to furnish users with information and assistance in managing symptoms through a self-paced course based on CBT. Furthermore, users are able to access therapist support via digital platforms. It is incumbent upon clinicians to discuss with their clients the reliability, features, cost, and confidentiality of the app in question. It is similarly vital to ensure the provision of the requisite support and ongoing guidance. In the absence of regulatory oversight, it is imperative to address ethical considerations such as accountability and the protection of individuals undergoing therapy. Further research and regulatory measures are required to enhance the efficacy and reach of online mental health applications.

AI may assist therapists in achieving a balance between the need to treat a greater number of individuals and the desire to spend less time with each of them. It is essential to monitor this carefully in order to ensure the provision of effective psychotherapy without compromising the quality of the client-therapist relationship. From a perspective of justice and care, the incorporation of AI into psychotherapy should be guided by the objective of fulfilling the right to autonomy and meeting the needs of vulnerable individuals. Nevertheless, it is imperative to consider care principles such as constant supervision, reassurance, prevention of risk of harm, and empathy. Let us posit that AI is integrated into psychotherapy in a manner that does not infringe upon the autonomy of the individual. In such a scenario, the objective would be to provide suitable therapy for all individuals experiencing mental health issues, while upholding the principle of equality with vulnerable groups. This supplementary methodology can facilitate the assurance of ethical and compassionate conduct with regard to AI. In instances where psychotherapy is not readily accessible, AI tools can facilitate increased access and reinforce the quality of care relationships. It is imperative that AI tools are designed in a manner that precludes any form of discriminatory practice and is responsive to the diverse needs of users (Roa et al. 2021).

The utilisation of AI in the field of mental healthcare presents a duality of advantages and ethical and social challenges. From a moral standpoint, embodied AI offers a number of advantages, including the potential to provide novel treatment modalities, engage populations that are difficult to reach, elicit more positive patient responses, and free up clinician time. However, there are also overarching ethical issues and concerns that require consideration, including the prevention of harm, questions of data ethics, the potential for the misuse of technologies, and the risk of exacerbating health inequalities. Furthermore, specific challenges must be addressed, including the risk assessment of embodied AI applications, the respect of patient autonomy, the transparency of algorithms, and the long-term impacts (Fiske et al. 2019).

Future Perspectives

Furthermore, AI has the potential to be a transformative force in the field of mental health. The application of AI to mental health has the potential to effect transformative change in a number of areas, including the early diagnosis of mental health issues, the creation of personalised treatment plans and the development of AI-powered virtual therapists. However, these advances also give rise to a number of ethical challenges, including those related to privacy, the mitigation of bias, and the role of the human element. In order to ensure the responsible development of AI in mental health, it is essential to prioritise work in key areas such as the establishment of transparent regulatory frameworks, the validation of AI models and the continuous advancement of research and development. The integration of AI into the field of mental health is a promising area of research, but the challenges associated with this endeavour are also multifaceted. It is of the utmost

importance to establish robust regulatory frameworks, ensure model validation and transparency, and invest in continuous research and development. It is imperative that these measures be implemented so that AI can play a pivotal role in enhancing the quality of mental health services and ensuring greater accessibility, practicality, and ethicality. It is similarly crucial to validate and ensure transparency regarding the AI models employed in clinical settings (Olawade et al. 2024).

The potential of AI models to support and automate psychotherapy is considerable, however, their use in a field such as clinical psychology is fraught with risk. Large language models represent a technology with the potential to facilitate and, in certain instances, even supplant psychotherapy. Nevertheless, concerns have been raised regarding the potential for harm, given the current lack of oversight and understanding of their use in the clinical field. Furthermore, it is imperative that the utilisation of large language models in psychotherapy is meticulously supervised and employed in a responsible and ethical manner, with the objective of safeguarding the well-being of patients (Stade et al. 2024).

