

ARTICLES

Balancing Ethics and Opportunities: The Role of AI in Psychotherapy and Counselling

Alexandra Bloch-Atefi^a

Keywords: Artificial Intelligence (AI), mental health, psychotherapy, counselling, ethical challenges, AI-assisted therapy, mental health technology
<https://doi.org/10.59158/001c.129884>

Psychotherapy and Counselling Journal of Australia

Vol. 13, Issue 1, 2025

The integration of artificial intelligence (AI) into psychotherapy and counselling practices presents both significant opportunities and ethical challenges. AI applications, such as automated scheduling, AI-assisted note taking, client engagement tools, and predictive analytics for client needs, are transforming the delivery of mental health care. However, the ethical considerations surrounding AI's role in mental health care are complex and multifaceted, necessitating careful scrutiny. This paper discusses the ethical considerations required for the responsible implementation of AI technologies in psychotherapy and counselling, focusing on data protection, client consent, and the preservation of the therapeutic relationship. It argues that professional bodies and educational institutions must collaborate to develop dynamic, adaptable ethical guidelines that ensure the safe and effective use of AI tools. Furthermore, it emphasises the need for robust data protection mechanisms to safeguard sensitive client information and proposes strategies to balance the benefits of AI with the preservation of human connection in mental health care.

Public Significance Statement

Recent advancements in artificial intelligence (AI) technologies, such as large language models like ChatGPT and conversational artificial intelligence systems (commonly referred to as *AI chatbots*), have garnered significant attention in the public sphere, driving innovation across various sectors including clinical and mental health care. While the responsible integration of AI tools holds promise for improving access and affordability of care, important considerations exist for their application in therapeutic settings. It is essential that AI administrative technologies, such as automated scheduling and data management systems, and AI complementary tools, like decision-

^a Dr. Alexandra Bloch-Atefi is a registered psychotherapist and clinical supervisor with more than 14 years of experience in higher education. She currently serves as Program Director for the Graduate Programs in Counselling and Psychotherapy within the School of Public Health at the University of Adelaide, soon to be known as Adelaide University. Alexandra is also the Program Lead for the trauma specialisation and research specialisation of the Master of Counselling and Psychotherapy. She has over 20 years of experience as a clinician in the mental health field and has worked across a variety of settings including community services, business organisations, hospitals, and private practice. Alexandra holds memberships with the Psychotherapy and Counselling Federation of Australia and the Australian Psychological Society. Correspondence concerning this article should be addressed to Dr. Alexandra Bloch-Atefi at alexandra.bloch-atefi@adelaide.edu.au

support algorithms and diagnostic aids, be introduced into clinical practice gradually. Therefore, a staged and thoughtful integration of AI—guided by ethical frameworks, informed consent, and ongoing evaluation—is crucial to ensuring that AI contributes meaningfully to addressing the global mental health crisis (Fiske et al., 2019). Psychotherapists and counsellors, like all mental health professionals, must be trained to address the ethical implications of AI use, ensuring that client autonomy, confidentiality, dignity, and informed consent are consistently upheld in alignment with the core principles of therapeutic practice. This approach will ensure that AI applications provide meaningful support, complementing and enhancing care quality while respecting the foundational principles of ethical clinical practice.

Introduction

Over 10 million Australians experience mental health challenges each year, evidenced by a consistent 3% annual increase in prevalence (Australian Bureau of Statistics, 2020–2022). This growing demand places considerable strain on an already overburdened mental health system which struggles to meet the need for care. Despite the availability of evidence-based treatments, including psychological therapies and medications, approximately one-third of individuals requiring treatment do not receive care within a 12-month period (Jorm, 2018). This gap represents a critical public health concern since untreated mental health conditions can lead to worsened symptoms, chronicity, and significant impacts on quality of life (Mohamed Ibrahim et al., 2020).

A meta-analysis of 34 studies found that 75% of individuals with mental health challenges preferred psychological treatment over medication (McHugh et al., 2013), thereby highlighting the need for expanded access to psychotherapy and counselling. However, psychological treatment is often unavailable when individuals need it most. In Australia, psychiatric medication is the most common treatment for mental health challenges, and approximately one in six Australians receive prescriptions, particularly in rural and remote areas where access to mental health professionals is limited (Australian Institute of Health and Welfare (AIHW), 2023; Bloch-Atefi et al., 2021). This issue is compounded by a global shortage of mental health professionals; for example, an estimated shortfall of 1.8 million providers exists worldwide (World Health Organization, 2022). Prior to the COVID-19 pandemic, anxiety and depression were estimated to affect 29% of the global population over their lifetime (Steel et al., 2014). The pandemic exacerbated the mental health crisis, significantly increasing the demand for support (Loosen et al., 2021; Ornell et al., 2021; Santomauro et al., 2021; Thome et al., 2021).

As the prevalence of mental health conditions rises globally, this shortage has been further strained, exacerbating disparities in care, particularly in low- and middle-income countries (Lange, 2021; Lora et al., 2020). In Australia, these disparities are also evident within culturally and linguistically diverse

(CALD) communities, the LGBTIQ+ community, and Aboriginal and Torres Strait Islander populations. These groups often face additional barriers to accessing mental health care, including discrimination, a lack of culturally appropriate services, and challenges in navigating mainstream health systems (Dune et al., 2022; Morgan et al., 2021).

