UNIVERSITY OF CANBERRA

ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND MENTAL HEALTHCARE

AN INTRODUCTION FOR MENTAL HEALTH SERVICES AND CLINICIANS

Kelly Mazzer, Sonia Curll, Danielle Hopkins, Debra Rickwood

Faculty of Health, University of Canberra

canberra.edu.au

July 2024

Acknowledgement

This work was supported by NHMRC Partnership Grant APP1153481.

Suggested Citation

Mazzer, K., Curll, S., Hopkins, D. & Rickwood, D. (2024). Artificial intelligence, machine learning and mental healthcare: an introduction for mental health services and clinicians. University of Canberra. doi:10.31234/osf.io/a52kr

EXECUTIVE SUMMARY

The increasing rates, severity, and complexity of mental health problems are putting immense strain on Australia's mental healthcare system. The rapidly advancing field of Machine Learning [ML] offers a promising pathway to more efficient and effective mental healthcare. Mental health services and clinicians need a basic understanding of ML to make informed decisions regarding the use of ML tools in practice.

ML is a type of Artificial Intelligence [AI] that enables algorithms to autonomously learn from data to perform well-defined tasks and make future predictions. Many varying ML approaches can be applied in mental healthcare and uptake is rapidly escalating. ML tools can provide benefit in every stage of a person's mental healthcare pathway, including assessment and early detection, diagnosis and classifications of mental disorders, prognosis, treatment, and suicide prevention. Some of the key benefits and opportunities for mental health services include:

- Improved screening and other diagnostic tools
- Individualised treatment selection
- More accurate prognosis
- Novel therapeutic tools (e.g., Virtual Reality-based exposure therapy)
- Improved monitoring tools (e.g., based on data from wearable devices)
- Ongoing and out-of-hours support (e.g., chatbots, mental health apps)
- Improved classification system for mental disorders
- Time and cost savings (e.g., task automation)
- Improved approaches to clinical training, professional development, and supervision

There are also considerable challenges and risks of ML in mental healthcare that must be considered:

- Data availability, security, and privacy must be ensured. Issues such as informed consent, confidentiality and data sharing are impacted.
- ML must balance accuracy with interpretability and mitigate bias in the models.
- Acceptability and trust of ML tools for both clients and clinicians are important.
- Legal and ethical issues relating to human versus machine decision making, misinformation, and regulatory oversight.
- Practical barriers to real-world implementation of ML tools such as the necessary data system support.

ML has the potential to create a smarter, more adaptive mental healthcare system. The real-world implementation of ML tools is increasingly recognised as the solution to key issues in mental healthcare, however there are multiple interacting practical and ethical issues to resolve before ML is likely to be clinically actionable for routine care.

The evolving role of ML in mental healthcare is not about replacing clinicians, but rather, it is about providing new insights and tools that can enhance their capabilities. As ML in mental healthcare advances, it is important for services and clinicians to stay abreast of the opportunities and risks it presents. ML tools should be considered aids to be utilised in conjunction with clinicians' professional judgement and expertise. Ongoing in-depth training is needed to ensure the mental health workforce is adequately prepared to harness the benefits and mitigate the risks of ML.

Checklist of Key Considerations

The following need to be considered before mental health services and clinicians recommend or use an ML-based tool or app in practice:	
1.	**VALIDITY:** Is there sufficient evidence for the tool's effectiveness? Has it been clinically validated?
2.	**INTERPRETABILITY:** Can you understand the features and processes it uses to make recommendations?
3.	**ACCURACY:** Does the tool present a confidence level with its recommendation?
5.	**IMPLEMENTATION**: How well does the tool integrate with your existing workflow in practice?
6.	**ETHICAL**: Are there any additional ethical issues to consider? Confidentiality? Data sharing?
7.	**BENEFIT VS RISK**: Do the benefits of using the tool outweigh the risks?
8.	**ACCEPTABILITY**: Do you trust the tool? Do you think your clients will trust it?
9.	**INFORMED CONSENT**: Is your client/s fully informed about the tool and do they need to provided consent?
10.	**RESPONSIBILITY**: What regulations or guidelines apply to the use of the tool? Who is responsible for any errors in the tool?
11.	**DATA SAFETY:** Is the data stored securely within Australia? Who will have access to the data?

TABLE OF CONTENTS

INTRODUCTION

The increasing rates, severity, and complexity of mental health problems are putting immense strain on Australia’s mental healthcare system. The rapidly advancing field of Machine Learning (ML) offers a promising pathway to more efficient and effective mental healthcare. Currently, however, there are multiple ethical and practical barriers to real-world implementation of ML-based tools.

This document aims to introduce mental health clinicians to the opportunities and challenges involved with bringing ML into practice. We provide an overview of the ML process and how ML methods can overcome some of the limitations of traditional statistical methods. We then describe how ML-driven tools have the potential to improve detection, diagnosis, prognosis, and treatment of mental health problems, as well as automate clinical administration and enhance clinicians’ professional development. We include applied examples from the literature that, while not a comprehensive review, offer a glimpse into the diversity of ML-driven innovations in the mental health field. Finally, we outline the key challenges and risks involved with translating ML research into clinical practice, and the key next steps toward overcoming them.

MENTAL HEALTHCARE AND MACHINE LEARNING: WHY AND WHY NOW?

Mental health services and clinicians need to know about ML in mental health because:

- The uptake of ML use in mental healthcare is rapidly escalating.
- Clinicians and services need to make informed decisions regarding the use of ML tools in practice.
- Issues such as informed consent, confidentiality and data sharing are impacted.
- Acceptability of ML tools and responsible decision making is important for both clients and clinicians.

This is an opportune time for ML in the mental health field. Over the last three decades, significant global investment in Electronic Health Records (EHRs) has led to the production of massive clinical datasets (Jabali et al., 2022). As computing power and data storage capabilities have grown more affordable and accessible, it has become feasible to harness big data using rapidly evolving ML techniques. Many ML approaches can be applied in a mental healthcare setting, with the recent popularity of ChatGPT (a form of generative AI) increasing accessibility and awareness in the healthcare context. Given the rapid pace of advancements in this field, it is critical for mental health services and clinicians to start preparing now.

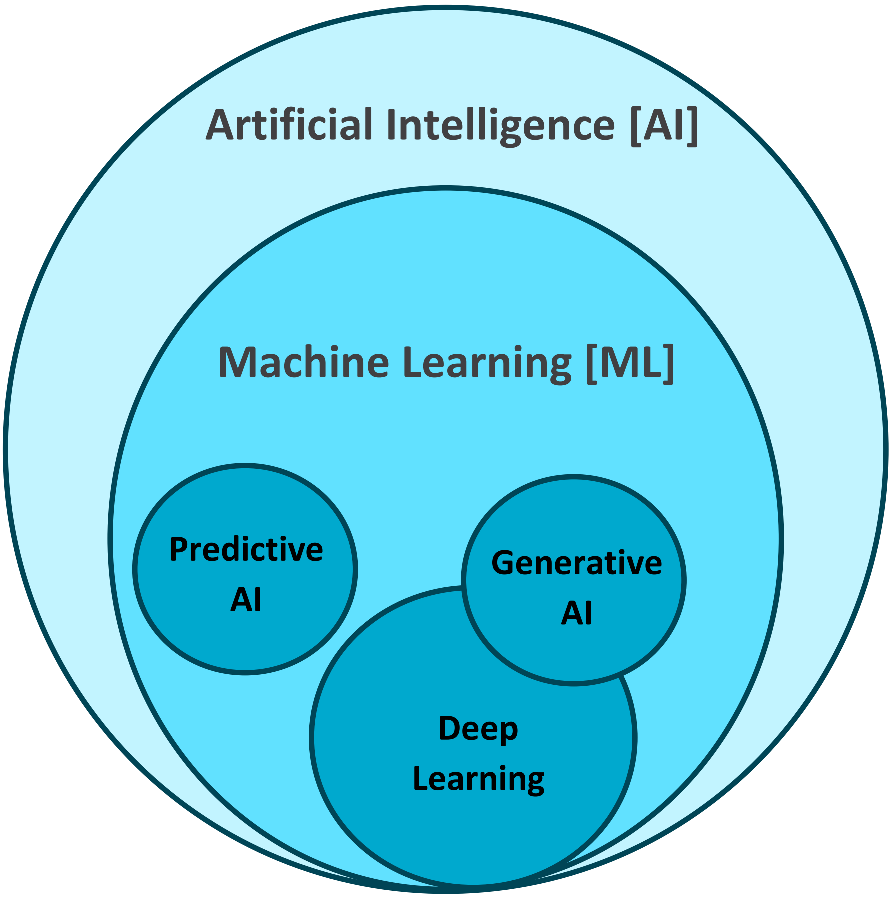
A thorough understanding of the benefits and risks is essential to using ML tools effectively and responsibly in clinical practice. Clinicians need to make informed decisions about which ML approaches they feel comfortable integrating into clinical care. A working knowledge of ML methods and processes, including potential inaccuracies, biases, and uncertainties, is an ethical responsibility. Clinicians need to be able to explain the risks and benefits to clients. Furthermore, existing obligations take on new dimensions when they involve the use of novel data collection methods (e.g., from smartphones or wearable devices) and other forms of ML technology. Services and clinicians must consider the impacts of AI and ML on responsibilities such as privacy, informed consent, safety, and accountability.