Conclusion

AI therapy is a service that employs sophisticated and comprehensive data to assist individuals on their mental health journey. The provision of AI therapy is frequently conducted via online platforms and mobile applications, thereby enhancing the accessibility of mental health assistance. The majority of AI therapy programmes are based on CBT or other established therapeutic approaches. The availability of AI-based mental health services is not constrained by geographical or temporal boundaries. These services are accessible 24 hours a day, seven days a week, offering counselling support. The AI therapy system analyses the input provided by the user and offers supportive, informative, and therapeutic responses. In this way, it provides a service that is not feasible for therapists in practice. The use of AI in mental health therapy provides a convenient and accessible avenue for individuals to obtain assistance, leveraging the advantages of modern technology. For individuals who are unable to access mental health services and are reluctant to engage in face-to-face psychotherapy, AI-based psychotherapy may offer a viable alternative (Bickman 2020). The application of AI has the potential to enhance the accuracy and efficacy of psychiatric and psychological disorder diagnosis and treatment. Furthermore, AI-based psychotherapy can facilitate remote access.

As a result of advances in AI, the utilisation of biological, demographic, clinical, and other forms of data can assume a significant role in the diagnosis, prognosis, and treatment of individuals with psychiatric and psychological disorders. Such an approach could facilitate more accurate assessment, stratification, and assignment of clients to the most appropriate treatment. Furthermore, this approach could enable therapists to utilise their time more efficiently, thereby allowing them to serve a greater number of clients. This could result in a reduction in waiting times and an improvement in clinical outcomes. It is therefore hypothesised that this novel approach may yield individual, economic and social benefits. The role of AI in therapeutic applications within the field of mental health services is becoming increasingly significant. While AI offers cost-effectiveness and accessibility, it may also pose a risk to human well-being and trust. AI may play a significant role in the diagnosis of psychiatric and psychological disorders, particularly for certain client groups.

Consequently, AI is rapidly becoming an effective tool in fulfilling numerous tasks in healthcare settings that were previously regarded as the domain of humans (Minerva and Giubilini 2023). Nevertheless, it is not anticipated that AI will entirely supplant the role of human therapists. AI offers a cost-effective solution within the mental health system, providing support to those who would otherwise be unable to access traditional forms of treatment. Furthermore, developments in tools and technology can facilitate enhancements and reinforce the provision of client care, although they are not intended to supplant the role of human mental health professionals. In light of the growing importance of AI in mental health services, it is imperative that counsellors possess a comprehensive understanding of the AI tools that are becoming increasingly prevalent in this field. Such innovations have the potential to enhance the standard and accessibility of mental health services by augmenting existing practices to provide more effective support to clients. Although the utilisation of AI in psychotherapy presents advantages in terms of accessibility and efficacy, it is imperative to acknowledge the constraints of algorithmic decision-making and to adhere to the ethical boundaries that govern such practices. It is essential to maintain a human touch while leveraging the advantages of technology.

References

Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M (2020) Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res, 22:e16021.

Abd-Alrazaq A, AlSaad R, Harfouche M, Aziz S, Ahmed A, Damseh R. et al. (2023) Wearable artificial intelligence for detecting anxiety: systematic review and meta-analysis. J Med Internet Res, 25:e48754.

Ahmed A, Ramesh J, Ganguly S, Aburukba R, Sagahyroon A, Aloul F (2022) Investigating the feasibility of assessing depression severity and valence-arousal with wearable sensors using discrete wavelet transforms and machine learning. Information, 13:406.

Andrews G, Issakidis C, Carter G. (2001) Shortfall in mental health service utilisation. Br J Psychiatry, 179:417-425.

Arbabshirani MR, Plis S, Sui J, Calhoun VD (2017) Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Neuroimage, 145:137-165.

Ashar YK, Clark J, Gunning FM, Goldin P, Gross JJ, Wager TD (2021) Brain markers predicting response to cognitive-behavioral therapy for social anxiety disorder: an independent replication of Whitfield-Gabrieli et al. 2015. Transl Psychiatry, 11:260.

Bain EE, Shafner L, Walling DP, Othman AA, Chuang-Stein C, Hinkle J et al. (2017) Use of a novel artificial intelligence platform on mobile devices to assess dosing compliance in a phase 2 clinical trial in subjects with schizophrenia. JMIR Mhealth Uhealth, 5:e18.

Bălan O, Moise G, Moldoveanu A, Leordeanu M, Moldoveanu F (2020) An investigation of various machine and deep learning techniques applied in automatic fear level detection and acrophobia virtual therapy. Sensors, 20:496.

Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela, JP (2018) Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology, 43:1660-1666.

Bayramlı I, Castro V, Barak-Corren Y, Madsen EM, Nock MK, Smoller JW et al. (2022) Predictive structured-unstructured interactions in EHR models: A case study of suicide prediction. NPJ Digit Med, 5:15.

Bendig E, Erb B, Schulze-Thuesing L, Baumeister H (2022) The next generation: chatbots in clinical psychology and psychotherapy to foster mental health–a scoping review. Verhaltenstherapie, 32:64-76.

Bickman L (2020) Improving mental health services: A 50-year journey from randomized experiments to artificial intelligence and precision mental health. Adm Policy Ment Health, 47:795-843.

Birk RH, Samuel G (2022) Digital phenotyping for mental health: reviewing the challenges of using data to monitor and predict mental health problems. Curr Psychiatry Rep, 24:523-528.

Biswas A, Talukdar W (2024) Intelligent clinical documentation: harnessing generative AI for patientcentric clinical note generation. Int J Innov Sci Res Technol, 9:994-1008.

Blader SL, Rothman NB (2014) Paving the road to preferential treatment with good intentions: Empathy, accountability and fairness. J Exp Soc Psychol, 50:65-81.

Braga A, Logan RK (2017) The emperor of strong AI has no clothes: limits to artificial intelligence. Information, 8:156.

Bzdok D, Meyer-Lindenberg A (2018) Machine learning for precision psychiatry: opportunities and challenges. Biol Psychiatry Cogn Neurosci Neuroimaging, 3:223-230.

Calderita LV, Manso LJ, Bustos P, Suárez CM, Fernández F, Bandera A (2014) Therapist: towards an autonomous socially interactive robot for motor and neurorehabilitation therapies for children. JMIR Rehabil Assist Technol, 1:e1.

Chekroud AM, Hawrilenko M, Loho H, Bondar J, Gueorguieva R, Hasan A et al. (2024) Illusory generalizability of clinical prediction models. Sci, 383:164-167.

Chen J, Li Y, Wu X, Liang Y, Jha S (2020) Robust out-of-distribution detection for neural networks. arXiv:2003.09711v6.

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. Digit Health, 9:20552076231183542.

Coiera E, Liu S (2022) Evidence synthesis, digital scribes, and translational challenges for artificial intelligence in healthcare. Cell Rep Med, 3:100860.

Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E et al. (2018) AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry, 5:31-40.

Crawford K (2021) Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Connecticut, Yale University Press.

Cuff BM, Brown SJ, Taylor L, Howat DJ (2016) Empathy: a review of the concept. Emotion, 8:144-153.

Cuthbert BN (2020) The role of RDoC in future classification of mental disorders. Dialogues Clin Neurosci, 22:81-85.

D'Alfonso S (2020) AI in mental health. Curr Opin Psychol, 36:112–117.

Dellazizzo L, Percie du Sert O, Phraxayavong K, Potvin S, O'Connor K, Dumais A (2018) Exploration of the dialogue components in Avatar Therapy for schizophrenia patients with refractory auditory hallucinations: a content analysis. Clin Psychol Psychother, 25:878-885.

Dogrucu A, Perucic A, Isaro A, Ball D, Toto E, Rundensteiner EA et al. (2020) Moodable: on feasibility of instantaneous depression assessment using machine learning on voice samples artificial intelligence, machine learning and mental healthcare 29 with retrospectively harvested smartphone and social media data. Smart Health, 17:100118.

Dong Y, Hou J, Zhang N, Zhang M (2020) Research on how human intelligence, consciousness, and cognitive computing affect the development of Artificial Intelligence. Complexity, 2020:1680845.

du Sert, OP, Potvin S, Lipp O, Dellazizzo L, Laurelli M, Breton R et al. (2018) Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: a pilot clinical trial. Schizophr Res, 97:176-181.

Dwyer DB, Falkai P, Koutsouleris N (2018) Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol, 14:91-118.

Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C (2021) Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health, 8:e31170.

Farahany NA (2023) The Battle for Your Brain. New York, St. Martin's Press.

