AI and Digital Mental Health Tools in Australia: Risks, Regulation and Consumer Leadership

Position Paper | May 2026

NATIONAL MENTAL HEALTH CONSUMER ALLIANCE

Acknowledgement of Country

The National Mental Health Consumer Alliance acknowledges the Traditional Custodians of the lands and waters across Australia where we live, work, and advocate.

We pay our deepest respects to Aboriginal and Torres Strait Islander peoples, and to their Elders past and present. We acknowledge that First Nations lived experience is inseparable from the impacts of colonisation, dispossession, racism, and structural inequity. These ongoing injustices must be named, understood, and addressed.

The National Mental Health Consumer Alliance works in solidarity with the Indigenous Australian Lived Experience Centre, recognising the critical leadership of First Nations peoples in truth-telling, healing, and social and emotional wellbeing.

All references to ‘Consumer’ and ‘lived experience’ in this position paper refer to mental health consumers with lived experience of mental health challenges and/or suicidality. We use the term “mental health consumers” as a catchall term due to its connection with our movement’s history, but we acknowledge that different people self-identify with different terms. We do not include family, carers, kin or the bereaved in our definition of lived experience as it appears in this paper.

About us

The Alliance is the national peak body representing mental health consumers. We work together to represent the voice of all mental health consumers on national issues. We are the people experiencing mental health issues/distress, at the table advocating with government and policy makers, and working with a robust network of grassroots communities.

More information is available on the Alliance's website: nmhca.org.au.

Alliance Position

Artificial intelligence (AI) tools, including those marketed as AI (such as ChatGPT and Gemini) and AI-enabled Digital Mental Health Tools (DMHTs), are impacting people with lived experience of mental health challenges. Without strong, enforceable safeguards and genuine consumer leadership in the governance of these technologies, including the guidance for their use, there is a significant risk of these tools causing harm, eroding trust, and further limiting access to meaningful, relational and appropriate supports.

People with lived experience of mental health challenges are not passive recipients of digitally driven change. We are decision-makers, co-designers and custodians of our own wellbeing. We have the right to choose the supports and services we use, while maintaining dignity of risk, autonomy and human agency.

The Alliance strongly asserts that people with lived experience of mental health challenges must be meaningfully involved at every stage of DMHTs governance including problem definition, design, development, regulation, implementation and evaluation of all mental health specific digital tools before they are rolled out or embedded within service systems.

While these tools may improve reach, they should not replace, or be implemented as subsitutes for, locally available, relationship-based care, especially for people experiencing crisis or more complex mental health challenges. People with high mental health needs require trusted, locally available, relationship-based care that can respond with human judgement, continuity, and contextual understanding.

Background and Scope

For the purpose of this document, we have used the CSIRO definition of Artificial Intelligence (AI): *"a collection of interrelated technologies used to solve problems autonomously and perform tasks to achieve defined objectives without explicit guidance from a human being*[i]*"*.

This position paper focuses on AI and AI-enabled Digital Mental Health Tools (DMHTs) that are used specifically for mental health supports, as well as general-purpose AI tools that are increasingly repurposed for use in mental health supports. We refer to DMHTs & LLMs as digital tools to both avoid confusion about the nature of these products and include the wide-array of new technologies that make use of them.

These tools are used in diverse ways in mental health care and support from administration tasks to communication platforms, professional decision support, digital therapies (such as chatbots and digital sensing technologies) and patient monitoring[ii]. These tools may replace, supplement or influence mental health supports accessed by people with lived experience of mental health challenges.

While acknowledging wider issues associated with AI—such as privacy, data governance, authenticity, environmental and energy use concerns, security vulnerabilities and confidentiality—this paper focuses specifically on the impacts, risks and implications for people with lived experience of mental health challenges.

The Current Context

The number of digital tools used in mental health contexts has increased rapidly over the past 5–10 years. People with lived experience of mental health challenges are actively using these tools, including both purpose-built DMHTs (for example, services promoted by private health insurers, Government funded mental health services such as Beyond Blue, or charities) and general-use large language models (LLMs).