More broadly, clinicians need to understand any ML tools implemented at a mental health system level and consider the societal and economic implications of using ML in mental healthcare. These include ethical questions about digital exclusion, potential malicious uses of ML, and disproportionate impacts on vulnerable populations (Thieme et al., 2020). As the role of ML in mental healthcare evolves, greater collaboration across stakeholders, including clinicians, clients, researchers, industry, and policy makers, in all stages of the ML process will be crucial to maximise the benefits and minimise the risks for both clinicians and clients.

WHAT IS MACHINE LEARNING [ML]?

ML is a type of Artificial Intelligence [AI] that enables algorithms to autonomously learn from data to perform well-defined tasks and make future predictions. This section explains some of the key terms related to ML and shows how they relate to each other (see Figure 1).

Figure 1. Concept map of key ML terms.



Artificial Intelligence (AI). A collection of techniques that enable computers to mimic human abilities, like robotics, computer vision, and machine learning.

Machine Learning (ML). A branch of AI techniques that enable algorithms to autonomously learn from data to perform well-defined tasks and make future predictions.

Deep Learning. A subset of ML algorithms seeking to mimic how the human brain works by enabling multilayered neural networks to perform complex tasks (Al-Zaiti et al., 2022).

Predictive AI. Aims to understand patterns and relationships in data to make informed predictions. The accuracy of a forecast solely depends on the quality and relevance of the data and the level of sophistication of the algorithm. Common terms related to predictive AI include**:**

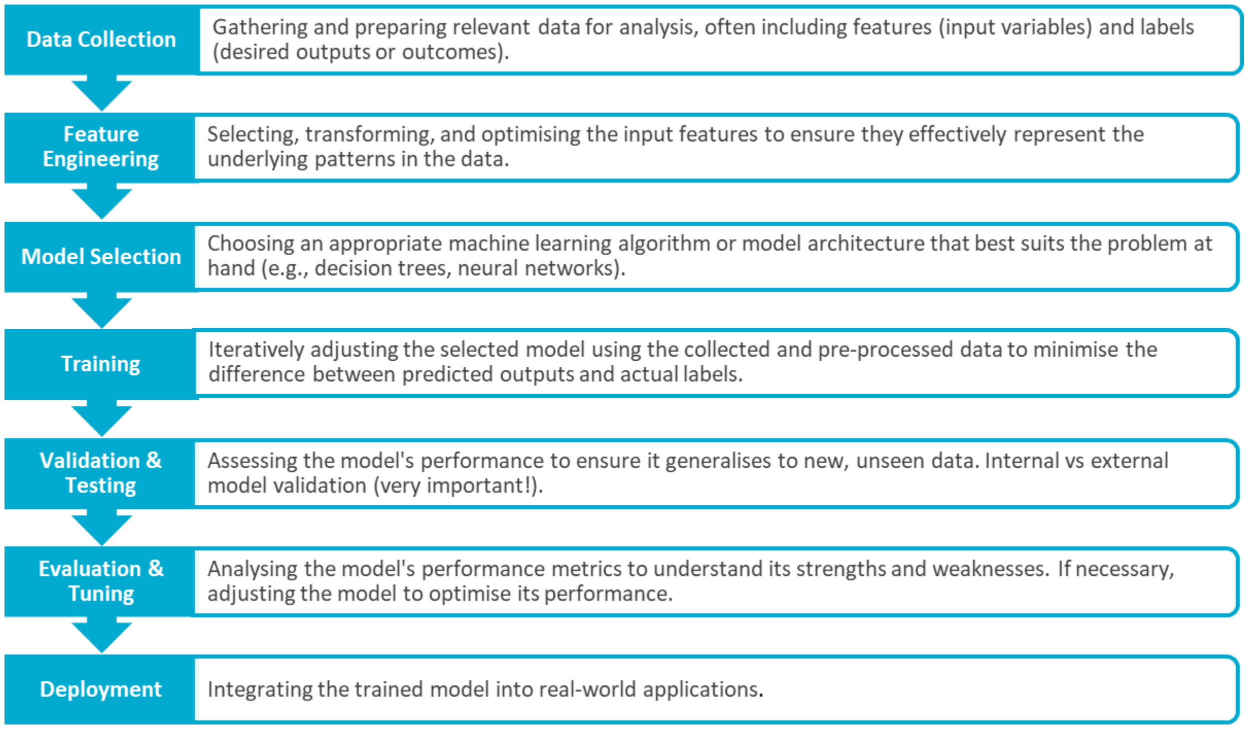
- **Unsupervised learning.** The model identifies patterns in unlabelled data. It is useful in description tasks, such as identifying dimensions and subgroups.
- **Supervised learning.** The model is trained on labelled data and is typically useful for prediction tasks where the goal is to forecast a specific outcome of interest.
- **Reinforcement learning.** The model learns by interacting with an environment and receiving feedback (Bzdok & Meyer-Lindenberg, 2018).
- **Labelled data.** Input data that has been tagged with one or more labels identifying their characteristics or classification.
- **Unlabelled data.** Raw data that has not been classified or tagged. For example, a large body of text without any specific annotations.
- **Features**. The input variables used to make predictions or classifications in ML.

Generative AI. Develops algorithms and models that can generate synthetic data, learning from itself, that closely resemble real-world data (Bandi et al., 2023). This type of AI uses a combination of ML and deep learning algorithms to produce similar but new content.

Natural Language Processing (NLP). A branch of AI that enables the representation, analysis, and generation of large quantities of unstructured language data from both written text and speech (e.g., conversation transcripts and medical records; Zhang et al., 2022).

ML PROCESS

The process of ML typically involves the following key steps:



Data Collection: Gathering and preparing relevant data for analysis, often including features (input variables) and labels (desired outputs or outcomes).

Feature Engineering: Selecting, transforming, and optimising the input features to ensure they effectively represent the underlying patterns in the data.

Model Selection: Choosing an appropriate machine learning algorithm or model architecture that best suits the problem at hand (e.g., decision trees, neural networks).

Training: Iteratively adjusting the selected model using the collected and pre-processed data to minimise the difference between predicted outputs and actual labels.

Validation & Testing: Assessing the model's performance to ensure it generalises to new, unseen data. Internal vs external model validation (very important!).

Evaluation & Tuning: Analysing the model's performance metrics to understand its strengths and weaknesses. If necessary, adjusting the model to optimise its performance.

Deployment: Integrating the trained model into real-world applications.

BENEFITS OF ML IN MENTAL HEALTHCARE

Traditional mental healthcare primarily relies on self-report by client as well as clinician assessment. The use of ML in applied mental healthcare research focuses on learning from complex data sets to make generalisable predictions about individuals. ML approaches are better able to handle non-linear associations (more complex data patterns) and work with data from multiple sources to predict outcomes at the individual level (Bzdok & Meyer-Lindenberg, 2018). In addition, ML methods can identify which variables in a dataset relate to an outcome of interest, whereas conventional methods rely on researcher input to specify which variables are relevant. This brings greater statistical power and generalisability, which leads to greater translational potential (Dwyer et al., 2018).

Data sources can include:

- E-health records
- Social media
- Smartphone data
- Wearable devices
- MH session recordings
- Imaging data
- Purpose-built data systems
- Much more …

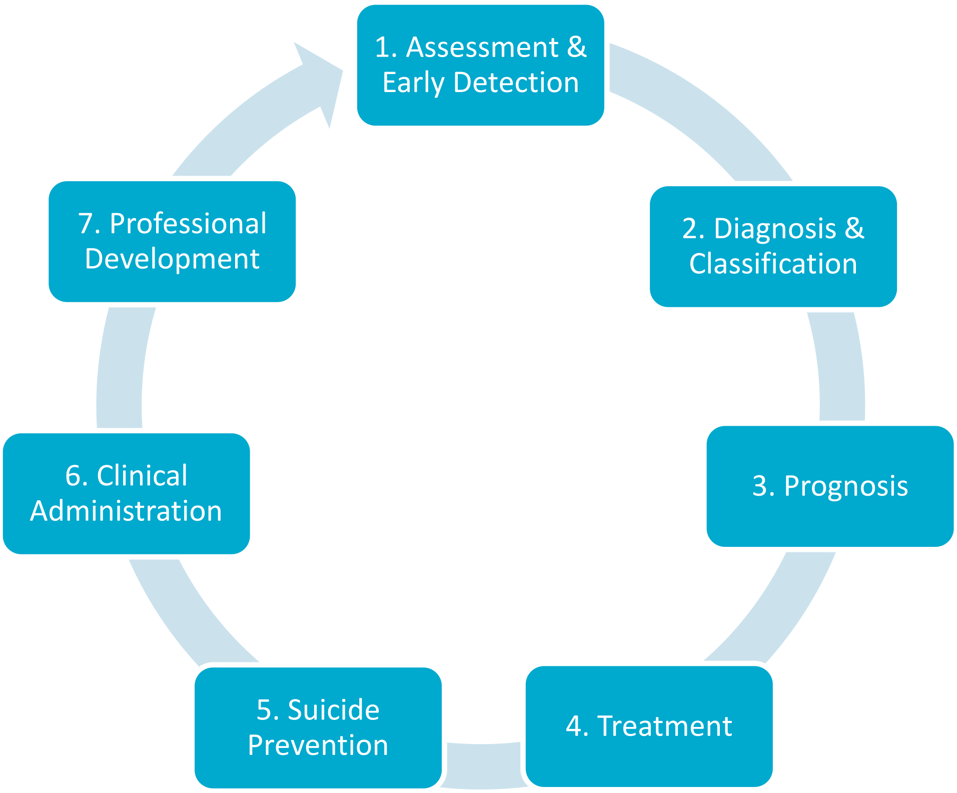
ML-driven tools offer multiple opportunities to improve the efficiency and quality of mental healthcare. Many varying ML approaches can be applied in mental healthcare and uptake is rapidly escalating. ML tools can provide benefit at every stage of a person's mental healthcare pathway, including early detection and assessment, diagnosis and classifications of mental disorders, prognosis, treatment, and suicide prevention.