Fiske A, Henningsen P, Buyx A (2019) Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res, 21:e13216.

Fitzpatrick KK, 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 Ment Health, 4:e19.

Freeman D, Haselton P, Freeman J, Spanlang B, Kishore S, Albery E et al. (2018) Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. Lancet Psychiatry, 5:625-632.

Giguère S, Potvin S, Beaudoin M, Dellazizzo L, Giguère CÉ, Furtos A et al. (2023) Avatar intervention for cannabis use disorder in ındividuals with severe mental disorders: A pilot study. J Pers Med, 13:766.

Góngora SA, Hamrioui S, de la Torre Díez, I., Motta Cruz E, López MC, Franco M (2019) Social robots for people with aging and dementia: A systematic review of literature. Telemed J E Health, 25:533-540.

Gual-Montolio P, Jaén I, Martínez-Borba V, Castilla D, Suso-Ribera C (2022) Using artificial intelligence to enhance ongoing psychological interventions for emotional problems in real or close to real-time: A systematic review. Int J Environ Res Public Health, 19:7737.

He Y, Yang L, Qian C, Li T, Su Z, Zhang Q et al. (2023) Conversational agent interventions for mental health problems: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res, 25:e43862.

Holohan M, Fiske A (2021) Like I'm talking to a real person: exploring the meaning of transference for the use and design of AI-based applications in psychotherapy. Front Psychol, 12:720476.

Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth, 6:e12106.

Islam MR, Sakib MKH, Ulhaq A, Akter S, Zhou J, Asirvathamt D (2023) Sidvis: designing visual interactive system for analyzing suicide ideation detection. 2023 27th International Conference Information Visualisation (IV), Tampere, Finland, 2023, 384-389.

Kessler RC, Aguilar-Gaxiola S, Alonso J, Chatterji S, Lee S, Ormel J et al. (2009) The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. Epidemiol Psychiatr Sci, 18:23-33.

Koutsouleris N, Dwyer DB, Degenhardt F, Maj C, Urquijo-Castro MF, Sanfelici R et al. (2021) Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. JAMA Psychiatry, 78:195-209.

Koutsouleris N, Hauser TU, Skvortsova V, De Choudhury M. (2022) From promise to practice: Towards the realisation of AI-informed mental health care. Lancet, 4:e829-e840.

Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann, S, Rosen M, Ruef A, Dwyer DB et al. (2018) Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. JAMA Psychiatry, 75:1156-1172

Lambert MJ (2004) Bergin and Garfield's Handbook of Psychotherapy and Behavior Change. New York, Wiley.

Lamichhane B, Moukaddam N, Sabharwal A. (2024) Mobile sensing-based depression severity assessment in participants with heterogeneous mental health conditions. Sci Rep, 14:18808.

Li H, Zhang R, Lee YC, Kraut RE, Mohr DC (2023) Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and wellbeing. NPJ Digit Med, 6:236-236.

Liu YS, Chokka S, Cao B, Chokka PR (2021a) Screening for bipolar disorder in a tertiary mental health centre using EarlyDetect: a machine learning-based pilot study. J Affect Disord Rep, 6:100215.

Liu Y, Hankey J, Cao B, Chokka P (2021b) Screening for major depressive disorder in a tertiary mental health centre using EarlyDetect: a machine learning-based pilot study. J Affect Disord Rep, 3:100062.

Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J et al. (2017) Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. Front Robot AI, 4:51.

Luxton DD (2014) Artificial intelligence in psychological practice: current and future applications an implications. Prof Psychol Res Pr, 45:332-339.

Martin NM, Kreitmair K (2018) Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. JMIR Ment Health, 5:e32.

Martinez-Martin N, Kreitmair K (2018) Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. JMIR Ment Health, 5:e9423.

Mehrotra S, Kumar S, Sudhir P, Rao GN, Thirthalli J, Gandotra A (2017) Unguided mental health self-help apps: reflections on challenges through a clinician's lens. Indian J Psychol Med, 39:707-711.

Meyerhoff J, Liu T, Kording KP, Ungar LH, Kaiser SM, Karr CJ et al. (2021) Evaluation of changes in depression, anxiety, and social anxiety using smartphone sensor features: longitudinal cohort study. J Med Internet Res, 23:e22844.

