AI Adoption in Large Mental Health Services

V1.0

Last updated July 2025

TACKLIT

AI Adoption in Large Mental Health Services: A Proposed Approach

Author: Isar Mazer *23 July 2025*

Introduction

The emergence of generative AI is driving significant innovation across all sectors of the economy. In each sector, senior leaders are trying to separate facts from hype, and understand what this new technology can do for their organisations. The mental health services sector is no exception. There is so much noise, how should leaders in the sector approach the AI question?

At Tacklit, we have been building and deploying technology and AI solutions for years, *exclusively* in the mental healthcare sector. From our many customer partnerships and conversations, we have developed a clear view on the safe use of AI in mental health services. We are sharing our learnings in this white paper, in the hope they are useful to other large mental health service organisations.

The Case for Deploying Artificial Intelligence in Mental Health Care Services

The mental health sector is grappling with an unprecedented crisis. Demand for services is very high and growing rapidly, far outpacing the sector's collective capacity to provide the necessary care. One in five Australians experiences a mental health incident every year, with the figure rising to two in five for individuals aged 16-24[1]. In the UK, it is one in four[2], and demand has grown by 15% just in the last two years[3].

A significant workforce shortage is exacerbating the issue, with Australia currently meeting only 35% of its psychology workforce target[4] and virtually all psychiatrists believing the workforce crisis directly impacts patient care[5]. In England, two million people are currently on waitlists for mental health services, while 19% of all mental health jobs remain unfilled[2].

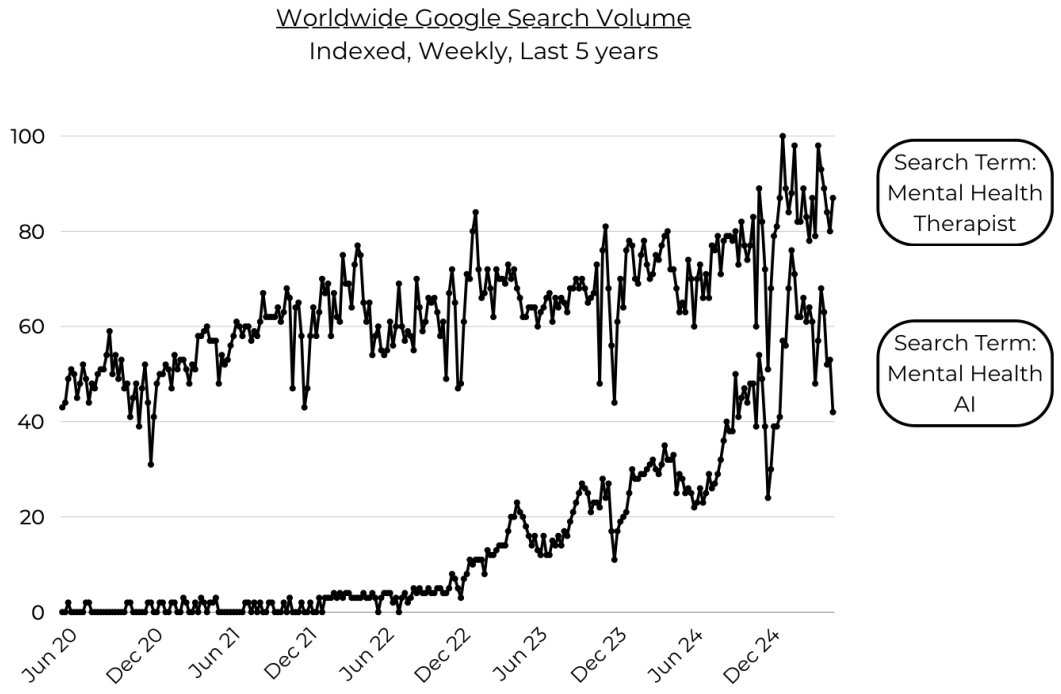
This gap between demand and available resources presents a critical challenge that must be addressed urgently. While scaling the workforce is an obvious part of the answer, technology has a critical role to play in amplifying access and quality of care. In this context, AI presents significant promise which cannot be ignored by service providers, funders and policy makers.