Some of the key benefits and opportunities for mental healthcare include:

- Improved screening and other diagnostic tools
- Individualised treatment selection
- More accurate prognosis
- Novel therapeutic tools (e.g., Virtual Reality-based exposure therapy)
- Continuous monitoring tools (e.g., based on data from wearable devices)
- Ongoing and out-of-hours support (e.g., chatbots, apps)
- Improved classification system for mental disorders
- Time and cost savings (e.g., task automation)
- Improved approaches to clinical training, professional development, and supervision

APPLICATIONS OF ML IN MENTAL HEALTHCARE

Figure 2. Stages of a mental healthcare pathway where ML tools can provide benefit.


1. Assessment & Early Detection
2. Diagnosis & Classification
3. Prognosis
4. Treatment
5. Suicide Prevention
6. Clinical Administration
7. Professional Development

1. Early Detection and Assessment

- **Improved efficiency of detection**
- **Increased objectivity in identification and decision making**
- **Clinician support tools**

One of the most researched applications of ML in the mental health field is in improving early mental illness detection. Mental illness often goes undetected and untreated for significant periods of time after initial onset. These delays stem from many factors, including under-recognition of early symptomology, non-specific presentations that challenge differential diagnosis, and lack of access to services. Delays in receiving a diagnosis can increase personal and societal burden, and lead to poorer outcomes.

ML methods hold promise to transform mental illness detection through improved predictive modelling and clinical decision support tools. Leveraging large datasets (e.g., EHR, social media) and multimodal data (e.g., imaging, genetic, clinical), ML can uncover novel signatures that characterise specific mental disorders. These signatures can then be used to predict illness onset and differentiate diagnoses. To date, studies have mostly used ML to detect symptoms of depression and suicide risk, but also schizophrenia, bipolar disorder, post-traumatic stress disorder, and autism spectrum disorder (Graham et al., 2019; Thieme et al., 2020).

Table 1 provides some applied examples. Current evidence suggests ML models can predict future onset of mental illness with reasonable accuracy using clinical and self-report data. Models discriminating between mental disorders have also achieved promising results using cognitive batteries and neuroimaging data. Future research needs to move beyond supervised learning models, enhance interpretability, and overcome ethical challenges such as data security and confidentiality concerns (Su et al., 2020; Zhang et al., 2022). While still an emerging field, current evidence suggests ML-based tools could reduce delays between onset and diagnosis, facilitating more timely intervention.

Table 1. Applied examples

For more information click on the link for each example

- **Detecting common disorders based on EHR data**
  ML methods were applied to biomedical and demographic (EHR) data to predict Generalised Anxiety Disorder (GAD) and Major Depressive Disorder (MDD). Advanced techniques were used to identify the most important risk factors for each disorder. This research shows promise for enhancing early detection and intervention of difficult to diagnose illnesses with easy to obtain information (Nemesure et al., 2021).
- **Predicting internalising disorders based on survey data**
  ML and prospective survey data were used to develop algorithms predicting adult onset of internalising disorders (generalised anxiety, panic, social phobia, depression, and mania). Results showed acceptable (depression) to excellent (social phobia) prediction accuracy, suggesting potential value for preventative interventions (Rosellini et al., 2020).
- **Early screening for depression using Instagram**
  ML techniques were used to analyse Instagram photos (content, colour, metadata) posted by individuals who had been diagnosed with depression. The resulting models predicted depression with 70% accuracy (higher than GPs' unassisted diagnostic success rate), suggesting potential for new approaches to early screening using social media data (Reese & Danforth, 2017).

2. Diagnosis and Classification

Diagnosis

- **Ability to draw on more data and assess more complex patterns to determine diagnoses**
- **ML-based tools may enable clinicians to improve diagnostic accuracy and efficiency**

Mental illness diagnosis is a time-consuming and subjective process that typically relies on client self-report (e.g., their thoughts and feelings, changes in symptoms, interactions with others) and professional judgement (observations and interpretations made by clinicians). Constraints of this approach include the varying skillsets of clinicians, variability in clinical presentation and symptomatology, and fluctuations in the course of illness, as well as stigma, help-negation, and resource limitations.

ML-driven tools offer hope for a more accurate and efficient approach to mental illness diagnosis**.** ML techniques are uniquely placed to address the complexities involved in mental disorders; that is, each client has a unique combination of relevant factors (Koutsouleris et al., 2022). ML has been increasingly used in efforts to develop novel and objective methods for diagnosis based on specific disorder-related mechanisms, rather than the traditional self-reported symptoms-based approach. While most research has used neuroimaging data, studies have begun exploring the use of genetic, neuropsychological, and EEG measures (Shatte et al., 2019).

Table 2 provides some applied examples. Overall, the current evidence suggests that ML-tools can be used to diagnose mental disorders with above 75% prediction accuracy (Abd-Alzaraq et al., 2022; Arbabshirani et al., 2017). However, some researchers have argued that the broad clinical definitions and degree of subjective judgement involved in mental illness diagnosis limit the ability of prediction models to make decisions (Schnack & Kahn, 2016; Su et al., 2020). Hence, it is often suggested that combining clinical expertise and ML models may be the best way forward. For now, the most valuable application of ML may be in developing decision support tools for differential diagnosis (Dwyer et al., 2018).

Table 2. Applied examples

- **Diagnostic support system**
  Tested a ML-based diagnostic support system aimed to differentiate between depression and anxiety disorders based on a cognitive battery and found 66-80% classification accuracy. This objective tool is expected be used alongside clinical interview to enable clinicians to achieve increased diagnostic specificity and precision (Richter et al., 2021).
- **Differentiating between depression and schizophrenia**
  An ML-based signature that separated depression from schizophrenia was effectively used in samples with uncertain diagnoses, such as first-episode psychosis and the high-risk state for psychosis (Koutsouleris et al., 2015).
- **EarlyDetect**
  A digital app consisting of composite measures to screen for risk factors improved specificity in mental illness screening in tertiary settings, including bipolar disorder (Lui, Chokka et al., 2021) and MDD (Lui, Hankey et al., 2021).

Classification

- **Potential to increase our understanding of the multi-dimensionality of mental illness**
- **May provide more useful grouping of presentations than DSM or ICD classifications**

Current diagnostic practices rely on classification frameworks outlined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) and the International Classification of Diseases (ICD-11) manuals. These classification systems group symptoms into disorders with the aim of understanding a person's difficulties and guiding decisions about optimal care. However, there are long-standing criticisms of mental illness diagnoses and the validity of traditional mental illness classifications are increasingly being questioned (Bzdok & Meyer-Lindenberg, 2018; Love, 2018).

One of the key challenges of characterising psychopathology is the heterogeneity of mental illness; this means that a given mental disorder can present differently with varying symptoms for different people. There are also gaps in the understanding of the complex and interacting causes of dysfunction. Further, diverse data sources (e.g., genetic, imaging, self-report) can be used to inform diagnosis and classification but these are not always available or interpretable for clinicians (Ferrante et al., 2018). ML and other computational approaches provide analytic tools that can help overcome these challenges and inform an improved classification system (Bzdok & Meyer-Lindenberg, 2018; Cuthbert, 2019; see also Table 3). This has important implications for the development and delivery of targeted interventions, with increasing evidence suggesting that data-derived subgroups of clients can better predict treatment outcomes than DSM/ICD diagnoses (Dwyer et al., 2018).

TABLE 3. APPLIED EXAMPLE

- **The Research Domain Criteria (RDoC) project**
  RDoC provides a complementary approach to understanding mental disorders. This research framework emphasises the underlying multidimensionality of psychopathology, including observable behaviour and neurobiological measures, that is consistent with ML approaches (Cuthbert, 2019).
- **Identifying schizophrenia subgroups**
  ML clustering methods were used to identify subgroups of clients with schizophrenia in early-stage psychosis based on cognitive performance and neuroanatomical patterns. Early detection of subgroups could help tailor interventions (Wenzel et al., 2021).