Large Language Models

LLMs including Chat GPT and Gemini are increasingly being used for mental health support, including simulating social companionship and therapeutic dialogue.

General-purpose large language model (LLM) services (e.g. ChatGPT, Claude, Gemini) are under-regulated yet increasingly imposed and adopted for everyday use. These tools are often used by the public as de facto therapeutic resources. Given the substantial risks associated with their use by the public as substitutes for mental health care, it is essential that people with lived experience of mental health challenges have a meaningful role in broader AI and LLM governance and regulatory processes.

Digital Mental Health Tools (DMHTs)

Consumers report using these tools for a range of reasons, including:

- difficulty accessing timely, appropriate mental health supports both generally and outside of business hours
- to access immediate support in working through distress and to facilitate self/emotional regulation

- shortages of services, particularly in regional and remote areas
- cost and affordability barriers
- experiences of bigotry, discrimination, vilification and harm within traditional service systems and a desire for judgement free space
- desire for autonomy and self-directed support
- desire to find information about mental health including tools, education and research and service navigation
- considerations of privacy (setting aside whether it is protected by various DMHTs)

Unfortunately, this indicates that consumers are turning to digital mental health technologies (DMHTs) to address systemic deficiencies that require policy-led solutions, not technological substitution.

General tools embedded in Mental Health services

At the same time, AI and other digital tools are being embedded into mental health and related service systems—often without consumer knowledge/understanding or consent—including:

- generative AI used to summarise clinical information
- ambient AI and digital scribes used to create clinical notes
- algorithmic tools used for triage, risk assessment and decision-making
- emerging tools linked to eligibility assessment and funding applications

In many cases, people seeking support have little or no choice about whether digital tools are used in their care. Consumer rights to care, privacy and their information need to be protected. The unliteral implementation of many of these technologies within mental health care threaten to directly and indirectly abrogate these rights.

Impact on In-Person and Peer-Led Supports

People with lived experience consistently report a preference for in-person and peer-led supports, particularly during times of distress or crisis. This is especially important for consumers who do not have access to relational support, such as supportive and respectful friends or family members. The expansion of telehealth during and after the COVID-19 pandemic has been accompanied by a measurable reduction in face-to-face services, especially in regional and remote Australia.

While AI-enabled DMHTs may increase reach or convenience for some people, unevaluated and poorly regulated tools cannot replace locally available, relationship-based care, particularly for people experiencing significant distress, trauma or crisis and present significant risks in their own right.

It is imperative that long-term planning and budgeting for expanding face to face, peer-led and community-based supports are not deprioritised in favour of DMHTs. The development of community service capacity and the peer-workforce is critical to supporting Australia's long-term mental health prevention and promotion efforts. Prioritising DMHTs at the expense of the community mental health sector, will have major long-term consequences for Australian communities.

If DMHTs are funded by the government, they must be implemented as technologies that complement—not substitute—face-to-face, peer-led and community-based supports. In addition, people with lived experience of mental health challenges must be involved in at all levels of decision making, including governance and co-design. Funding should not be diverted from in person supports to develop these tools.

Risks and Harms

AI and AI-enabled DMHTs pose specific risks to people with lived experience of mental health challenges, including:

- provision of potentially incorrect and harmful information and advice[iii]
- exacerbation of known and unknown pre-existing conditions
- lack of meaningful informed consent
- opaque data collection, storage and secondary use
- bias, discrimination and cultural harm from non-representative training data
- over-reliance on unreliable probabilistic outputs presented as authoritative or clinical, including diagnosis based on online questionnaires[iv]
- erosion of human judgement and therapeutic relationships
- reduced access to in-person services
- actively harmful or unsafe results/responses that worsen mental health and wellbeing
- lack of support to build "AI literacy" resulting in people using these tools without information on the limitations/risks

Regulation of AI-enabled digital mental health tools in Australia is spread across several areas -such as therapeutic goods, privacy, consumer law, and professional standards - with different bodies responsible for different aspects, rather than a single, unified framework[v].