The global burden of mental health challenges is immense. An estimated one in eight individuals are affected at some point in their lives, resulting in significant personal, social, and economic consequences (World Health Organization (WHO), 2022). Untreated mental health conditions contribute substantially to productivity losses and increased healthcare costs, underscoring the urgent need for scalable solutions (Lange, 2021; Lora et al., 2020). These conditions currently cost the global economy approximately USD1 trillion annually, and projections indicate that by 2030, the economic cost will rise to USD6 trillion, further highlighting the critical need for effective interventions (WHO, 2022).

In Australia, the shortage of qualified mental health professionals is equally concerning; the current 32% shortfall has been projected to increase to 42% by 2030 (Department of Health and Aged Care, 2022). These figures illustrate the widening gap between demand for mental health services and the available workforce, raising questions about the sustainability of traditional, in-person care models to meet the full scope of demand.

One potential solution to address this disparity is the integration of AI technologies into mental healthcare. For example, large language models (LLMs) like ChatGPT can be used to provide therapeutic support through text-based interactions, machine learning algorithms for diagnostic support, natural language processing (NLP) tools for clinical documentation, and AI-powered chatbots that can deliver therapeutic interventions and provide psychoeducational support between therapy sessions. These technologies have demonstrated promise in supporting mental health care by assisting with administrative tasks (Topol, 2019), diagnostic support (Bzdok & Meyer-Lindenberg, 2018), and therapeutic interventions (Fitzpatrick et al., 2017). The market for mental health apps is rapidly expanding, and over 10,000 options are currently available (King et al., 2023). While many of these apps offer basic functionalities such as mood tracking, progress monitoring, medication reminders, journaling, access to prerecorded guided meditations and breathing exercises (Fiske et al., 2019; Wasil et al., 2022), chatbots more closely resemble conventional talk therapies, offering interactive and personalised support that simulates therapeutic conversations (Gaffney et al., 2019; Miner et al., 2019).

AI technologies have the potential to bridge the gap in mental health service delivery, especially in underserved regions, by providing timely, scalable support and alleviating the pressure on overburdened professionals (Thakkar et al., 2024). However, the integration of AI into mental health care must be approached with caution. While AI holds promise, its use in therapeutic settings requires careful ethical consideration, oversight by experienced

clinicians, and alignment with established therapeutic principles (Elyoseph et al., 2024). AI tools can complement but not replace the human interaction essential to effective psychotherapy and counselling (Grodiewicz & Hohol, 2024; Montemayor et al., 2022). Therefore, a staged and thoughtful integration of AI—guided by ethical frameworks, informed consent, and ongoing evaluation—is crucial to ensuring that AI contributes meaningfully to addressing the global mental health crisis (Fiske et al., 2019).

The Role of AI in Mental Health Care

While mental health professionals have traditionally relied on human interaction to provide care, there is increasing recognition of the role AI can play in alleviating administrative burdens, improving accessibility, and reducing costs (Elyoseph et al., 2024; Lee et al., 2021). Strong clinician–client relationships are central to mental health care but can be hindered by time constraints and limited interaction. AI technologies can automate routine tasks that do not require a human touch, thereby allowing clinicians to focus more on delivering empathetic, person-centred care. By freeing up clinicians to prioritise these aspects of care, AI can enhance the therapeutic relationship, humanising mental health practice and ensuring clients receive more meaningful and personalised support (Topol, 2019). AI tools, including client booking systems, session planning applications, and AI-assisted note-taking technologies, offer significant potential to optimise clinical workflow, allowing therapists to focus more on direct client care (Elyoseph et al., 2024; Lee et al., 2021; Topol, 2019). In recent years, AI has been increasingly integrated into various areas of mental health care (D’Alfonso, 2020; Fiske et al., 2019). For example, AI-driven solutions are used to enhance the diagnosis of conditions such as depression (Mastoras et al., 2019; Ware et al., 2020) and schizophrenia (Kalmady et al., 2019), as well as to predict treatment outcomes (Thieme et al., 2020). Additionally, robots are employed to support children with autism spectrum disorders (Islam et al., 2023), while virtual reality avatars help clients manage their auditory hallucinations (O’Neill et al., 2024). AI applications—such as Woebot, a cognitive-behavioural therapy (CBT) chatbot—are already being used in clinical settings to provide low-cost, accessible mental health support (Grodiewicz & Hohol, 2024). Additionally, Replika, an avatar-based therapy app, engages users in therapeutic conversations, reconstructing a personality profile from users’ digital footprint or text exchanges. It provides a judgement-free space for users to engage in vulnerable conversations and gain insights into their own personality (Pham et al., 2022). This is just a snapshot of the growing role AI plays in mental health care.

The integration of AI into therapeutic settings has shown promising results, particularly in addressing conditions like anxiety and depression. Studies have indicated that AI-based interventions, such as internet-delivered CBT (iCBT) and computerised CBT (cCBT), offer significant symptom reduction, often comparable to traditional therapeutic approaches (Botella et al., 2010; Fitzpatrick et al., 2017; Merry et al., 2012). For example, iCBT

has proven effective in alleviating anxiety and depression, demonstrating sustained improvements at follow-up (Botella et al., 2010). However, challenges related to engagement and adherence persist, particularly among younger or less engaged populations, suggesting that more interactive or supportive features could improve retention (Fleming et al., 2012; Kretzschmar et al., 2019). Additionally, AI interventions have been found effective for young adults, students, and individuals experiencing mild to moderate symptoms, although these studies typically involved small, homogeneous groups, thus limiting their generalisability (Bowler et al., 2012; Clarke et al., 2009). While AI-based interventions have demonstrated positive effects over time (Lenhard et al., 2017), concerns about the applicability of these findings to broader populations remain. Many studies focused on narrow demographic groups, and there has been limited representation of older or more diverse populations (Danieli et al., 2022). Despite these limitations, the evidence highlights the potential for AI interventions to offer effective, scalable, and accessible solutions for mental health care. As AI continues to evolve, it is essential to balance its potential with careful consideration of its limitations and the ethical implications surrounding its use in mental health settings. The growing integration of AI into therapeutic practice holds transformative potential for mental health care, yet it requires ongoing refinement to address challenges such as user engagement and to ensure its ethical implementation (Cross et al., 2024).