AI Questions Senior Leaders Should Be Asking

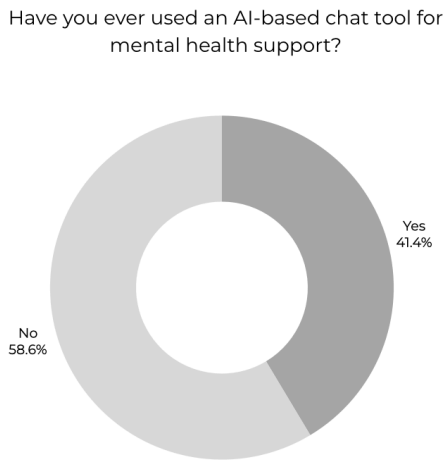
Most of the discussion to date in the industry has been about whether or not AI can and should replace human therapists.

Arguments for the potential replacement of human-led therapy include the high existing unmet demand, the early adoption, and emerging user satisfaction with AI. Positive attributes of AI-led interventions include ease/cost of access, 24x7 availability, and AI's non judgemental nature.

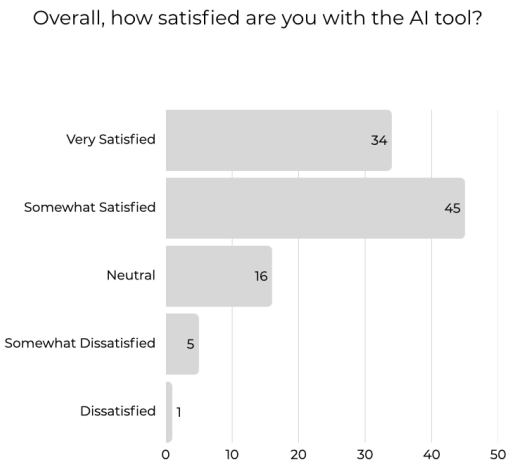
As an indication of overall consumer interest, we looked at Google search volumes for the term 'Mental Health AI', compared to the term 'Mental Health Therapist' (exhibit 1). While search volume for therapists has grown strongly, doubling in the last 5 years, interest in AI has come from nowhere three years ago, and is fast approaching similar levels.



In a recent survey by The Hemingway Group[6], over 40% of respondents state they have already used an AI chatbot for mental health, with 80% of them somewhat/very satisfied (exhibit 2).



Source: The Hemingway Group; n=452, May 2025



Source: The Hemingway Group; n=185, May 2025

Further, some contend that there may be no difference in quality or effectiveness between human and AI-powered interventions. In a fascinating study recently published in the International Journal of Human-Computer Interaction[7], 63 mental health professionals reviewed anonymised therapy transcripts, in the early, problem exploration phase. Some were from real human therapists, others were generated by an AI chatbot. The mental health professionals were asked to identify which transcripts were human-led and rate their quality. Therapists could only tell AI from human transcripts 54% of the time, no better than random guessing. Further, on average therapists scored AI-powered sessions as higher quality as those delivered by fellow human therapists.

As compelling as the arguments for AI-led therapy are, there are also warnings against its risk. A recent survey[6] suggests as many as 80% of people using generative AI for mental health do so with ChatGPT and other generic models, as opposed to specifically designed therapy agents. It has also been widely reported that AI can display sycophantic behaviour, telling users what they want to hear. Gen AI tools are known to ‘hallucinate’ and are at risk of providing rogue advice. There are reports of delusion inducement and other concerning outcomes, particularly for more severe cases such as psychosis[8].

The lack of human oversight, risk management frameworks and escalation paths in generative AI currently present unacceptable risks. Clearly, a lot further evidence and debate are required before AI can be offered at scale for unsupervised, direct care delivery.

While the questions above deserve high focus from academic and policy leaders, senior service delivery operators are best to focus on a different set of questions to ensure their organisations don’t miss out on the significant AI opportunity. For there is much AI can do to lift access and quality of care which does not entail the substantial risks outlined above.

The right question for service delivery leaders to focus on is how to start their organisation on its journey towards AI enablement, delivering on lower risk but very valuable opportunities, while we collectively learn more.

From our work with leading service delivery organisations, we see two parts to this:

1. Identifying the right roadmap of AI opportunities, based on their risk and reward profile.
2. Building the long term enablement ‘fabric’ for AI, across systems, policies and partnerships.

Assessing Risk and Return of AI Opportunities

Whilst the most important part of a frontline workers’ day is the time spent caring for people, there are many other tasks they need to perform on a daily basis. Administration, case noting, report writing, training, etc, all consume material time. The majority, if not all of these tasks have the potential to be supported by AI. The same is true in the context of service delivery organisations as a whole, where different departments will have different levels of direct contact with the people receiving care, yet could all likely benefit from AI in some way.

