# AI in psychiatric care offers promise but raises ethical and privacy concerns



The integration of artificial intelligence (AI) into psychiatric care presents both promising opportunities and significant challenges, as explored in a recent discussion aired by the Australian Broadcasting Corporation’s program All in the Mind. The conversation featured Daniel Oberhaus, a science and technology reporter and author of the book *The Silicon Shrink*, which examines the complex role AI plays in mental health diagnosis and treatment.

Daniel Oberhaus opens the dialogue by sharing a deeply personal story: his sister Paige’s lifelong struggles with mental health following traumatic bullying experiences in early childhood, which tragically culminated in her taking her own life in 2018 shortly after her 22nd birthday. Oberhaus, who was working as a journalist for Wired magazine at the time, found himself reviewing her digital footprint, reflecting on how technology might have helped or hindered her situation.

He notes that while AI applications in psychiatry have existed since the inception of the field in the 1950s and 60s, including the early development of therapeutic chatbots, contemporary tools remain grounded on unstable foundations. Oberhaus highlights several applications of AI in psychiatry: diagnostic tools, treatment methods, and research efforts aimed at better understanding mental disorders.

One early example he mentions is the 2010 MONARCA study in Denmark, which used digital phenotyping—analysing behavioural data such as call logs and geolocation from smartphones—to monitor bipolar disorder symptoms. This approach initially showed promise by differentiating between manic and depressive states, but subsequent studies, including follow-ups, tempered expectations.

Oberhaus also points to the now-defunct US-based startup MindStrong, which aimed to commercialise AI diagnostics by monitoring smartphone behaviour to enable early intervention in psychiatric care. Despite substantial funding and leadership by a former head of the National Institute of Mental Health, MindStrong ceased operations in 2023.

Promising though some studies are, such as a 2019 Australian experiment using a simple colour-clicking game to diagnose bipolar disorder more accurately than clinicians, many fail to progress beyond initial research stages. Oberhaus emphasises that follow-up studies are lacking and real-world applications remain limited.

One active area of AI in mental health involves chatbots, such as the popular Woebot, which employ cognitive behavioural therapy (CBT) techniques. Oberhaus explains that CBT's structured approach lends itself well to scripted digital tools and is among the most effective therapies for various mental disorders. However, he is cautious about the efficacy of chatbots compared to human therapists, noting that evidence of chatbots improving patient outcomes is insufficient.

Further concerns centre on data privacy and patient autonomy. Unlike human clinicians, AI systems – especially commercial chatbots – are not bound by strict laws protecting patient data, such as the US’s Health Insurance Portability and Accountability Act (HIPAA) or similar legislation in Australia. Oberhaus cites instances where sensitive data from crisis hotlines was exploited for AI training, and points to risks of data breaches.

Moreover, there is apprehension about mandated AI monitoring in sensitive environments like workplaces, prisons, and schools, potentially infringing on individuals' rights. Oberhaus describes scenarios where AI tools may run continuously on devices without explicit consent, purportedly to detect emerging mental health crises, thereby transforming all individuals into ‘potential patients’ subjected to intrusive surveillance.

Oberhaus also warns about overreliance on AI diagnostic tools, drawing parallels to aviation where pilots became dependent on autopilot systems, diminishing manual skills. Since many AI algorithms operate as ‘black boxes’ with opaque decision-making processes, psychiatrists may find it difficult to scrutinise or trust AI-generated recommendations, raising significant ethical and clinical questions.

Illuminating the historical context, Oberhaus references Joseph Weizenbaum, creator of ELIZA, the first AI ‘therapist’ chatbot from the 1960s, who was deeply disturbed by how lonely people turned to machines for intimacy and therapy, viewing it as dehumanising. This legacy underpins current debates about the role of AI in therapy.

The discussion also highlights whistleblower testimony by Sarah Wynn Williams, former global policy director at Meta (Facebook’s parent company), who accused the company of using AI algorithms to target advertisements at vulnerable teenagers by detecting when young people felt “worthless or helpless.” Meta has denied these allegations and criticised Williams’ claims as false and defamatory.

Reflecting on whether AI could have aided his sister’s care, Oberhaus is sceptical. He doubts the effectiveness of current systems and emphasises the importance of respecting patient autonomy and privacy in any treatment approach.

For those considering AI chatbots for mental health support, Oberhaus advises caution, recommending users educate themselves about the tools' limitations and data risks.

The ABC’s All in the Mind programme, hosted by Sana Qadar, provided a detailed exploration of AI’s potential and pitfalls in psychiatry through Oberhaus’s insights, underscoring the urgency of nuanced discussions on this rapidly evolving frontier. Oberhaus’s book *The Silicon Shrink: How Artificial Intelligence Made the World an Asylum* offers an in-depth examination of these themes.

Source: [Noah Wire Services](https://www.noahwire.com)

## References

* <https://pmc.ncbi.nlm.nih.gov/articles/PMC9924259/> - This article discusses the challenges and opportunities of integrating AI in mental healthcare, highlighting its potential to support diagnosis and treatment while acknowledging implementation hurdles.
* <https://www.psychologytoday.com/us/blog/invisible-bruises/202407/the-impact-of-ai-in-the-mental-health-field> - The article explores the mixed views of clinicians on AI's integration into mental health, emphasizing both potential benefits and significant challenges.
* <https://builtin.com/artificial-intelligence/ai-mental-health> - This piece explores how AI tools can be valuable in mental health but lack genuine human empathy, making balanced integration with human interaction crucial for effective care.
* <https://www.activeminds.org/blog/exploring-the-pros-and-cons-of-ai-in-mental-health-care/> - The blog post delves into both the advantages and disadvantages of AI in mental health care, including early diagnosis potential and privacy concerns.
* <https://beetroot.co/healthcare/ai-in-mental-health-care-solutions-opportunities-and-challenges-for-tech-companies/> - This article highlights the transformative potential of AI in mental health care while addressing challenges such as data privacy, bias, and the need for balancing technology with human empathy.