Challenges, Ethical Considerations, and the Complementary Role of AI in Therapy

The integration of AI into mental health care represents a significant development in the field, offering a range of potential benefits in terms of efficiency, accessibility, and cost-effectiveness. However, because the application of AI in therapeutic settings is still in the early stages, it is imperative to approach its adoption with caution and foresight. While the capabilities of AI systems—such as LLMs and decision-support algorithms—are advancing at an unprecedented pace, it must be recognised that these technologies, at present, fall short in replicating the core components of therapeutic practice that are fundamentally human: emotional intelligence, empathy, and the nuanced understanding required to form a therapeutic alliance.

AI tools have the potential to improve the efficiency of mental health professionals by automating routine tasks such as appointment scheduling, client progress tracking, and data entry. This would enable therapists to allocate more time to direct client care, focusing on the core aspects of therapy that require emotional intelligence, clinical expertise, and human connection. Nevertheless, there is a risk that therapists may become overly reliant on AI technologies, thereby potentially undermining the therapeutic relationship. Clinicians must remain vigilant and recognise the limitations of AI and the importance of human judgement in clinical decision-making. AI

can certainly assist with diagnosis and treatment planning but its use should be framed as a tool to augment, not replace, the therapist's professional expertise (Alfano et al., 2024; Farmer et al., 2024). The integration of AI into psychotherapy and counselling offers substantial opportunities to enhance the quality and accessibility of care, yet it also presents several ethical and practical challenges that must be addressed (Elyoseph et al., 2024; Luxton, 2020; Manriquez Roa et al., 2021). AI can assist with administrative tasks, improve diagnostic accuracy, and streamline certain aspects of care, but it is critical that it does not replace the human therapist (Stade et al., 2024). While AI systems, such as LLMs like ChatGPT, can assist with generating responses to client inquiries and providing diagnostic support, these responses are based on algorithms and data, rather than the emotional intelligence or human experience that a therapist can offer. Unlike human therapists, AI lacks the emotional depth, empathy, and nuanced understanding vital to building a therapeutic relationship. Empathy, trust, and emotional connection are essential for creating a safe and healing environment, which AI cannot replicate (Montemayor et al., 2022).

Thus, this article proposes that AI should be viewed as a complementary tool, not a replacement for a human therapist. The unique human-to-human interactions that elicit neurological responses tied to empathy and emotional connection are irreplaceable in therapy. This is further evidenced in how human interaction and AI systems—such as LLMs like ChatGPT or AI chatbots—engage the brain in fundamentally different ways, particularly in terms of empathy, emotional regulation, and social bonding (Eslinger et al., 2021; Tang et al., 2023). In real-life therapeutic conversations, empathy involves understanding complex, layered emotions and offering responses that align with a client's deeper emotional needs (Harris, 2024). Humans can recognise and respond to subtle emotional cues, such as changes in tone or slight expressions of distress, which may not be explicitly communicated. Brain regions like the medial prefrontal cortex and anterior insula are activated during human interactions, facilitating emotional processing and empathy (Eslinger et al., 2021). While AI can simulate conversational cues, such as altering tone or using emotive language, it does not engage the mirror neuron system in the same way. AI-generated conversations, even those mediated by sophisticated LLMs, do not elicit the same automatic emotional resonance from the listener owing to the absence of genuine emotional expression and the contextual understanding that human interactions provide (Montemayor et al., 2022). Consequently, although humans may cognitively process an AI's responses, the emotional mirroring central to social connection is not replicated in these interactions. This highlights the gap between human and AI interaction, since AI cannot trigger the neurological responses associated with human empathy, such as the release of oxytocin, which plays a key role in bonding and trust. Instead, AI interactions tend to evoke more mechanical or cognitive brain activation, primarily linked to language processing and task completion (Tang et al., 2023). Therefore,

while AI can complement human therapeutic practices, it cannot replace the authentic emotional connection and neurobiological engagement that human therapists provide. Rather than replacing human involvement, AI should serve to complement and enhance the therapeutic process (Grodiewicz & Hohol, 2024; Montemayor et al., 2022).

One significant ethical concern with AI is the potential for bias in AI systems, which can arise from the datasets upon which they are trained. If AI tools are used in psychological assessments, such as diagnostic tests or evaluations of mental health conditions, or in therapeutic interventions, like automated therapy sessions or symptom tracking, there is a risk that these systems could reinforce existing societal biases. This risk is particularly concerning for clients from underrepresented or vulnerable groups, since AI-driven tools may not be attuned to the unique cultural, social or individual needs of such clients (Farmer et al., 2024). Therefore, it is crucial that AI tools are rigorously evaluated for bias, and practitioners must be aware of the potential for discrimination or inequity. Ensuring that AI systems are culturally safe and appropriate for diverse client populations is an essential responsibility for mental health professionals (Cross et al., 2024).