Digital Phenotyping

- **New ways to access rich real-time, continuous, non-invasive data about clients' daily lives**
- **Smartphones and wearables can provide self-report, behavioural, and physiological data (e.g., mood, exercise, sleep)**
- **Remote monitoring and real-time feedback**
- **ML-driven tools can then use that data to support diagnostic and treatment decisions**

Smartphones and wearable sensors (e.g., FitBit, Garmin) can provide access to rich real-time data that are usually inaccessible to clinicians. Data can be collected continuously and non-invasively and can include a multitude of indicators related to sleep, exercise, stress levels, location, phone usage, social avoidance, fatigue, and mood. ML-driven diagnostic e-tools could triangulate this data with traditional sources to support more accurate and comprehensive diagnosis (Roberts et al., 2018).

Digital phenotyping also offers new ways for clinicians to monitor clients between sessions. For example, changes in movement or mobile usage could reflect changes in mood that serve as early warning signs of a mental health crisis or relapse, allowing more timely intervention (Meyerhoff et al., 2021). Similar tools could also be used to monitor clients' adherence to medication schedules, helping to identify potential lapses and providing reminders. Reporting of behavioural data can help clinicians guide clients toward established goals by providing user-friendly feedback on how to improve as well as increasing motivation and engagement in the treatment process.

Digital phenotyping for mental health is a rapidly evolving field. Table 4 presents some applied examples. Currently, however, there are very few diagnostic e-tools that have been validated or implemented on a wide scale (Roberts et al., 2018). Recent reviews concluded that although wearable AI has the potential to detect anxiety and depression, it is not yet advanced enough for clinical use (Abd-Alzaraq et al., 2023a; 2023b). Further, many experts have questioned claims that digital data can be objective and unbiased, and emphasised the ethical and practical challenges that must be overcome before relying on digital phenotyping to monitor and predict mental health problems (Birk & Samuel, 2022; Moura et al., 2023).

TABLE 4. APPLIED EXAMPLES

- **Relapse warnings based on smartphone data**
  Passively collected smartphone data can be used to monitor clients with schizophrenia to identify warning signs of relapse in real-time, supporting early intervention and improving client outcomes (Barnett et al., 2018).
- **Assessing depression severity from wearable devices**
  ML methods can be applied to data from wearable devices to assess depression severity, with implications for screening and treatment selection (Ahmed et al., 2022).
- **Moodable**
  The Mood Assessment Framework (Moodable) provides instantaneous depression and suicide risk screening based on retrospectively harvested smartphone data, opening new avenues for mental health screening at the population level (Dogrucu et al., 2020).

3. Prognosis

- **Improved ability to predict client outcome trajectories**
- **Enables greater learning from complex data sets to make generalisable predictions about subgroups, and eventually individuals**
- **Potential to support personalised preventative treatments**

Determining an individual's prognostic outcome is important for psychoeducation, management, and treatment planning. Currently, however, there is limited capacity for accurate prognosis, especially in the early stages of mental illness. This is because accurate prognosis requires personalised profiles, including environmental, biological, and genetic information, which can predict key outcomes (Dwyer et al., 2018).

The evolution of ML methods has made it realistic to achieve subgroup profiles. This progress offers the potential for ML-driven prognostic tools capable of predicting individual outcomes, such as changes in symptom severity, daily functioning and quality of life, relapses or remissions, and transitions (e.g., to psychosis). These tools could significantly improve the capacity of clinicians to deliver preventative treatments tailored to an individual's specific needs and risk level.

There are some encouraging initial results in this emerging field (see Table 5 for applied examples). Notably, one review found average predictive accuracy above 70% in predicting transition from mild cognitive impairment to Alzheimer's disease (Arbabshirani et al., 2017). However, significant practical and ethical challenges (e.g., interpretability, responsibility) must be overcome before such tools can be implemented for routine use in clinical practice (Chekroud et al., 2021).

TABLE 5. APPLIED EXAMPLES

- **PRONIA - The Project**
  A multi-site European study applying ML methods to clinical, cognitive, neuroimaging and other data from recent onset psychosis clients. The project's findings demonstrate how leveraging multimodal data and ML can help predict treatment outcomes, enabling individualised outcome trajectories and optimising treatment decisions (Koutsouleris et al., 2021).
- **Predicting functional outcomes**
  An ML-based tool to predict the functional outcomes of individuals with a first episode of psychosis using routinely reported client information has been developed and validated (Koutsouleris et al., 2018). More recent work expanded this model to include environmental factors (Antonucci et al., 2022).
- **Predicting transition to psychosis**
  ML models sequentially combining clinical and biological data with clinicians' prognostic estimates reduced clinicians' false-negative rate of psychosis prediction by over 15% (Koutsouleris et al., 2021).

4. Treatment

Individualised Treatment Plans

- **Potential to improve effectiveness of mental health treatments by targeting them toward groups of people identified as likely to respond to that approach**
- **Clinical decision support tools to improve individual treatment selection**

Treatment selection in current clinical models is often a "trial and error" approach. There are many reasons for this, including a reliance on group-level statistical evidence, therapeutic traditions (education, training, supervision), and resources. This approach can unnecessarily prolong treatment and recovery, waste resources (e.g., rebated Medicare sessions), and undermine trust in the mental healthcare system (Dwyer et al., 2018; Koutsouleris et al., 2022).

ML has the potential to improve treatment selection by advancing our understanding of the complex interplay between biological, genetic, behavioural, personality, and social and environmental factors (Bzdok & Meyer-Lindenberg, 2018). Individual treatment selection should involve weighing the benefits against the risks, side effects, possible treatment resistance, time and financial costs. Ultimately, these factors could be learned and incorporated into a clinical decision tool (Dwyer et al., 2018), making precision medicine tools a reality.

Recent ML research has used large-scale and advanced data sources to aid treatment decisions (see Table 6). While findings demonstrate potential, more work is needed before translation into clinical practice. For example, Sajjardian et al. (2021) reviewed the performance of ML methods in delivering replicable predictions of treatment outcomes for MDD (N=46 studies). They found an inverse relationship between study quality and prediction accuracy, with better quality studies showing poorer accuracy, and noted a lack of independent replication of findings. They concluded that a cautious approach is warranted when assessing the potential of ML applications for the treatment of depression.

Table 6. Applied examples

- **Antidepressant selection**
  Based on the response patterns of clients to three antidepressant medications, a predictive ML model was developed based on genetic, clinical, and demographic factors (Taliaz et al., 2021). The promising results suggest ML models could improve accuracy in antidepressant selection.
- **Clinical decision-making tool**
  A combination of ML and traditional techniques was used to develop a treatment selection algorithm to recommend cognitive-behavioural therapy (CBT) or psychodynamic therapy based on pre-treatment characteristics, to support therapists' clinical decision-making (Schwartz et al., 2021).
- **Predicting response to CBT for social anxiety disorder**
  This replication study highlights some of the exciting opportunities and challenges in developing generalisable predictive biomarkers, in this case for predicting response to CBT for social anxiety disorder (Ashar et al., 2021).

Virtual and Augmented Realities

- **Virtual reality and augmented reality applications (powered by ML) create immersive and interactive environments that offer novel diagnostic and treatment options**
- **Evidence supports VR exposure therapy in anxiety disorders and PTSD, less for other disorders (so far)**

ML-powered Virtual Reality (VR) and Augmented Reality (AR) applications can create immersive and interactive environments that offer novel diagnostic and treatment options for anxiety disorders, fears, and phobias. For example, VR exposure therapy allows individuals to experience anxiety-evoking stimuli in a safe environment, recognise specific triggers, and gradually increase their exposure to perceived threats (Bălan et al., 2020). Applied examples are presented in Table 7.

A recent systematic review of VR for mental disorders found convincing evidence for VR exposure therapy in anxiety disorders and PTSD, and promising but less robust results for VR interventions in other disorders (Weibe et al., 2022). The study noted high study variability regarding methodological rigour and application maturity, significant lack of research for obsessive-compulsive disorder and depression, and markedly more studies with adults than with children or adolescents.

TABLE 7. APPLIED EXAMPLES

- **VR exposure therapy**
  Rahman et al. (2023) explored the use of ML models to predict arousal states from physiological data. They implemented a biofeedback tool in the context of VR exposure therapy for public speaking anxiety. Similar tools could be used in other domains where arousal detection is important, allowing for more individualised and effective interventions.
- **ADHD assessment tool**
  Wiebe et al. (2023) investigated a new VR-based multimodal assessment tool for adult ADHD, involving a virtual continuous performance task with visual, auditive, and audiovisual distractions. Results suggest it was a valid approach to more accurately assess the disorder and may also be a useful approach for assessing medication effects within the ADHD population.

Chatbots

- **AI-powered chatbots can provide anonymous and unbiased support, potentially increasing motivation to seek help for mental health concerns**
- **May promote self-care and help meet growing demand for mental healthcare**
- **May be especially valuable for young people and hard-to-reach populations**
- **Chatbots have the potential to improve mental health, but current evidence is weak**

Chatbots, or conversational agents, are tools that use ML and AI to simulate human communication, either through voice or text communication. Voice-based chatbots are accessed via mobile devices, computers, and smart speakers such as Amazon Alexa and Google Home. Text-based chatbots can be accessed through channels such as Messenger, Kik, Slack, and Telegram, or in a web or mobile app. Table 8 provides some research applications.