The Australian National Safety and Quality Digital Mental Health Standards, written in 2020, aim to improve the quality of digital mental health services offerings[vi]. However as these standards are voluntary they are structurally insufficient resulting in AI and AI-enabled DMHTs that can be routinely used in Australia without meeting the standards. DMHTs are, theoretically, required to undergo a regulation process by the Therapeutic Goods Administration (TGA) if they meet the criteria of a medical device. That is they are intended for diagnosis, prevention, monitoring, prognosis, treatment or investigation of a physiological process[vii]. In fact, there is currently an exemption to AI therapy chatbots or any digital mental health technologies being regulated as a "medical device" if the providers take steps such as working from widely accepted cognitive behavioural therapy (CBT) models and not providing experimental therapy.

While not mandatory, this regulatory process effectively operates as a soft accreditation requirement as federal funding is limited to accredited digital mental health services.

This regulation model leaves out several AI and digital tools such as AI minute takers, Chatbots and other tools that can be used for mental health supports. There is currently no transparent or reliable mechanism for users to identify whether a DMHT has been subject to regulation and no way for users to identify whether the DMHT does what it says it does. Harm can be minimised if the developers of AI and AI-enabled DMHTs embed key ethical and legal principles - privacy and confidentiality, informed consent, fairness and bias mitigation, transparency and accountability, autonomy and human agency, and safety and efficacy - into design. Regulations need to be developed and implemented so that safety, duty of care and ethical standards are essential conditions of product service and use. Provider self-enforcement of recommended best standards are not and will not be sufficient to ensure that Australians are protected from unsafe and harmful DMHTs[viii].

General-purpose AI tools, including LLMs such as Chat GPT, Gemini, are entirely unregulated in both general public and mental health contexts.[ix] People are currently using these LLMs in place of, or in addition to, their other supports, including psychiatry, psychology and counselling therapists. People are sharing their most private thoughts with a product that emulates empathy, can reinforce harmful beliefs, is unable to assess

when someone is in danger, and is programmed to maximise engagement and/or profit rather than respond to distress safely. LLMs pose significant risks due to data harvesting, prompt injection vulnerabilities, and the potential for back-end manipulation, including steering users towards particular interventions, such as specific pharmaceuticals.

Equity, Intersectionality and Inclusion

LLM systems and DMHTs are trained on large datasets that do not adequately represent the full diversity of cultures, languages, neurodiversity, identities and lived experiences. Without deep and ongoing involvement from people with intersectional experiences—including First Nations people, culturally and linguistically diverse communities, LGBTQIA+ people, people with disability and people living in poverty—they risk embedding and amplifying existing inequities.[x]

Meaningful consumer inclusion must extend across governance, design, development, regulation, evaluation and monitoring, not be limited to consultation after key decisions have already been made. Consumers should also be involved in the development of robust, enforceable national protections for people with lived experience of mental health challenges, including:

- **National legislation**, co-designed with mental health consumers, that protects people using AI-enabled DMHTs and includes clear, accessible and meaningful informed consent requirements
- **National literacy education**, designed and delivered with people with lived experience of mental health challenges, that supports informed choice and understanding of risks and limitations
- **National Harms Register:** A publicly accessible, independent register to record and monitor harms associated with AI-enabled DMHTs, including adverse events, near misses, and patterns of risk
- **Mandatory national standards**, developed with mental health consumers, supported by a regulatory and enforcement framework that ensures transparency, accountability, and real consequences for breaches.

Language Matters

The Alliance strongly opposes the use of the term **"hallucination"** to describe incorrect or fabricated LLM or DMHT outputs, particularly in mental health contexts. This term is stigmatising and misleading. We advocate for the use of alternatives such as:

- False outputs
- Fabricated responses
- Model-generated errors

Language must reflect technical reality without reinforcing bigotry or misunderstanding about mental health.[xi]

Actions

The Alliance calls for the following actions:

Governments and Policy-Makers

- Design, develop and implement a national digital tool literacy strategy, co-designed with people with lived experience of mental health challenges, with targeted support for populations most at risk
- Guarantee ongoing investment in affordable, accessible, in-person and peer-led supports, including navigation assistance
- Introduce mandatory, enforceable Digital Mental Health Tool Standards, including clear accountability mechanisms for reporting & responding to harms arising from DMHTs
- Clarify regulatory responsibilities across agencies, including explicit guidance on what constitutes a medical device with respect to DMHTs

Service Providers and Clinicians

- Obtain explicit, informed, ongoing consent, rather than a set-and-forget arrangement before using AI tools, including LLM note-taking or decision-support tools
- Retain consumers ability to opt out with no penalty or stigma and provide genuine alternatives where consumers do not consent to LLM and/or DMHT use

- Clearly explain where data is stored, how it is used, and who is responsible for decisions informed by LLM outputs
- Provide consumers with clear guidance about when and how DMHTs or LLMs are used when engaging with a service up to and including the design and outputs of services. E.g., a service's practice includes LLM produced materials, outputs or guidance for a consumer's care, this should be stipulated and noted prior to service provision as a matter of ensuring informed consent.

Researchers and Developers

- Embed lived experience expertise at every stage of LLM and DMHT development, deployment, monitoring and governance
- Ensure LLMs and DMHTs are evaluated for safety, effectiveness, cultural relevance and bias before deployment
- Avoid language, such as 'hallucinations' that pathologises and trivialises harm.

Recognition of Lived Experience

As a consumer lived experience-led organisation, the National Mental Health Consumer Alliance values the skill and expertise of consumers with lived experience. We pay tribute to those we have lost for the work that they have done to advocate for our rights. We acknowledge that we stand on the shoulders of giants who have paved the way for the rights we have today, and we will continue their work today and every day until the mental health system recognises and upholds our human rights.

Nothing about us without us.

Submission prepared May 2026. National Mental Health Consumer Alliance.

See nmhca.org.au for more information about the Alliance.

For questions about this submission, please contact us at policy@nmhca.org.au.

References

[i] Hajkowicz SA et al (2019) Artificial intelligence: Solving problems, growing the economy and improving our quality of life. CSIRO Data61, Australia.**:/** _.pdf

[ii] https://www.mentalhealtheurope.org/wp-content/uploads/2025/02/Study-on-AI-in-mental-health-care-final-for-publication.pdf

[iii]Kalai AT et al (2025) Why language models hallucinate, arXiv:2509.04664v1. https://arxiv.org/pdf/2509.04664v1 (April 2026)

[iv] Https://Iview.abc.net.au/Video/NC2603H010S00. (2026, April 20). *Attention Deficit* [Video]. ABC iView. https://iview.abc.net.au/

[v] Gooding, P., & Pink, G. (2026). Regulating digital mental health services in Australia: Strengthening oversight and clarifying complaints mechanisms. Australian Journal of Public Administration, 85, 206–211. https://doi.org/10.1111/1467-8500.70013 (March 2026)

[vi] National Safety and Quality Digital Mental Health Standards, https://www.safetyandquality.gov.au/sites/default/files/2025-08/national-safety-and-quality-digital-mental-health-nsqdmh-standards.pdf (March 2026)

[vii] Artificial intelligence (AI) and medical device software regulation | Therapeutic Goods Administration (TGA) (March 2026)

[viii] Australian Commission on Safety and Quality in Healthcare, AI Clinical Use Guide – Guidance for clinicians, Version 1.0, August 2025, https://www.safetyandquality.gov.au/sites/default/files/2025-08/ai-clinical-use-guide.pdf (March 2026)

[ix] ibid

[x] Gilbert, Kellie; Muchamore, I; Katterl, S; Purdon, H; Allen, A; Ozols, I; et al. (2025). Digital futures in mind: Why lived experience collaboration must guide digital mental health technologies. Australian Journal of Social Issues, 60, 196–215.

[xi] Søren Dinesen Østergaard, Kristoffer Laigaard Nielbo, False Responses From Artificial Intelligence Models Are Not Hallucinations, *Schizophrenia Bulletin*, Volume 49, Issue 5, September 2023, Pages 1105–1107, https://doi.org/10.1093/schbul/sbad068 (April 2026)