Moreover, as AI technologies become more integrated into clinical practice, concerns about client data protection and confidentiality grow. AI systems rely on vast datasets, often containing sensitive personal information, which increases the risk of data breaches or unauthorised access. Protecting client data is, therefore, a primary ethical consideration. The potential for data breaches, unauthorised access, or misuse of client data is a significant concern because mental health data is particularly sensitive and private (Sedlakova & Trachsel, 2022). To safeguard client confidentiality and ensure trust in the therapeutic relationship, it is critical to implement stringent data protection protocols. Clients must be fully informed about how their data will be used, and clear consent protocols should be implemented that allow them to opt out if they so choose (Cross et al., 2024; Lustgarten et al., 2020). The autonomy of clients to make informed decisions about the role AI plays in their care is paramount, and clients should not feel pressured to accept AI involvement against their will.

As outlined in the Psychotherapy and Counselling Federation of Australia's (PACFA's) (2024) practice guidelines regarding AI use, within the Australian context, practitioners must ensure that their use of AI complies with regulatory standards set by agencies such as the Office of the Australian Information Commissioner, the Digital Transformation Agency, and the Australian Human Rights Commission. Practitioners are also encouraged to engage in continuous professional development and contribute to research on AI's effectiveness in practice. Mental health professionals must be vigilant in ensuring that AI-driven tools comply with privacy regulations, such as the *Australian Privacy Principles* (Office of the Australian Information Commissioner, 2022), which govern the collection, storage, and use of

personal data (Mennella et al., 2024). Stricter regulatory frameworks are essential to safeguard client information and ensure that AI technologies are used responsibly within the mental health sector.

Training and Professional Development for Mental Health Practitioners

As artificial intelligence (AI) continues to make significant advancements, its integration into psychotherapy and mental health care is becoming increasingly relevant. However, for mental health practitioners to effectively incorporate AI into their practice, it is crucial that professional training programs evolve to include AI literacy and address the ethical implications associated with AI technology (Sedlakova & Trachsel, 2022). Future therapists must be equipped not only with the technical skills required to use AI tools effectively but also with the critical thinking abilities necessary to evaluate these tools, recognise their limitations, and apply them responsibly in clinical settings. Consequently, alongside technical training, it is essential that professional training frameworks also prioritise the development of ethical guidelines that regulate the responsible use of AI in psychotherapy and counselling, ensuring that such tools enhance rather than undermine the core human elements of therapeutic practice (Montemayor et al., 2022).

Discussion

This paper examines the integration of AI into counselling and psychotherapy, highlighting both the significant opportunities it offers and the ethical challenges it presents. While human interaction has long been the cornerstone of therapeutic care, the increasing demand for mental health services, alongside rising clinician burnout, has spurred interest in AI's potential to enhance service delivery. AI has the capacity to alleviate administrative burdens, improve diagnostic accuracy, and expand access to care, particularly in remote or underserved regions (Elyoseph et al., 2024; Lee et al., 2021). While human connection remains essential to therapy, AI technologies offer considerable promise in streamlining clinical workflows by allowing clinicians to focus on providing personalised, person-centred care.

AI tools, such as automated client scheduling systems, session tracking applications, and virtual therapeutic assistants, hold substantial potential to reduce the time clinicians spend on administrative tasks, which have become an increasing burden in contemporary mental health practice (Lee et al., 2021). In the context of heightened demand for services, AI could enable clinicians to allocate more time for direct client care, enhancing overall efficiency. For example, AI-driven platforms offer cost-effective, accessible therapeutic support for individuals who might otherwise face barriers to traditional face-to-face therapy (Grodiewicz & Hohol, 2024). Similarly, conversational AI systems can provide users with a private, non-judgemental space for therapeutic conversations, thereby addressing the stigma surrounding mental health and extending support to individuals who might otherwise be hesitant to seek help (Pham et al., 2022).

However, despite these opportunities, the integration of AI into mental health care presents a number of significant ethical challenges. One of the most pressing concerns is the potential for algorithmic bias. AI systems are only as reliable as the data upon which they are trained. If AI tools are used in assessment or intervention, there is a risk that these systems could perpetuate existing societal biases, which may be particularly detrimental to clients from underrepresented or vulnerable groups (Farmer et al., 2024). Therefore, it is crucial that AI tools are rigorously evaluated for bias, and practitioners must be aware of their potential for discrimination or inequity. Ensuring that AI systems are culturally safe and equitable, while prioritising the needs of diverse populations, is an essential responsibility for mental health professionals (Cross et al., 2024).

Another significant ethical concern involves the protection of client data. AI systems process vast amounts of data, including sensitive personal information, to function effectively. Given the confidential nature of mental health data, legitimate concerns exist about the potential for data breaches or misuse. Safeguarding client privacy is crucial for maintaining trust in the therapeutic relationship and for ensuring that individuals feel comfortable engaging with AI tools. Informed consent processes must be transparent, clearly outlining how client data will be used and ensuring clients retain the option to opt out of AI involvement in their care if they so wish (Sedlakova & Trachsel, 2022). Additionally, AI systems must comply with stringent data protection standards, such as the *Australian Privacy Principles* (Office of the Australian Information Commissioner, 2022), to ensure that data is collected, stored, and used securely.

A further ethical challenge is the preservation of empathy in the therapeutic process. While AI can support clinicians in tasks such as symptom tracking and diagnostic assistance, it inherently lacks the emotional intelligence and nuanced understanding required to foster the deep, therapeutic connections that are central to effective therapy (Montemayor et al., 2022). Empathy remains a cornerstone of therapy and is essential for building trust, emotional safety, and a strong therapeutic alliance. Therefore, AI should be viewed as a complementary tool that supports rather than replaces the human therapist. The therapeutic relationship, grounded in empathy and understanding, is unique in its ability to attune to the emotional and psychological needs of the client (Harris, 2024). While AI can support and augment certain aspects of therapeutic practice, it should be regarded as a complementary tool that enhances, rather than replaces, the human connection essential to an effective therapeutic process (Fiske et al., 2019).