Many chatbots are now available for use in mental healthcare service provision, including Tess, Woebot and Wysa. Most currently available mental health chatbots aim to reduce symptoms of psychological distress (stress, depression, anxiety) or promote wellbeing though improved self-awareness and coping skills. They also have the potential to provide higher-level therapeutic interventions (Bendig et al., 2022).

Chatbots offer several benefits in the realm of mental healthcare. One significant advantage lies in their 24/7 availability, extending support beyond traditional in-person sessions and addressing the growing demand for accessible mental health services. Additionally, chatbots offer anonymous and non-judgemental support, potentially alleviating stigma and increasing motivation to seek help. Indeed, research suggests that some people may be more comfortable speaking to a computer than a clinician (e.g., Lucas et al., 2017). Chatbots may be an especially valuable source of support for young people due to their greater acceptance and comfort with technology, and for hard-to-reach populations such as those living in remote areas.

Another use of chatbots in mental healthcare is to assist chat operators of mental healthcare services (e.g., Madeira et al., 2020). In this context, chatbot tools aim to improve the quality and reduce the time of interactions. Hybrid models have also been developed, where a chatbot service is combined with a real-life coach or therapist (e.g., Joyable, Talkspace in USA), minimising some of the risks of a fully automated service.

Overall, however, the current evidence for chatbots is weak. Meta-analytic reviews have highlighted several key limitations in the literature, including insufficient number of studies, risk of bias, and conflicting results (Abd-Alzaraq et al., 2020; Li et al., 2023). Personalisation, empathic response, and longer duration of interaction have been shown to facilitate efficacy (He et al., 2023), while user experience with chatbots appears to be largely shaped by the quality of human-AI therapeutic relationship, content engagement, and effective communication (Li et al., 2023).

At present, multiple issues and challenges need to be addressed. Technical issues include the risk of misinterpreting nuances in emotional expression that could lead to inappropriate or inadequate responses. Ethical issues include privacy and data security concerns (Coghlan et al., 2023). There is also a risk of over-reliance on chatbots, potentially diminishing the importance of human interaction in mental health treatment. The blurring of boundaries between reality and fiction by chatbots may have complex effects on users. This is particularly concerning for vulnerable populations who might form attachments with chatbots out of loneliness or desire for care, potentially leading to emotional transference or other negative outcomes (Fiske et al., 2019). Ensuring adequate regulation, ongoing monitoring, and the incorporation of human oversight in chatbot interactions are crucial strategies to mitigate these risks and ensure their responsible integration within mental healthcare contexts.

TABLE 8. APPLIED EXAMPLES

- **Improving treatment motivation for eating disorders**
  Shah et al. (2022) developed and evaluated a chatbot designed for pairing with online eating disorder screening. The tool aimed to improve motivation for treatment and self-efficacy among individuals with eating disorders.
- **Woebot**
  Fitzpatrick et al. (2017) found support for the feasibility, acceptability and preliminary efficacy of a fully automated conversational agent (Woebot) to deliver a CBT-based self-help program for reducing symptoms of depression and anxiety among college students.
- **Avatar intervention for cannabis use disorder**
  Giguère et al. (2023) demonstrated the short-term efficacy of an avatar intervention for cannabis use disorder. The intervention integrates virtual reality with existing techniques (CBT, motivational interviewing), allowing participants to practice them in real-time.

Mental Health Apps

- **Widely available**
- **Often target emotion regulation, well-being, mindfulness, mood, tracking**
- **Largely unregulated**
- **Not all MH apps are effective**

There is a proliferation of mental health apps available offering diverse functionalities. Some of these have the potential to enhance wellbeing and complement therapeutic interventions. Table 9 gives a few examples of available apps. The types of apps include tracking apps (e.g., mood, sleep, and symptom trackers), mindfulness and meditation exercises, CBT-based treatment apps, apps to boost mental health literacy, and apps designed for peer support that enable users to connect with others experiencing similar challenges.

Like chatbots, mental health apps can potentially assist in improving access to mental healthcare for the many people who would otherwise not have the resources or ability to connect with a therapist. Research applying ML methods to user reviews suggests that most users find mental health apps helpful, with key factors influencing perceived effectiveness including usability, personalised content, privacy and security, and customer support (Oyebode et al., 2020).

However, while in the traditional therapeutic relationship there are ethical obligations to protect client interests, in a therapy app there are no ethical obligations or clear lines of accountability to protect the user. The lack of adequate regulation in this area exacerbates concerns over how safety, privacy, accountability, and other ethical obligations to protect an individual in therapy are addressed within these services (Martinez-Martin & Kreitmair, 2018).

Moreover, as yet, there is little evidence for the effectiveness of mental health apps (Eisenstadt et al., 2021; Marshall et al., 2020). In a review of 28 popular English language mental health apps, Wang et al. (2020) found that only five had empirical support. They recommended that to maximise clinical effectiveness, clinicians should discuss with clients the credibility of the app, features and cost, privacy cost, and support required, and provide ongoing guidance.

TABLE 9. APPLIED EXAMPLES

- **Refresh**
  Refresh is an app-based unguided sleep intervention. Vollert et al. (2023) evaluated the effects of Refresh in a randomised controlled trial and concluded that while it may be a feasible and effective option, uptake and adherence need careful consideration.
- **WeClick**
  WeClick is a relationship-focused app for improving wellbeing and help-seeking intentions among adolescents from Australia (O'Dea et al., 2020).
- **Made4Me Program**
  Aims to provide information and help users manage symptoms though a self-paced CBT-based course. It also offers access to therapist support (via digital platforms).

5. Suicide Prevention

- **ML is well suited to suicide prediction and treatment decision support tools**
- **ML-based tools have outperformed existing models in identifying high-risk individuals**
- **Early research highlights potential value of ML in suicide prevention but needs further clinical guidance and external validation**

Accurate prediction of suicide risk is a vital but challenging task. ML approaches are well-suited to developing improved suicide prediction tools due to their ability to handle rare outcomes, large numbers of predictors, and correlated predictors. For example, neural network models, a powerful type of ML, have been applied to publicly available Twitter data to successfully detect suicide risk (Roy et al., 2020). Additional examples are presented in Table 10.

Beyond identifying individuals at risk of suicide, ML may play a transformative role in developing support tools for clinicians making treatment decisions for at-risk individuals, such as whether to hospitalise a client after a suicide attempt (Kessler et al., 2020; Kirtley et al., 2022). ML-based tools may also support continuous risk assessment, allowing for real-time monitoring of at-risk individuals and enabling timely interventions.

Despite the potential, research in this area is currently limited by methodological issues, most notably the lack of cross-validation (Kessler et al., 2020; Kirtley et al., 2022). There is a clear need for research using best-practice methodologies to harness the value of ML in suicide prevention.

TABLE 10. APPLIED EXAMPLES

- **Identifying individuals at risk of suicide**
  Research has shown that EHR data can be used to create predictive models for suicide risk among children and adolescents (Su et al., 2021). Other studies have found that ML models outperformed existing models in identifying high-risk individuals in vulnerable communities based on self-report data, including military personnel (Rozek et al., 2020) and Native Americans (Haroz et al., 2019). Crucial next steps include developing clinical guidance and external validation.
- **Facilitating crisis helpline triage**
  A recent study trained and deployed an NLP-based system to improve response times to help-seekers in crisis accessing a national crisis helpline (Swaminathan et al., 2023).
- **SIDVis**
  Suicide Ideation Detection Visual Interactive Systems (SIDVis) is an interactive visual dashboard designed to detect suicide ideation from social media data, enabling proactive interventions and support (Islam et al., 2023).

6. Clinical Administration

- **ML-based practice management tools could improve service quality, increase productivity, and reduce costs**
- **Improved administrative efficiency through task automation**

A key way that ML can support clinical practice is through AI-supported practice management tools, which are rapidly developing and being made commercially available (see Table 11). These tools can streamline administrative processes such as appointment scheduling, invoicing, and electronic health record management. This reduced administrative burden allows clinicians to focus on direct care and complex decision making, benefitting both clients and clinicians (e.g., less burnout; Coeira & Liu, 2022).

Large Language Models (LLMs) are likely to find application in digital scribes, assisting clinicians to create health records by listening to conversations and creating summaries of the clinical content. Furthermore, AI-based tools can transfer pertinent elements of previous sessions (such as homework, assessments, clients' goals) into future treatment plans and subsequent documentation. This can support continuity of treatment fidelity, organisation of sessions, and agenda setting (Biswas & Talukdar, 2024).

TABLE 11. APPLIED EXAMPLES

- **Otter.ai**
  Automated meeting notes & real-time transcription.
- **Heidi**
  Automated medical scribe.
- **Patient Notes**
  Automated clinical note-taking tool.