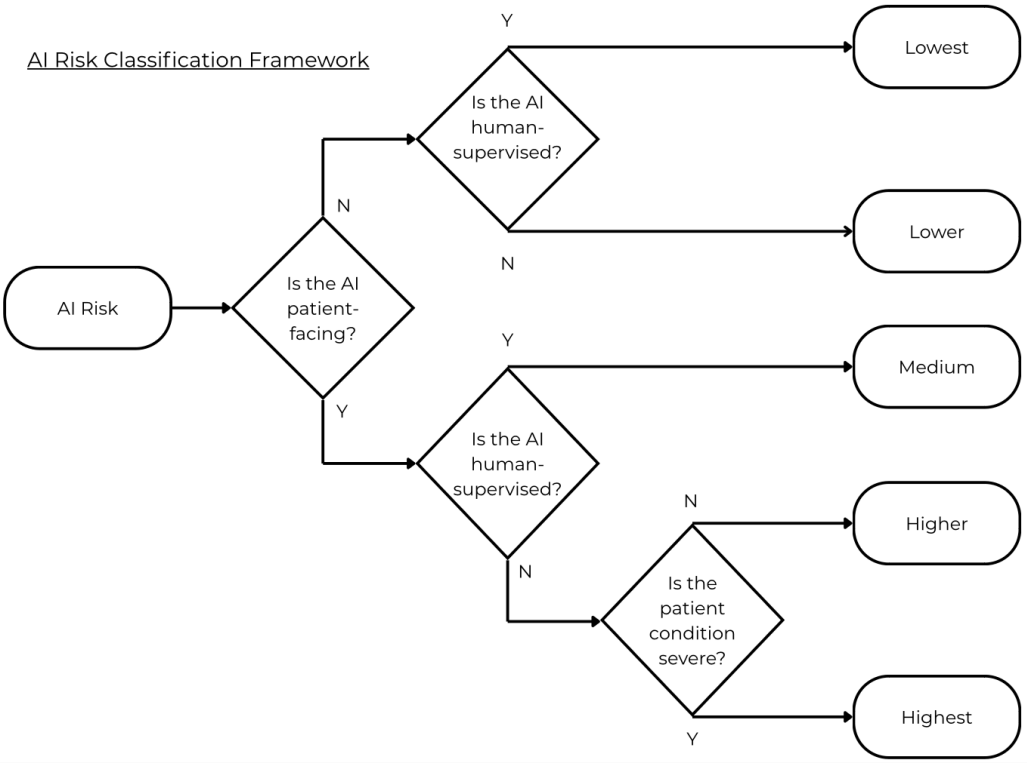
Each potential AI opportunity will hold a different level of potential to increase access and/or improve quality, and also a different level of inherent risk. Senior leaders would be well advised to understand their portfolio of opportunities, and their risk and reward characteristics, so they can pursue AI opportunities in a prudent, risk aware fashion.

Estimating AI Risk

Each organisation has its own method to understanding and managing risk, and their understanding of AI risk should be nested within their broader approach. However, from discussions with senior leaders at our largest customers, we can make some general assertions regarding the inherent risk of various AI opportunities, based on a few key factors.

First, is the AI capability in question patient-facing? Generally speaking, AI capabilities which interact directly with patients will carry higher inherent risk. Second, is the AI capability supervised by humans? If humans are reviewing the work of AI, risk will likely be lower. Finally, the severity of a patient condition is also a driver of risk, particularly in cases where the AI is patient-facing and unsupervised.

Bringing all these factors together, and customising for their individual context, service providers can develop a risk assessment framework for AI initiatives, as exemplified in the following diagram (exhibit 3).