Despite the promise that AI holds, it is important to acknowledge its limitations. AI technologies are still evolving and there are intrinsic challenges in their ability to comprehend context, understand cultural nuances, and recognise complex emotional states. While AI can assist in routine tasks

and offer support for specific conditions, it cannot replicate the clinical judgement, intuition, and empathy that human therapists bring to the therapeutic process.

AI interventions, such as iCBT and cCBT (Botella et al., 2010; Fitzpatrick et al., 2017; Merry et al., 2012), have shown positive results and are being increasingly used to enhance the diagnosis and treatment of various mental health conditions. However, while these interventions can be beneficial for many individuals, they may not be suitable for those with more complex or severe conditions, particularly when tailored, personalised care is required. In contrast, other AI-driven solutions, such as robots to support children with autism spectrum disorders (Islam et al., 2023) and virtual reality avatars to help clients manage auditory hallucinations, have been successfully employed in more specialised contexts (O'Neill et al., 2024). These applications demonstrate that AI can offer innovative and valuable support for diverse mental health challenges, although careful consideration of each individual's needs is essential. Moreover, many studies on AI-based interventions have involved homogeneous groups, which limits the generalisability of their findings across diverse demographic populations (Danieli et al., 2022).

In light of these challenges, it is crucial that AI literacy becomes a central component of mental health professional training. As the use of AI in psychotherapy continues to evolve, professional training programs for mental health practitioners must be updated to include AI literacy and the ethical implications of AI technology (Sedlakova & Trachsel, 2022). It is essential that future therapists are equipped with the knowledge and skills required to evaluate AI tools critically, understand their limitations, and apply them responsibly in clinical settings. Incorporating AI into psychotherapy and counselling training programs will ensure that practitioners are able to navigate the complex ethical issues that arise when integrating these technologies into their work.

Training programs should focus on not only how to use AI tools effectively but also how to recognise situations in which AI is inappropriate or human judgement and empathy are paramount. For example, while AI can assist in diagnosing certain conditions, it cannot replace the human capacity to form meaningful therapeutic relationships or exercise clinical judgement in nuanced cases (Montemayor et al., 2022). Additionally, ethical guidelines must be developed to govern the use of AI in clinical practice, ensuring that these tools prioritise client welfare, uphold confidentiality, and safeguard the integrity of the therapeutic process.

Ongoing professional development and regular evaluation of AI tools will ensure they continue to align with best practices and uphold the highest ethical standards. Mental health professionals must be supported to keep up to date with the latest developments in AI tools and their ethical implications. This ongoing education will allow practitioners to refine their skills, understand emerging risks, and continue to apply AI responsibly in their

practice. Furthermore, regular evaluation of AI tools is necessary to assess their impact on client outcomes and ensure the tools remain aligned with best practices in mental health care.

Finally, the use of AI in mental health care demands the collaborative engagement of professional bodies, academic institutions, and AI developers in the formulation of comprehensive, robust, and flexible ethical frameworks. These frameworks should be designed to guide the responsible use of AI in psychotherapy and counselling and ensure that the technologies safeguard the integrity of the therapeutic process. These frameworks should not only address data protection and transparency but also consider the impact of AI on the therapeutic relationship and the long-term effects on client outcomes. Collaboration between AI developers, mental health professionals, and regulatory bodies will be crucial to ensure that AI remains a force for good in mental health care. These frameworks should promote inclusivity, ensuring that AI-based services are accessible to all, including individuals with limited access to technology or digital literacy. By establishing clear ethical guidelines and promoting AI literacy among mental health practitioners, the profession can harness the potential of AI while mitigating its risks. This approach will ensure that AI functions as a tool that enhances, rather than disrupts, the human-centric elements of therapeutic practice.

Conclusion

In summary, AI holds significant potential to improve the efficiency, accessibility, and affordability of mental health care, particularly for underserved and remote populations. By automating routine administrative tasks, such as appointment scheduling, client progress tracking, and data entry, AI can alleviate some of the increasing administrative burdens that clinicians face and thus help address the growing demand for mental health services. Furthermore, AI technologies offer the potential for low-cost, accessible therapeutic options, for instance, AI-powered chatbots or virtual assistants, which can support individuals who might otherwise face barriers to traditional face-to-face therapy. This would be especially beneficial for those in need of ongoing, accessible support between sessions. However, its integration into psychotherapy and counselling practice requires careful consideration of ethical issues related to bias, data privacy, and the preservation of essential human qualities such as empathy, connection, and attunement. As AI continues to evolve, it will be crucial for clinicians to remain informed and discerning, ensuring that these technologies enhance, rather than replace, the human elements of therapy. By equipping therapists with AI literacy and promoting ethical guidelines, AI can be leveraged as a complementary tool that supports the therapeutic process while safeguarding the integrity of the therapeutic relationship.