7. Professional Development

- **Enhanced training opportunities**
- **ML supported supervision and self-reflection**
- **Improved data analytic efficiency for clinical research**

ML can also enhance the professional development of mental health professionals through ML supported training, supervision, and self-reflection. A common element of training is for clinicians to role-play together (taking turns to act as client and clinician) to develop and practice new skills. ML-based tools can now simulate the role of the client, and better depict diverse client scenarios, enabling clinicians to build their skills in a safe and controlled environment with less demand on resources (i.e., without needing to involve multiple clinicians; Luxton, 2014). Clinicalnotes.ai can offer earlier career clinicians a form of "AI supervision" by analysing notes and treatment progression and suggesting evidence based next steps to deliver cohesive interventions. Such tools could also help supervisors to plan for the progression of clinical skills development for their supervisees by analysing and organising evidence for clinical competencies.

NLP verbal or text analysis can be used to analyse client sentiment (expression of emotion) and provide objective feedback to clinicians. For example, accessible and easy-to-use tools are now available to record client engagement in interventions and reactions to clinical delivery and communication (e.g., Dovetail). Other tools can produce and analyse session feedback summaries based on verbal data from sessions. This could assist clinicians in self-reflection and adjustment of clinical practice. AI-based tools can also more efficiently analyse complex patterns in client data (e.g., WhyHive), which could enable the dissemination of more novel and innovative clinical research on interventions, therapeutic methods, group programs and client outcomes. Table 12 provides examples.

TABLE 12. APPLIED EXAMPLES

- **Clinicalnotes.ai**
  A platform trained to understand clinical context and personalise automated documentation.
- **Dovetail**
  Client feedback tool that synthesizes data into actionable insights.
- **WhyHive**
  A simple to use NLP data analysis tool.

CHALLENGES AND RISKS OF ML IN MENTAL HEALTH

Data

- Data availability – lack of suitable data. Robust ML models require high-quality, reliable, representative data, which can be costly.
- Data security – cautions around sharing data and who can access it.
- Data privacy – confidentiality concerns.
- Data storage – Australian health-based data must be securely stored and remain within Australia (OAIC, 2019).

Interpretability

- ML models often function as “black boxes”, making it difficult for clinicians to understand how the algorithms arrive at specific predictions or recommendations.
- This lack of transparency limits trust and raises ethical concerns and accountability issues.

Bias

- There is a risk of bias in ML models, which can exacerbate existing health inequalities among groups (gender, racial, ethnic, socioeconomic).
- Safeguards must be built into models to mitigate bias.

Accuracy

- There is a risk of algorithms producing and spreading misinformation (algorithmic manipulation or “hallucinations”; Farquhar et al., 2024).
- At times, AI-based tools (e.g., ChatGPT, chatbots) can produce incorrect or inappropriate information. This could have consequences such as delaying clients seeking or receiving care, recommending inappropriate management, and worsening clients’ emotional state.
- Validity of diagnostic and prognostic labels provided by ML.

Acceptability

- Successful real-world implementation requires acceptance from clinicians, clients, and the community.
- Barriers to acceptability include:
  - Cultural norms within clinical practice
  - People’s capacity to trust and believe in decisions made by a machine
  - Widespread belief that clinical judgement is superior to quantitative data
  - Well-established value of person-to-person connection and the therapeutic relationship
  - Cognitive dissonance from having a bot for a therapist.

Legal and Ethical

- Machine versus human decision-making. For example, how should a conflict between human expert judgement and an ML model be resolved? Clear protocols need to be established.
- Misleading marketing.
- Not all MH apps are helpful. There is a lack of evidence for many widely available products and tools.

- Many MH apps bypass human service providers and are then not subject to existing ethical and professional standards of care.
- A comprehensive regulatory system is needed to minimise potential negative social and psychological impacts, and Australia is lagging behind other countries in this regard (Coeira et al., 2023).
- There are currently no national frameworks for an AI-ready workforce (APS, 2024).

Responsibility

- At present, ML tools are largely unregulated. Regulations and guidelines are required, including to determine:
  - Who is responsible for any errors in an ML model or tool?
  - Who takes precedence in a conflict between expert human judgement and an ML model?
  - Whose assessment takes precedence, under what circumstances, and with what accountability measures?
- Clinicians must retain the ability to overrule automated decisions that breach professional standards of care.
- Support must be available to anyone identified as at-risk by an ML model. For example, it would be irresponsible and detrimental to care if an ML model identified someone at-risk of suicide and no follow-up action was taken in a timely manner.

Implementation [Practical]

- Identifying what systems need to be in place to implement and support ML-based interventions
- Protocols to inform and support someone identified as at-risk
- Knowledge building and ongoing training for clinicians and the broader mental healthcare workforce

FUTURE – WHAT'S COMING

- **Early days, more research on ML in mental healthcare is needed**
- **Minimising error and bias in models**
- **Guidelines for transparency in ML research, including ethical sharing of data and code**
- **Establishment of regulatory oversight**
- **Development of trust and acceptability among clinicians and the wider community**
- **Rapid skill acquisition for clinicians and other mental healthcare providers to understand how to use ML-based tools effectively and responsibly**
- **The future of AI in mental healthcare is not about replacing clinicians, but rather providing tools that enhance and support their capabilities**

Many ML studies have shown the potential for translation into mental healthcare, including diagnosis, prognosis, and treatment. In the next decade, advancements in ML methods and increasing availability of data are expected to enable more individualised and proactive mental healthcare, improving outcomes and reducing the burden on mental healthcare systems. However, there are serious practical and ethical issues that need to be resolved.

From a research standpoint, significantly more investigation is required to establish the real-world reliability and effectiveness of ML tools across different mental health populations and contexts. A critical priority is developing more robust and fairer ML models. This will require generating large-scale, high-quality datasets and data sharing initiatives (see e.g., the European Open Science Cloud). To leverage this data, improved modelling techniques are needed to minimise errors, uncertainties and biases that could perpetuate health disparities. Researchers must prioritise gaining the trust and confidence of clinicians and consumers, including by involving them in all stages of research, development and implementation of ML in service delivery (e.g., Higgins et al., 2023).

From a systems perspective, regulatory oversight will play a critical role in the safe and effective implementation of ML in mental healthcare. Australia needs investment and development of a national approach. To avoid burdening our health system, clear guidelines and standards for the development, validation, and deployment of ML tools are needed. This includes ethical sharing of data and code, ongoing monitoring, and clear accountability. Collaborative efforts between regulators, healthcare providers, product developers, and people with lived experience of mental health problems are essential to create a trustworthy ecosystem for ML in mental health. For detailed recommendations, readers are directed to ’A Roadmap for AI in Healthcare for Australia’.

Clinicians must be prepared for the integration of ML into mental healthcare. Real-world deployment of ML-based tools will require mental healthcare providers to understand their risks and benefits. Future research needs to determine the minimum skill requirements for the safe integration of AI algorithms in mental healthcare services, including retraining of the workforce, retooling health services, and transforming workflows.

To leverage ML in mental healthcare systems, consumers must feel confident that their data is used ethically, and that ML will enhance their care experience. Consumers also need opportunities to develop digital skills and contribute to research. Ongoing communication between stakeholders is essential, and may generate a need for additional healthcare roles, such as digital navigators who would serve as an interface between the technology and the clinician and client (Koutsouleris et al., 2022).

CONCLUSION

ML has the potential to create a smarter, more adaptive mental healthcare system. The real-world implementation of ML tools is increasingly recognised as the solution to key issues in mental healthcare, such as delayed, inaccurate, and inefficient service delivery. Similarly, ML methodologies are becoming widely used in mental health research because they are well-placed to enhance our understanding of the biopsychosocial complexity of mental disorders, and hence to improve preventative, diagnostic, prognostic, and treatment frameworks.

However, the encouraging proof-of-concept studies to date do not yet translate to improved clinical practice. Currently, the actual impact of ML models is mostly speculative; in terms of effectiveness and relevance for mental health and use and acceptance by clinicians and clients. There are multiple interacting practical and ethical issues to resolve before AI can be clinically actionable for routine care. In addition to the cultural norms of clinical practice, key barriers include the availability and representativeness of training data, interpretability, risk of inaccuracies and bias, and practical implementation barriers.

Experts agree that the future role of ML in mental healthcare is not about replacing clinicians. Rather, it is about providing new insights and tools that can enhance their capabilities, including their efficiency and effectiveness. As AI and ML in mental health continues to evolve, it will be crucial for clinicians to stay abreast of the opportunities and risks it presents. Ongoing and in-depth training is needed to ensure the mental health workforce is adequately prepared to safely harness the benefits of AI and ML to enable more efficient and effective care and ensure a positive client experience.