Minerva F, Giubilini A (2023) Is AI the future of mental healthcare? Topoi, 42:809-817.

Moura I, Teles A, Viana D, Marques J, Coutinho L, Silva F (2023) Digital phenotyping of mental health using multimodal sensing of multiple situations of interest: a systematic literature review. J Biomed Inform, 138:104278.

Nathan PE, Gorman JM (2015) A Guide to Treatments That Work. Oxford, Oxford University Press.

Nemesure MD, Heinz MV, Huang R, Jacobson NC (2021) Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. Sci Rep, 11:1980.

Olawade DB, Wada OZ, Odetayo A, David-Olawade AC, Asaolu F, Eberhardt J (2024) Enhancing mental health with artificial intelligence: current trends and future prospects. J Med Surg Public Health, 3:100099.

Oyebode O, Alqahtani F, Orji R (2020) Using machine learning and thematic analysis methods to evaluate mental health apps based on user reviews. IEEE Access, 8:111141-111158.

Raamkumar AS, Yang Y (2022) Empathetic conversational systems: a review of current advances, gaps, and opportunities. IEEE Trans Affect Comput, 14:2722-2739.

Reece AG, Danforth CM (2017) Instagram photos reveal predictive markers of depression. EPJ Data Sci, 6:15.

Rein BA, McNeil DW, Hayes AR, Hawkins TA, Ng H. Mei, Yura CA (2018) Evaluation of an avatar-based training program to promote suicide prevention awareness in a college setting. J Am Coll Health, 66:401-411.

Richards D (2024) Artificial intelligence and psychotherapy: a counterpoint. Couns Psychother Res, https://doi.org/10.1002/capr.12758

Richter T, Fishbain B, Fruchter E, Richter-Levin G, Okon-Singer H (2021) Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders. J Psychiatr Res, 141:199-205.

Roa TM, Biller-Andorno N, Trachsel M (2021) The ethics of artificial intelligence in psychotherapy. In The Oxford Handbook of Psychotherapy Ethics. (Eds M Trachsel, J Gaab, N Biller-Andorno, Ş Tekin, JZ Sadler):744-758. Oxford, Oxford University Press.

Roberts LW, Chan S, Torous J (2018) New tests, new tools: mobile and connected technologies in advancing psychiatric diagnosis. NPJ Digit Med, 1:20176.

Rosellini AJ, Liu S, Anderson GN, Sbi S, Tung ES, Knyazhanskaya E (2020) Developing algorithms to predict adult onset internalizing disorders: an ensemble learning approach. J Psychiatr Res, 121:189-196.

Roy A, Nikolitch K, McGinn R, Jinah S, Klement W, Kaminsky ZA (2020) A machine learning approach predicts future risk to suicidal ideation from social media data. NPJ Digit Med, 3:1-12.

Rozek DC, Andres WC, Smith NB, Leifker FR, Arne K, Jennings G et al. (2020) Using machine learning to predict suicide attempts in military personnel. Psychiatry Res, 294:113515.

Sachan D (2018) Self-help robots drive blues away. Lancet Psychiatry, 5:547.

Sajjadian M, Lam RW, Milev R, Rotzinger S, Frey BN, Soares CN et al. (2021) Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis. Psychol Med, 51:2742-2751.

Schwartz B, Cohen, ZD, Rubel JA, Zimmermann D, Wittmann WW, Lutz W (2021) Personalized treatment selection in routine care: integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. Psychother Res, 31:33-51.

Shah J, DePietro B, D'Adamo L, Firebaugh ML, Laing O, Fowler LA et al. (2022) Development and usability testing of a chatbot to promote mental health services use among individuals with eating disorders following screening. Int J Eat Disord, 55:1229-1244.

Shatte AB, Hutchinson D, Teague SJ. (2019) Machine learning in mental health: a scoping review of methods and applications. Psychol Med, 49:1426-1448.

Torjesen I. (2017) Sixty seconds on . . . sex with robots. Br Med J, 358:j3353.

Sönmez D, Hocaoğlu C. (2024) Metaverse and psychiatry: a review. Psikiyatride Güncel Yaklaşımlar, 16:225-238.