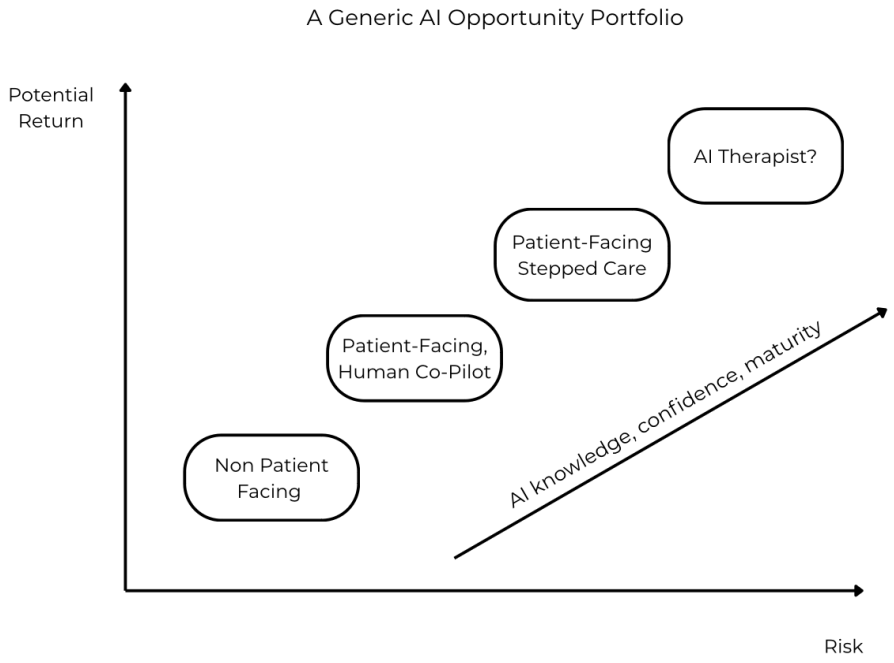
Estimating AI Potential Return

The measure of potential return of a specific opportunity will be given by the extent to which it can increase access to, or quality of care. Measures of this will vary from service to service, and depend on the type of AI opportunity. Measures may include psychometrics and other patient reported metrics, user engagement, volume of people supported for fixed resource sets, and so on. We argue that a good starting point, at least for non-patient-facing opportunities, is simply the volume of hours saved by clinical and non clinical staff.

A General AI Opportunity Portfolio

By bringing together the risk and return dimensions of an AI opportunity portfolio, service providers can identify opportunities which combine meaningful return potential with acceptable risk. Service leaders can then focus on successfully implementing these ‘low hanging fruits’, to deliver benefits whilst increasing the organisation's knowledge and support for AI, before the next opportunity is tackled.

A generic AI opportunity portfolio might look like the following picture (exhibit 4).



The generic AI opportunity portfolio above is made up of four categories:

I) Non Patient-Facing

There are many significant opportunities to save time of clinical and non clinical staff by deploying non-patient-facing AI agents at various stages of a patient journey. The most prevalent initial use case, offered by Tacklit as well as numerous other point solutions, is note taking and letter writing, where clinician time is saved, but all documentation is still reviewed and approved by its human author. Another area where Tacklit clients are already benefiting from AI agents is in the processing of referrals, which is cutting down significant admin work and reducing wait times, with all work output being reviewed and quality checked by human administrators.

II) Patient-Facing, Human Co-Pilot

A second category on the AI maturity journey could include situations where the AI is patient-facing, however working in direct service of the human in charge. In this category, Tacklit customers are already benefiting from AI risk monitoring tools, where the AI listens to clinician and peer worker interventions, and can raise alarm flags and escalation procedures according to pre-determined rules. Important to note, this is an additional risk management layer, on top of the human-led risk management procedures. By adding an AI layer on top of the pre-existing human layer, risk is reduced. Another relevant opportunity in this category is AI-supported triage, where initial steps are performed by AI only to be reviewed, progressed and finalised by the appropriate clinical staff.

III) Patient-Facing, Stepped Care

A third category is then envisaged, where the AI is allowed to interact with people receiving care in an unsupervised manner, however within a strict stepped care model. In this situation, people suffering from mild mental health symptoms can benefit from evidence-based interventions, with clear risk flags and escalation procedures to humans when necessary. One way to understand this category is to realise that today, organisations already offer evidence-based content in the form of text and video, which are consumed without any supervision. If AI can make this content more engaging, and introduce an escalation procedure for higher risk cases, then we have a clear improvement on the status quo.

IV) AI Therapist?

Conceptually, as previously discussed in this paper, a fourth category emerges where AI is able to support patients of varying levels of severity, in an unsupervised manner. We argue that significant evidence must be amassed in controlled situations before such capability can be deployed at scale. Fortunately for service providers, significant opportunity, at lower risk, is available while the debate and study of this final category rages on.

Building the Long Term Enablement Fabric for AI

In addition to the roadmap of specific AI opportunities, to achieve the full long-term potential of AI, senior leaders must also focus on the underlying enablement of the organisation, in the areas of systems, policy and partnerships.