Published: March 03, 2025 AEST.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Alfano, L., Malcotti, I., & Ciliberti, R. (2024). Psychotherapy, artificial intelligence and adolescents: Ethical aspects. *Journal of Preventive Medicine and Hygiene*, 64(4), E438–E442. <https://doi.org/10.15167/2421-4248/jpmh2023.64.4.3135>
- Australian Bureau of Statistics. (2020–2022). *National study of mental health and wellbeing*. <https://www.abs.gov.au/statistics/health/mental-health/national-study-mental-health-and-wellbeing/2020-2022>
- Australian Institute of Health and Welfare. (2023). *Workforce*. <https://www.aihw.gov.au/mental-health/topic-areas/workforce>
- Bloch-Atefi, A., Day, E., Snell, T., & O'Neill, G. (2021). A snapshot of the counselling and psychotherapy workforce in Australia in 2020: Underutilised and poorly remunerated, yet highly qualified and desperately needed. *Psychotherapy and Counselling Journal of Australia*, 9(2). <https://doi.org/10.59158/001c.71216>
- Botella, C., Gallego, M. J., Garcia-Palacios, A., Guillen, V., Baños, R. M., Quero, S., & Alcañiz, M. (2010). An internet-based self-help treatment for fear of public speaking: A controlled trial. *Cyberpsychology, Behavior and Social Networking*, 13(4), 407–421. <https://doi.org/10.1089/cyber.2009.0224>
- Bowler, J. O., Mackintosh, B., Dunn, B. D., Mathews, A., Dalgleish, T., & Hoppitt, L. (2012). A comparison of cognitive bias modification for interpretation and computerized cognitive behavior therapy: Effects on anxiety, depression, attentional control, and interpretive bias. *Journal of Consulting and Clinical Psychology*, 80(6), 1021–1033. <https://doi.org/10.1037/a0029932>
- Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(3), 223–230. <https://doi.org/10.1016/j.bpsc.2017.11.007>
- Clarke, G., Kelleher, C., Hornbrook, M., Debar, L., Dickerson, J., & Gullion, C. (2009). Randomized effectiveness trial of an internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. *Cognitive Behaviour Therapy*, 38(4), 222–234. <https://doi.org/10.1080/16506070802675353>
- Cross, S., Bell, I., Nicholas, J., Valentine, L., Mangelsdorf, S., Baker, S., Titov, N., & Alvarez-Jimenez, M. (2024). Use of AI in mental health care: Community and mental health professionals survey. *JMIR Mental Health*, 11, Article e60589. <https://doi.org/10.2196/60589>
- D'Alfonso, S. (2020). AI in mental health. *Current Opinion in Psychology*, 36, 112–117. <https://doi.org/10.1016/j.copsyc.2020.04.005>
- Danieli, M., Ciulli, T., Mousavi, S. M., Silvestri, G., Barbato, S., Di Natale, L., & Riccardi, G. (2022). Assessing the impact of conversational artificial intelligence in the treatment of stress and anxiety in aging adults: Randomized controlled trial. *JMIR Mental Health*, 9(9), Article e38067. <https://doi.org/10.2196/38067>
- Department of Health and Aged Care. (2022). *National mental health workforce strategy 2022–2032*. <https://www.health.gov.au/sites/default/files/2023-10/national-mental-health-workforce-strategy-2022-2032.pdf>
- Dune, T., Chimoriya, R., Caputi, P., MacPhail, C., Olcon, K., & Ogbeide, A. (2022). White and non-White Australian mental health care practitioners' desirable responding, cultural competence, and racial/ethnic attitudes. *BMC Psychology*, 10(1), Article 119. <https://doi.org/10.1186/s40359-022-00818-4>

- Elyoseph, Z., Gur, T., Haber, Y., Simon, T., Angert, T., Navon, Y., Tal, A., & Asman, O. (2024). An ethical perspective on the democratization of mental health with generative AI. *JMIR Mental Health*, 11, Article e58011. <https://doi.org/10.2196/58011>
- Eslinger, P. J., Anders, S., Ballarini, T., Boutros, S., Krach, S., Mayer, A. V., Moll, J., Newton, T. L., Schroeter, M. L., de Oliveira-Souza, R., Raber, J., Sullivan, G. B., Swain, J. E., Lowe, L., & Zahn, R. (2021). The neuroscience of social feelings: Mechanisms of adaptive social functioning. *Neuroscience and Biobehavioral Reviews*, 128, 592–620. <https://doi.org/10.1016/j.neubiorev.2021.05.028>
- Farmer, R. L., Lockwood, A. B., Goforth, A., & Thomas, C. (2024). Artificial intelligence in practice: Opportunities, challenges, and ethical considerations. *Professional Psychology: Research and Practice*, 56(1), 19–32. <https://doi.org/10.1037/pro0000595>
- 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. *Journal of Medical Internet Research*, 21(5), Article e13216. <https://doi.org/10.2196/13216>
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, 4(2), Article e19. <https://doi.org/10.2196/mental.7785>
- Fleming, T., Dixon, R., Frampton, C., & Merry, S. (2012). A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. *Behavioural and Cognitive Psychotherapy*, 40(5), 529–541. <https://doi.org/10.1017/S1352465811000695>
- Gaffney, H., Mansell, W., & Tai, S. (2019). Conversational agents in the treatment of mental health problems: Mixed-method systematic review. *JMIR Mental Health*, 6(10), Article e14166. <https://doi.org/10.2196/14166>
- Grodiewicz, J. P., & Hohol, M. (2024). Therapeutic chatbots as cognitive-affective artifacts. *Topoi*, 43, 795–807. <https://doi.org/10.1007/s11245-024-10018-x>
- Harris, L. T. (2024). The neuroscience of human and artificial intelligence presence. *Annual Review of Psychology*, 75, 433–466. <https://doi.org/10.1146/annurev-psych-013123-123421>
- Islam, M. A., Hasan, M. M., & Deowan, S. A. (2023). Robot-assisted training for children with autism spectrum disorder: A review. *Journal of Intelligent & Robotic Systems*, 108(3), Article 41. <https://doi.org/10.1007/s10846-023-01872-9>
- Jorm, A. F. (2018). Australia's "Better Access" scheme: Has it had an impact on population mental health? *Australian & New Zealand Journal of Psychiatry*, 52(11), 1057–1062. <https://doi.org/10.1177/0004867418804066>
- Kalmady, S. V., Greiner, R., Agrawal, R., Shivakumar, V., Narayanaswamy, J. C., Brown, M. R. G., Greenshaw, A. J., Dursun, S. M., & Venkatasubramanian, G. (2019). Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning. *NPJ Schizophrenia*, 5, Article 2. <https://doi.org/10.1038/s41537-018-0070-8>
- King, D. R., Emerson, M. R., Tartaglia, J., Nanda, G., & Tatro, N. A. (2023). Methods for navigating the mobile mental health app landscape for clinical use. *Current Treatment Options in Psychiatry*, 10, 72–86. <https://doi.org/10.1007/s40501-023-00288-4>

- Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., Singh, I., & NeurOx Young People's Advisory Group. (2019). Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical Informatics Insights*, 11, Article 1178222619829083. <https://doi.org/10.1177/1178222619829083>
- Lange, K. W. (2021). Task sharing in psychotherapy as a viable global mental health approach in resource-poor countries and also in high-resource settings. *Global Health Journal*, 5(3), 120–127. <https://doi.org/10.1016/j.glohj.2021.07.001>
- Lee, E. E., Torous, J., De Choudhury, M., Deep, C. A., Graham, S. A., Kim, H., Paulus, M. P., Krystal, J. H., & Jeste, D. V. (2021). Artificial intelligence for mental healthcare: Clinical applications, barriers, facilitators and artificial wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6(9), 856–864. <https://doi.org/10.1016/j.bpsc.2021.02.001>
- Lenhard, F., Andersson, E., Mataix-Cols, D., Rück, C., Vigerland, S., Högström, J., Hillborg, M., Brander, G., Ljungström, M., Ljótsson, B., & Serlachius, E. (2017). Therapist-guided, internet-delivered cognitive-behavioural therapy for adolescents with obsessive-compulsive disorder: A randomized controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 56(1), Article e2. <https://doi.org/10.1016/j.jaac.2016.09.515>
- Loosen, A. M., Skvortsova, V., & Hauser, T. U. (2021). Obsessive-compulsive symptoms and information seeking during the Covid-19 pandemic. *Translational Psychiatry*, 11(1), Article 309. <https://doi.org/10.1038/s41398-021-01410-x>
- Lora, A., Hanna, F., & Chisholm, D. (2020). Mental health service availability and delivery at the global level: An analysis by countries' income level from WHO's Mental Health Atlas 2014. *Epidemiology and Psychiatric Sciences*, 29, Article e2. <https://doi.org/10.1017/S2045796017000075>
- Lustgarten, S. D., Garrison, Y. L., Sinnard, M. T., & Flynn, A. W. (2020). Digital privacy in mental healthcare: Current issues and recommendations for technology use. *Current Opinion in Psychology*, 36, 25–31. <https://doi.org/10.1016/j.copsyc.2020.03.012>
- Luxton, D. D. (2020). Ethical implications of conversational agents in global public health. *Bulletin of the World Health Organization*, 98(4), 285–287. <https://doi.org/10.2471/BLT.19.237636>
- Manriquez Roa, T., Biller-Andorno, N., & Trachsel, M. (2021). The ethics of artificial intelligence in psychotherapy. In M. Trachsel, Ş. Tekin, N. Biller-Andorno, J. Gaab, & J. Z. Sadler (Eds.), *The Oxford handbook of psychotherapy ethics* (pp. 744–758). Oxford University Press.
- Mastoras, R. E., Iakovakis, D., Hadjidimitriou, S., Charisis, V., Kassie, S., Alsaadi, T., Khandoker, A., & Hadjileontiadis, L. J. (2019). Touchscreen typing pattern analysis for remote detection of the depressive tendency. *Scientific Reports*, 9(1), Article 13414. <https://doi.org/10.1038/s41598-019-50002-9>
- McHugh, R. K., Whitton, S. W., Peckham, A. D., Welge, J. A., & Otto, M. W. (2013). Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: A meta-analytic review. *The Journal of Clinical Psychiatry*, 74(6), 595–602. <https://doi.org/10.4088/JCP.12r07757>
- Mennella, C., Maniscalco, U., De Pietro, G., & Esposito, M. (2024). Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon*, 10(4), Article e26297. <https://doi.org/10.1016/j.heliyon.2024.e26297>
- Merry, S. N., Stasiak, K., Shepherd, M. J., Frampton, C., Fleming, T., & Lucassen, M. F. (2012). The effectiveness of SPARX, a computerised self-help intervention for adolescents seeking help for depression: Randomised controlled non-inferiority trial. *The BMJ*, 344, Article e2598. <https://doi.org/10.1136/bmj.e2598>