FURTHER READING

AI terminology

The AI Terms Cheat Sheet (Marketing AI Institute, 2021)

Mental health apps and chatbots

‘They thought they were doing good but it made people worse’: Why mental health apps are under scrutiny (The Guardian, 2024)

Not all mental health apps are helpful. Experts explain the risks, and how to choose one wisely (The Conversation, 2023)

AI chatbots are still far from replacing human therapists (The Conversation, 2023)

Digital therapy

deprexis - overcome depression effectively

Working towards an AI-enabled healthcare system in Australia

Leveraging Digital Technology in Healthcare (Productivity Commission, 2024)

A Roadmap for AI in Healthcare for Australia (Australian Alliance for AI in Healthcare, 2021)

Safe and Responsible AI in Australia (Department of Industry, Science, & Resources, 2023)

APS Response to the Safe and Responsible AI in Australia Discussion Paper (Australian Psychological Society, 2023)

Other resources

The Tech Savvy Psych (Meagher & Cavenett, 2024)

Harnessing the Power of AI in Psychology (APS, 2024)

REFERENCES

Abd-Alrazaq, A., Alhuwail, D., Schneider, J., Toro, C. T., Ahmed, A., Alzubaidi, M., ... & Househ, M. (2022). The performance of artificial intelligence-driven technologies in diagnosing mental disorders: An umbrella review. *NPJ Digital Medicine*, *5*(1), 87. https://doi.org/10.1038/s41746-022-00631-8

Abd-Alrazaq, A., AlSaad, R., Harfouche, M., Aziz, S., Ahmed, A., Damseh, R., & Sheikh, J. (2023a). Wearable artificial intelligence for detecting anxiety: systematic review and meta-analysis. *Journal of Medical Internet Research, 25*, e48754. https://doi.org/10.2196/48754

Abd-Alrazaq, A., AlSaad, R., Shuweihdi, F., Ahmed, A., Aziz, S., & Sheikh, J. (2023b). Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. *NPJ Digital Medicine, 6*(1), 84. https://doi.org/10.1038/s41746-023-00828-5

Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. *Journal of Medical Internet Research, 22*(7), e16021. https://doi.org/10.2196/16021

Ahmed, A., Ramesh, J., Ganguly, S., Aburukba, R., Sagahyroon, A., & Aloul, F. (2022). Investigating the feasibility of assessing depression severity and valence-arousal with wearable sensors using discrete wavelet transforms and machine learning. *Information*, *13*(9), 406. https://doi.org/10.3390/info13090406

Al-Zaiti, S. S., Alghwiri, A. A., Hu, X., Clermont, G., Peace, A., Macfarlane, P., and Bond R. (2022). A clinician's guide to understanding and critically appraising machine learning studies: A checklist for Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML). *European Heart Journal - Digital Health, 3*(2), 125–140. https://doi.org/10.1093/ehjdh/ztac016

Antonucci, L. A., Penzel, N., Sanfelici, R., Pigoni, A., Kambeitz-Ilankovic, L., Dwyer, D., ... & PRONIA Consortium. (2022). Using combined environmental–clinical classification models to predict role functioning outcome in clinical high-risk states for psychosis and recent-onset depression. *The British Journal of Psychiatry, 220*(4), 229-245. https://doi.org/10.1192/bjp.2022.16

Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*, *145*, 137-165. https://doi.org/10.1016/j.neuroimage.2016.02.079

Ashar, Y. K., Clark, J., Gunning, F. M., Goldin, P., Gross, J. J., & Wager, T. D. (2021). Brain markers predicting response to cognitive-behavioral therapy for social anxiety disorder: an independent replication of Whitfield-Gabrieli et al. 2015. *Translational Psychiatry*, *11*(1), 260. https://doi.org/10.1038/s41398-021-01366-y

Australian Psychological Society (APS). (2024). Harnessing the power of AI in psychology. https://psychology.org.au/insights/harnessing-the-power-of-ai-in-psychology

Bălan, O., Moise, G., Moldoveanu, A., Leordeanu, M., & Moldoveanu, F. (2020). An investigation of various machine and deep learning techniques applied in automatic fear level detection and acrophobia virtual therapy. *Sensors, 20*(2), 496. https://doi.org/10.3390/s20020496

Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The power of generative ai: A review of requirements, models, input–output formats, evaluation metrics, and challenges. *Future Internet, 15*(8), 260. https://doi.org/10.3390/fi15080260

Barnett, I., Torous, J., Staples, P., Sandoval, L., Keshavan, M., & Onnela, J. P. (2018). Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology*, *43*(8), 1660-1666. https://doi.org/10.1038/s41386-018-0030-z

Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2022). The next generation: chatbots in clinical psychology and psychotherapy to foster mental health–a scoping review. *Verhaltenstherapie, 32*(Suppl. 1), 64-76. https://doi.org/10.1159/000501812

Birk, R. H., & Samuel, G. (2022). Digital phenotyping for mental health: reviewing the challenges of using data to monitor and predict mental health problems. *Current Psychiatry Reports, 24*(10), 523-528. https://doi.org/10.1007/s11920-022-01358-9

Biswas, A., Talukdar, W. (2024). Intelligent clinical documentation: Harnessing generative AI for patient-centric clinical note generation. *International Journal of Innovative Science and Research Technology, 9*(5), 994-1008. https://doi.org/10.38124/ijisrt/IJISRT24MAY1483

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

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., ... & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry, 20*(2), 154-170. https://doi.org/10.1002/wps.20882

Coghlan, S., Leins, K., Sheldrick, S., Cheong, M., Gooding, P., & D'Alfonso, S. (2023). To chat or bot to chat: Ethical issues with using chatbots in mental health. *Digital Health, 9*, 20552076231183542. https://doi.org/10.1177/20552076231183542

Coiera, E., & Liu, S. (2022). Evidence synthesis, digital scribes, and translational challenges for artificial intelligence in healthcare. *Cell Reports Medicine*, *3*(12). https://doi.org/10.1016/j.xcrm.2022.100860

Coiera, E., Magrabi, F., Hansen, D., & Verspoor, K. (2023). A National Policy Roadmap for Artificial Intelligence in Healthcare. https://aihealthalliance.org/wp-content/uploads/2023/11/AAAiH_NationalPolicyRoadmap_FINAL.pdf

Cuthbert, B. N. (2020). The role of RDoC in future classification of mental disorders. *Dialogues in Clinical Neuroscience, 22*(1), 81-85. https://doi.org/10.31887/DCNS.2020.22.1/bcuthbert

Dogrucu, A., Perucic, A., Isaro, A., Ball, D., Toto, E., Rundensteiner, E. A., ... & Boudreaux, E. (2020). Moodable: On feasibility of instantaneous depression assessment using machine learning on voice samples with retrospectively harvested smartphone and social media data. *Smart Health*, *17*, 100118. https://doi.org/10.1016/j.smhl.2020.100118

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*, 91-118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Eisenstadt, M., Liverpool, S., Infanti, E., Ciuvat, R. M., & Carlsson, C. (2021). Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. *JMIR Mental Health, 8*(11), e31170. https://doi.org/10.2196/31170

Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, *630*(8017), 625-630. https://doi.org/10.1038/s41586-024-07421-0

Ferrante, M., Redish, A. D., Oquendo, M. A., Averbeck, B. B., Kinnane, M. E., & Gordon, J. A. (2019). Computational psychiatry: a report from the 2017 NIMH workshop on opportunities and challenges. *Molecular Psychiatry, 24*(4), 479. https://doi.org/10.1038/s41380-018-0063-z

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), e13216. https://doi.org/10.2196/13216

Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., & Jeste, D. V. (2019). Artificial intelligence for mental health and mental illnesses: An overview. *Current Psychiatry Reports*, *21*, 1-18. https://doi.org/10.1007/s11920-019-1094-0

Haroz, E. E., Walsh, C. G., Goklish, N., Cwik, M. F., O'Keefe, V., & Barlow, A. (2020). Reaching those at highest risk for suicide: development of a model using machine learning methods for use with Native American communities. *Suicide and Life-Threatening Behavior, 50*(2), 422-436. https://doi.org/10.1111/sltb.12598

He, Y., Yang, L., Qian, C., Li, T., Su, Z., Zhang, Q., & Hou, X. (2023). Conversational agent interventions for mental health problems: systematic review and meta-analysis of randomized controlled trials. *Journal of Medical Internet Research*, *25*, e43862. https://doi.org/10.2196/43862

Higgins, O., Short, B. L., Chalup, S. K., & Wilson, R. L. (2023). Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. *International Journal of Mental Health Nursing, 32*(4), 966-978. https://doi.org/10.1111/inm.13114

Islam, M. R., Sakib, M. K. H., Ulhaq, A., Akter, S., Zhou, J., & Asirvathamt, D. (2023). Sidvis: Designing visual interactive system for analyzing suicide ideation detection. In *27th International Conference Information Visualisation* (pp. 384-389). IEEE. https://doi.org/10.1109/IV60283.2023.00071

Jabali, A. K., Waris, A., Khan, D. I., Ahmed, S., & Hourani, R. J. (2022). Electronic health records: Three decades of bibliometric research productivity analysis and some insights. *Informatics in Medicine Unlocked*, *29*, 100872. https://doi.org/10.1016/j.imu.2022.100872

Kessler, R. C., Bossarte, R. M., Luedtke, A., Zaslavsky, A. M., & Zubizarreta, J. R. (2020). Suicide prediction models: A critical review of recent research with recommendations for the way forward. *Molecular Psychiatry*, *25*(1), 168-179. https://doi.org/10.1038/s41380-019-0531-0