End-to-End, Integrated System and Data

It is somewhat ironic that, while justifiable excitement exists for the potential of AI, many service delivery organisations are yet to achieve the benefits of long established technologies. The IT environment of large service providers is often very fragmented, composed of multiple poorly integrated systems. Frontline staff often need to toggle through multiple screens and copy data across. The resulting data set is therefore fragmented, and reporting is difficult. In many instances, the people receiving care have no patient portal or any other way to digitally interact with the service provider.

In this context, the overall return of any AI initiatives will be significantly constrained. As an example, while an AI scribe might still bring some benefit, it will not be able to incorporate data from the patient file, and might still require front line workers to copy and paste content across systems, wasting valuable time. More obviously, none of the many patient-facing opportunities can be properly implemented and improve patient experience, without the foundation of a consumer grade, intuitive patient-portal being in place.

To maximise the potential of AI, organisations must concurrently work to ensure that their critical workflows are seamlessly supported by integrated technology; that frontline workers, people receiving care and administrators can interact with the system via specifically designed portals, and that the data comes together in a way which can be leveraged by AI

agents and humans alike. In our conversations with customers, we refer to this as building the ‘fabric’ upon which AI agents can be ‘hung’, in a way that simplifies workflows and improves experiences, rather than making things even more complicated.

Policy, Governance and Security

Alongside the systems ‘fabric’, organisations would do well to clearly define how AI will be governed. Who decides on the risk of potential AI opportunities, and who monitors it after implementation? Some of our leading customers are putting in place AI Ethics Committees and developing AI policies to ensure good governance and management of this journey.

Of particular importance are the policies and related processes and infrastructure regarding data security. Organisations must ensure that any AI solutions do not introduce additional privacy or security risks, and avoid any personally identifiable information being shared beyond its health data cyber defences.

Trusted Vendors

To achieve the full potential of AI in mental health service delivery, and to avoid its many pitfalls, a multidisciplinary skillset is required. Clinical, service delivery, technology, data and change management skills need to come together in productive teamwork to achieve success. Almost without exception, mental health organisations will need to involve technology vendors to achieve their vision.

When assessing whether the right vendor partnerships are in place, senior leaders should consider many questions, including:

- Do the vendor(s) collectively have the necessary technical skills to support both the AI-specific as well as ‘fabric’ initiatives?
- Do the vendor(s) provide only a software platform, or do they provide the co-design, implementation, support and change management services which will be critical to project success?
- If multiple vendors are involved, does the internal team have the skills to manage the overall program and the vendor interfaces?
- Is the vendor(s) specifically focused on mental healthcare, or is it a generalist?
- Most critically, how important is your organisation likely to be for the vendor? Will the vendor stick by your side over multiple years and help you overcome the many obstacles you will no doubt encounter?

Assembling the right team is critical to AI success, and selecting the right vendors is an integral part of this.

In Conclusion

The mental health sector is grappling with an unprecedented crisis, with service demand far outstripping existing capacity. In this context, technology and AI represent a significant opportunity, which should be considered - and likely pursued - by all large service delivery

organisations. While most of the current discussion seems to be about whether or not AI can ultimately replace therapists and other frontline workers, senior leaders are best to focus on different questions: 1. What AI opportunities present the most attractive return vs. risk profile and thus should be pursued first? And 2. What are the key components of long term AI enablement, and how to pursue those? To successfully answer these questions, and ultimately succeed in implementation, senior leaders are advised to focus on assembling the correct team, including external vendors with the right skills and motivation to see them to success over the long term.

About the author
Isar Mazer is the Chairman of Tacklit, and an experienced technology and AI senior leader. You can reach him on isar@tacklit.com

About Tacklit
Tacklit is an end-to-end operating system for the delivery of mental health care. You can learn more at www.tacklit.com

References:
1 Australian Bureau of Statistics: National Study of Health and Wellbeing 2022
2 Mind UK: The Big Mental Health Report 2024
3 NHS: Mental Health Bulletin Annual Report 2023-2024
4 Australian Psychological Society: Media Release 10 July 2024
5 Royal Australia and New Zealand College of Psychiatrists: Media Release 12 February 2024
6 The Hemingway Group; Mental Health in an AI World Report; Published 9 May 2025
7 International Journal of Human Computer Interaction; Human-Human vs Human-AI Therapy: An Empirical Study; Published 16 August 2024
8 Schizophrenia Bulletin; Will Generative Artificial Intelligence Chatbots Generate Delusions in Individuals Prone to Psychosis; Published November 2023