Srinivasan R, González BSM (2022) The role of empathy for artificial intelligence accountability. J Responsib Technol, 9:100021.

Stade EC, Stirman, SW, Ungar LH, Boland CL, Schwartz HA, Yaden DB et al. (2024) Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. Npj Ment Health Res, 3:12.

Roa TM, Biller-Andorno N, Trachsel M (2021) The ethics of artificial intelligence in psychotherapy. In The Oxford Handbook of Psychotherapy Ethics. (Eds M Trachsel, J Gaab, N Biller-Andorno, Ş Tekin, JZ Sadler):744-758. Oxford, Oxford University Press.

Stark L, Hoey J. (2021) The ethics of emotion in artificial intelligence systems. In FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency: 782-793. New York, Association for Computing Machinery. New York, United States.

Su C, Aseltine R, Doshi R, Chen K, Rogers SC, Wang F. (2020) Machine learning for suicide risk prediction in children and adolescents with electronic health records. Transl Psychiatry, 10:413.

Swaminathan A, López I, Mar RAG, Heist T, McClintock T, Caoili K et al. (2023) Natural language processing system for rapid detection and intervention of mental health crisis chat messages. NPJ Med, 6:213.

Taliaz D, Spinrad A, Barzilay R, Barnett-Itzhaki Z, Averbuch D, Teltsh O et al. (2021) Optimizing prediction of response to antidepressant medications using machine learning and integrated genetic, clinical, and demographic data. Transl Psychiatry, 11:381.

Teding van Berkhout E, Malouff JM (2016) The efficacy of empathy training: a meta-analysis of randomized controlled trials. J Couns Psychol, 63:32-41.

Thakkar A, Gupta A, De Sousa A. (2024) Artificial intelligence in positive mental health: a narrative review. Front Digit Health, 6:1280235.

Turing AM (1950) Computing machinery and intelligence. Mind, 59:433-460.

Tutun S, Johnson ME, Ahmed A, Albizri A, Irgil S, Yesilkaya I et al. (2023) An AI-based decision support system for predicting mental health disorders. Inf Syst Front, 25:1261-1276.

van den Bosch K, Bronkhorst K (2019) Six challenges for human-AI co-learning. Adaptive Instructional Systems, 11597:572–589.

Wada K, Shibata T (2007) Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house. IEEE Trans Robot, 23:972-980.

Wang L, Fagan C, Yu CI (2020) Popular mental health apps (MH apps) as a complement to telepsychotherapy: guidelines for consideration. J Psychother Integr, 30:265-273.

Ward T, Rus-Calafell M, Ramadhan Z, Soumelidou O, Fornells-Ambrojo M, Garety P et al. (2020) AVATAR therapy for distressing voices: a comprehensive account of therapeutic targets. Schizophr Bull, 46:1038-1044.

Weizenbaum J. (1976) Computer Power and Human Reason: From Judgment to Calculation. London, Freeman.

Wenzel J, Haas SS, Dwyer DB, Ruef A, Oeztuerk OF, Antonucci LA. et al. (2021) Cognitive subtypes in recent onset psychosis: distinct neurobiological fingerprints? Neuropsychopharmacology, 46:1475-1483.

Wiebe A, Aslan B, Brockmann C, Lepartz A, Dudek D, Kannen K et al. (2023) Multimodal assessment of adult attention-deficit hyperactivity disorder: a controlled virtual seminar room study. Clin Psychol Psychother, 30:1111-1129.

Wong WH (2023) We, The Data: Human Rights in The Digital Age. Massachusetts, MIT Press.

Yu R, Hui E, Lee J, Poon D, Ng A, Sit K et al. (2015) Use of a therapeutic, socially assistive pet robot (PARO) in improving mood and stimulating social interaction and communication for people with dementia: study protocol for a randomized controlled trial. JMIR Res Protoc, 4:e45.

Authors Contributions: The author(s) have declared that they have made a significant scientific contribution to the study and have assisted in the preparation or revision of the manuscript
Peer-review: Externally peer-reviewed.
Conflict of Interest: No conflict of interest was declared.
Financial Disclosure: No financial support was declared for this study.