Kirtley, O. J., van Mens, K., Hoogendoorn, M., Kapur, N., & de Beurs, D. (2022). Translating promise into practice: a review of machine learning in suicide research and prevention. *The Lancet Psychiatry*, *9*(3), 243-252. https://doi.org/10.1016/S2215-0366(21)00254-6

Koutsouleris, N., Dwyer, D. B., Degenhardt, F., Maj, C., Urquijo-Castro, M. F., Sanfelici, R., ... & PRONIA Consortium. (2021). Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry, 78*(2), 195-209. https://doi.org/10.1001/jamapsychiatry.2020.3604

Koutsouleris, N., Hauser, T. U., Skvortsova, V., & De Choudhury, M. (2022). From promise to practice: Towards the realisation of AI-informed mental health care. *The Lancet Digital Health*, *4*(11), e829-e840. https://doi.org/10.1016/S2589-7500(22)00153-4

Koutsouleris, N., Kambeitz-Ilankovic, L., Ruhrmann, S., Rosen, M., Ruef, A., Dwyer, D. B., ... & Pronia Consortium. (2018). Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. *JAMA Psychiatry*, *75*(11), 1156-1172. https://doi.org/10.1001/jamapsychiatry.2018.2165

Li, H., Zhang, R., Lee, Y. C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *NPJ Digital Medicine, 6*(1), 236. https://doi.org/10.1038/s41746-023-00979-5

Liu, Y. S., Chokka, S., Cao, B., & Chokka, P. R. (2021). Screening for bipolar disorder in a tertiary mental health centre using EarlyDetect: A machine learning-based pilot study. *Journal of Affective Disorders Reports, 6*, 100215. https://doi.org/10.1016/j.jadr.2021.100215

Liu, Y., Hankey, J., Cao, B., & Chokka, P. (2021). Screening for major depressive disorder in a tertiary mental health centre using EarlyDetect: A machine learning-based pilot study. *Journal of Affective Disorders Reports*, *3*, 100062. https://doi.org/10.1016/j.jadr.2020.100062

Love, A. (2018). The diagnostic dilemma. *InPsych, 40*(1). https://psychology.org.au/for-members/publications/inpsych/2018/feb/the-diagnostic-dilemma

Lucas, G. M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., & Morency, L. P. (2017). Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI, 4*, 51. https://doi.org/10.3389/frobt.2017.00051

Luxton, D. D. (2014). Artificial intelligence in psychological practice: Current and future applications and implications. *Professional Psychology: Research and Practice, 45*, 332-339. https://doi.org/10.1037/a0034559

Martinez-Martin, N., & Kreitmair, K. (2018). Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. *JMIR Mental Health, 5*(2), e9423. https://doi.org/10.2196/mental.9423

Meyerhoff, J., Liu, T., Kording, K. P., Ungar, L. H., Kaiser, S. M., Karr, C. J., & Mohr, D. C. (2021). Evaluation of changes in depression, anxiety, and social anxiety using smartphone sensor features: longitudinal cohort study. *Journal of Medical Internet Research*, *23*(9), e22844. https://doi.org/10.2196/22844

Moura, I., Teles, A., Viana, D., Marques, J., Coutinho, L., & Silva, F. (2023). Digital phenotyping of mental health using multimodal sensing of multiple situations of interest: A systematic literature review. *Journal of Biomedical Informatics*, *138*, 104278. https://doi.org/10.1016/j.jbi.2022.104278

Nemesure, M. D., Heinz, M. V., Huang, R., & Jacobson, N. C. (2021). Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. *Scientific Reports*, *11*(1), 1980. https://doi.org/10.1038/s41598-021-81368-4

Office of the Australian Information Commissioner (OAIC). (2019). Guide to Health Privacy. https://www.oaic.gov.au/__data/assets/pdf_file/0011/2090/guide-to-health-privacy.pdf

Oyebode, O., Alqahtani, F., & Orji, R. (2020). Using machine learning and thematic analysis methods to evaluate mental health apps based on user reviews. *IEEE Access*, *8*, 111141-111158. https://doi.org/10.1109/ACCESS.2020.3002176

Rahman, M. A., Brown, D. J., Mahmud, M., Harris, M., Shopland, N., Heym, N., ... & Lewis, J. (2023). Enhancing biofeedback-driven self-guided virtual reality exposure therapy through arousal detection from multimodal data using machine learning. *Brain Informatics, 10*(1), 14. https://doi.org/10.1186/s40708-023-00193-9

Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science, 6*(1), 15. https://doi.org/10.1140/epjds/s13688-017-0110-z

Richter, T., Fishbain, B., Fruchter, E., Richter-Levin, G., & Okon-Singer, H. (2021). Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders. *Journal of Psychiatric Research*, *141*, 199-205. https://doi.org/10.1016/j.jpsychires.2021.06.044

Roberts, L. W., Chan, S., & Torous, J. (2018). New tests, new tools: mobile and connected technologies in advancing psychiatric diagnosis. *NPJ Digital Medicine*, *1*(1), 20176. https://doi.org/10.1038/s41746-017-0006-0

Rosellini, A. J., Liu, S., Anderson, G. N., Sbi, S., Tung, E. S., & Knyazhanskaya, E. (2020). Developing algorithms to predict adult onset internalizing disorders: An ensemble learning approach. *Journal of Psychiatric Research, 121*, 189-196. https://doi.org/10.1016/j.jpsychires.2019.12.006

Roy, A., Nikolitch, K., McGinn, R., Jinah, S., Klement, W., & Kaminsky, Z. A. (2020). A machine learning approach predicts future risk to suicidal ideation from social media data. *NPJ Digital Medicine*, *3*(1), 1-12. https://doi.org/10.1038/s41746-020-0287-6

Rozek, D. C., Andres, W. C., Smith, N. B., Leifker, F. R., Arne, K., Jennings, G., ... & Rudd, M. D. (2020). Using machine learning to predict suicide attempts in military personnel. *Psychiatry Research, 294*, 113515. https://doi.org/10.1016/j.psychres.2020.113515

Schnack, H. G., Van Haren, N. E., Brouwer, R. M., Evans, A., Durston, S., Boomsma, D. I., ... & Hulshoff Pol, H. E. (2015). Changes in thickness and surface area of the human cortex and their relationship with intelligence. *Cerebral Cortex, 25*(6), 1608-1617. https://doi.org/10.3389/fpsyt.2016.00050

Schwartz, B., Cohen, Z. D., Rubel, J. A., Zimmermann, D., Wittmann, W. W., & Lutz, W. (2021). Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. *Psychotherapy Research, 31*(1), 33-51. https://doi.org/10.1080/10503307.2020.1769219

Shatte, A. B., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*, *49*(9), 1426-1448. https://doi.org/10.1017/S0033291719000151

Su, C., Aseltine, R., Doshi, R., Chen, K., Rogers, S. C., & Wang, F. (2020). Machine learning for suicide risk prediction in children and adolescents with electronic health records. *Translational Psychiatry, 10*(1), 413. https://doi.org/10.1038/s41398-020-01100-0

Su, C., Xu, Z., Pathak, J., & Wang, F. (2020). Deep learning in mental health outcome research: A scoping review. *Translational Psychiatry*, *10*(1), 116. https://doi.org/10.1038/s41398-020-0780-3

Swaminathan, A., López, I., Mar, R. A. G., Heist, T., McClintock, T., Caoili, K., ... & Nock, M. K. (2023). Natural language processing system for rapid detection and intervention of mental health crisis chat messages. *NPJ Digital Medicine, 6*(1), 213. https://doi.org/10.1038/s41746-023-00951-3

Sweeney, C., Ennis, E., Mulvenna, M. D., Bond, R., & O'Neill, S. (2024). Insights derived from text-based digital media, in relation to mental health and suicide prevention, using data analysis and machine learning: systematic review. *JMIR Mental Health, 11*, e55747. https://doi.org/10.2196/55747

Taliaz, D., Spinrad, A., Barzilay, R., Barnett-Itzhaki, Z., Averbuch, D., Teltsh, O., ... & Lerer, B. (2021). Optimizing prediction of response to antidepressant medications using machine learning and integrated genetic, clinical, and demographic data. *Translational Psychiatry*, *11*(1), 381. https://doi.org/10.1038/s41398-021-01488-3

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 (TOCHI)*, *27*(5), 1-53. https://doi.org/10.1145/3398069

Wenzel, J., Haas, S. S., Dwyer, D. B., Ruef, A., Oeztuerk, O. F., Antonucci, L. A., ... & Kambeitz-Ilankovic, L. (2021). Cognitive subtypes in recent onset psychosis: distinct neurobiological fingerprints?. *Neuropsychopharmacology*, *46*(8), 1475-1483. https://doi.org/10.1038/s41386-021-00963-1

Wiebe, A., Aslan, B., Brockmann, C., Lepartz, A., Dudek, D., Kannen, K., ... & Braun, N. (2023). Multimodal assessment of adult attention-deficit hyperactivity disorder: A controlled virtual seminar room study. *Clinical Psychology & Psychotherapy, 30*(5), 1111-1129. https://doi.org/10.1002/cpp.2863

Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: A narrative review. *NPJ Digital Medicine*, *5*(1), 46. https://doi.org/10.1038/s41746-022-00589-7