- Miner, A. S., Shah, N., Bullock, K. D., Arnow, B. A., Bailenson, J., & Hancock, J. (2019). Key considerations for incorporating conversational AI in psychotherapy. *Frontiers in Psychiatry*, 10, Article 746. <https://doi.org/10.3389/fpsy.2019.00746>
- Mohamed Ibrahim, O. H., Ibrahim, R. M., Al-Tameemi, N. K., & Riley, K. (2020). Challenges associated with mental health management: Barriers and consequences. *Saudi Pharmaceutical Journal*, 28(8), 971–976. <https://doi.org/10.1016/j.jsps.2020.06.018>
- Montemayor, C., Halpern, J., & Fairweather, A. (2022). In principle obstacles for empathic AI: Why we can't replace human empathy in healthcare. *AI & Society*, 37(4), 1353–1359. <https://doi.org/10.1007/s00146-021-01230-z>
- Morgan, A. J., Wright, J., & Reavley, N. J. (2021). Review of Australian initiatives to reduce stigma towards people with complex mental illness: What exists and what works? *International Journal of Mental Health Systems*, 15, Article 10. <https://doi.org/10.1186/s13033-020-00423-1>
- Office of the Australian Information Commissioner. (2022). *Australian Privacy Principles*. <https://www.oaic.gov.au/privacy/australian-privacy-principles>
- O'Neill, E., Bird, M., & Riches, S. (2024). Technology-enhanced psychological assessment and treatment of distressing auditory hallucinations: a systematic review. *Mental Health and Digital Technologies*, 1(2), 118–140. <https://doi.org/10.1108/MHDT-03-2024-0009>
- Ornell, F., Borelli, W. V., Benzano, D., Schuch, J. B., Moura, H. F., Sordi, A. O., Kessler, F. H. P., Scherer, J. N., & von Diemen, L. (2021). The next pandemic: impact of COVID-19 in mental healthcare assistance in a nationwide epidemiological study. *The Lancet Regional Health—Americas*, 4, Article 100061. <https://doi.org/10.1016/j.lana.2021.100061>
- Pham, K. T., Nabizadeh, A., & Selek, S. (2022). Artificial intelligence and chatbots in psychiatry. *Psychiatric Quarterly*, 93, 249–253. <https://doi.org/10.1007/s11126-022-09973-8>
- Psychotherapy and Counselling Federation of Australia. (2024). *Good practice guidelines: Integrating artificial intelligence in counselling and psychotherapy*. https://pacfa.org.au/common/Uploaded%20files/PCFA/Documents/GoodPracticeGuidelines-o_IntegratingAI.pdf
- Santomauro, D. F., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., Adolph, C., Amlag, J. O., Aravkin, A. Y., Bang-Jensen, B. L., Bertolacci, G. J., Bloom, S. S., Castellano, R., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R. M., Collins, J. K., ... Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*, 398(10312), 1700–1712. [https://doi.org/10.1016/S0140-6736\(21\)02143-7](https://doi.org/10.1016/S0140-6736(21)02143-7)
- Sedlakova, J., & Trachsel, M. (2022). Conversational artificial intelligence in psychotherapy: A new therapeutic tool or agent? *The American Journal of Bioethics*, 23(5), 4–13. <https://doi.org/10.1080/15265161.2022.2048739>
- Stade, E. C., Stirman, S. W., Ungar, L. H., Boland, C. L., Schwartz, H. A., Yaden, D. B., Sedoc, J., DeRubeis, R., Willer, R., & Eichstaedt, J. C. (2024). Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *NPI Mental Health Research*, 3(1), Article 12. <https://doi.org/10.31234/osf.io/cuzvr>
- Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: A systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology*, 43(2), 476–493. <https://doi.org/10.1093/ije/dyu038>
- Tang, P. M., Koopman, J., Mai, K. M., De Cremer, D., Zhang, J. H., Reynders, P., Ng, C. T. S., & Chen, I. H. (2023). No person is an island: Unpacking the work and after-work consequences of interacting with artificial intelligence. *The Journal of Applied Psychology*, 108(11), 1766–1789. <https://doi.org/10.1037/apl0001103>

- Thakkar, A., Gupta, A., & De Sousa, A. (2024). Artificial intelligence in positive mental health: A narrative review. *Frontiers in Digital Health*, 6, Article 1280235. <https://doi.org/10.3389/fdgth.2024.1280235>
- Thieme, A., Belgrave, D., & Doherty, G. (2020). Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. *ACM Transactions on Computer–Human Interaction*, 27(5), Article 34. <https://doi.org/10.1145/3398069>
- Thome, J., Deloyer, J., Coogan, A. N., Bailey-Rodriguez, D., da Cruz e Silva, O. A. B., Faltraco, F., Grima, C., Gudjonsson, S. O., Hanon, C., Hollý, M., Joosten, J., Karlsson, I., Kelemen, G., Korman, M., Krysta, K., Lichterman, B., Loganovsky, K., Marazziti, D., Maraitou, M., ... Fond-Harmant, L. (2021). The impact of the early phase of the COVID-19 pandemic on mental-health services in Europe. *The World Journal of Biological Psychiatry*, 22(7), 516–525. <https://doi.org/10.1080/15622975.2020.1844290>
- Topol, E. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.
- Ware, S., Yue, C., Morillo, R., Lu, J., Shang, C., Bi, J., Kamath, J., Russel, A., Bamis, A., & Wang, B. (2020). Predicting depressive symptoms using smartphone data. *Smart Health*, 15, Article 100093. <https://doi.org/10.1016/j.smhl.2019.100093>
- Wasil, A. R., Palermo, E. H., Lorenzo-Luaces, L., & DeRubeis, R. J. (2022). Is there an app for that? A review of popular apps for depression, anxiety, and well-being. *Cognitive and Behavioral Practice*, 29(4), 883–901. <https://doi.org/10.1016/j.cbpra.2021.07.001>
- World Health Organization. (2022). *World mental health report*. <https://www.who.int/publications/i/item/9789240049338>