International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE	ARTIFICIAL INTELLIGENCE IN PSYCHIATRIC DIAGNOSIS: CURRENT APPLICATIONS, CLINICAL POTENTIAL, AND EMERGING CHALLENGES - A COMPREHENSIVE REVIEW
DOI	https://doi.org/10.31435/ijitss.3(51).2026.6118
RECEIVED	13 June 2026
ACCEPTED	15 August 2026
PUBLISHED	22 August 2026
LICENSE	The article is licensed under a **Creative Commons Attribution 4.0 International License.**

ARTIFICIAL INTELLIGENCE IN PSYCHIATRIC DIAGNOSIS: CURRENT APPLICATIONS, CLINICAL POTENTIAL, AND EMERGING CHALLENGES - A COMPREHENSIVE REVIEW

Rafał Marcin Wojtan *(Corresponding Author, Email: rwojtan2@gmail.com)*
Nikolay Pirogov Specialized District Hospital, Lodz, Poland
ORCID ID: 0009-0000-9647-2245

Anna Maria Szeszko
University Clinical Hospital No. 2, Medical University of Lodz, Poland
ORCID ID: 0009-0008-2062-2188

Weronika Maria Czernek
University Clinical Hospital No. 2, Medical University of Lodz, Poland
ORCID ID: 0009-0009-0457-8970

Magdalena Cichorzewska
University Clinical Hospital No. 2, Medical University of Lodz, Poland
ORCID ID: 0009-0005-0701-8742

Martyna Płachta
University Clinical Hospital No. 2, Medical University of Lodz, Poland
ORCID ID: 0009-0008-2471-598X

Małgorzata Lubowiecka
Military Clinical Hospital and Polyclinic IPHC, Wroclaw, Poland
ORCID ID: 0009-0005-0424-2095

ABSTRACT

Artificial intelligence is being considered as a means of assisting psychiatric diagnosis, risk prediction, and long-term monitoring. Such applications in psychiatry have clinical relevance due to the fact that most psychiatric diagnosis is still based on interview, observation, and self-report, and also because it relies on symptom-based categorization that suffers from overlapping symptoms, tardiness in detection of illness, and discrepancies among patients given a single diagnosis. In this narrative review, we discuss applications of artificial intelligence in psychiatric assessment including machine learning, deep learning, natural language processing, digital phenotyping, large language models, and multimodal modelling. Papers published predominantly from 2020-2025 were included, with preference given to systematic reviews and meta-analyses, multicenter trials, and clinically significant publications.
It is evident that various computational models can detect valuable signatures in data obtained from electronic health records, neuroimaging, speech, clinical text, smartphone usage, wearable sensors, and social media. Such techniques can aid in the identification of depression, schizophrenia, bipolar disorder, anxiety disorders, and suicide risk. Multimodal approaches are particularly interesting due to the fact that they take the biological, psychological and social aspects of mental illness into consideration. On the other hand, the domain is currently suffering from insufficient and non-representative data sets, over-fitting, poor external validation, low interpretability, privacy issues, algorithmic bias, and unclear regulations. These aspects suggest that AI should be viewed as a supportive tool and not a replacement for the clinician.

KEYWORDS

Artificial Intelligence, Psychiatry, Machine Learning, Precision Psychiatry, Mental Health, Clinical Decision Support

CITATION

Rafał Marcin Wojtan, Anna Maria Szeszko, Weronika Maria Czernek, Magdalena Cichorzewska, Martyna Płachta, Małgorzata Lubowiecka. (2026). Artificial Intelligence in Psychiatric Diagnosis: Current Applications, Clinical Potential, and Emerging Challenges - A Comprehensive Review. *International Journal of Innovative Technologies in Social Science*. 3(51). doi: 10.31435/ijitss.3(51).2026.6118

1. Introduction

Mental disorders are one of the largest causes of morbidity and early death. Depression, anxiety, bipolar, schizophrenia, substance abuse disorders and suicide impact on all ages, social class and countries. It has been reinforced by the World Health Organization (2022) that mental health is a component of health, but it also represents social participation, social production and social welfare. In fact the impacts of mental disorders are more extensive than simply psychological suffering. Psychiatric morbidity is associated with disruption to education, social exclusion, unemployment, familial tension, social stigma, physical morbidity, shorter life span, etc.

The process of psychiatric diagnosis is unique and distinct from diagnosis in many other medical disciplines. In infectious disease, cardiology or oncology, the diagnosis is often guided by findings on a variety of different forms of investigation, for example, results of laboratory investigations, radiology, histopathology or molecular investigations. Psychiatry has moved a great deal in its approach to diagnosis, but the routine practice is still largely based on a clinical interview, behavioral observations and standardized symptom criteria. While diagnostic classifications like DSM-5-TR and ICD-11 provide a shared language and enhance diagnostic reliability, they do not completely solve the problem of heterogeneity (American Psychiatric Association, 2022; World Health Organization, 2019). Two patients receiving a diagnosis of major depressive disorder could appear very dissimilar from a clinical perspective; one exhibiting insomnia, guilt and psychomotor retardation, the other suffering from hypersomnia, irritability and cognitive slowness. The diagnostic label is similar but the underlying mechanisms and the optimal route to treatment may be vastly different.

This heterogeneity leads to a range of problems in practice. Symptoms like sleep disturbance, fatigue, concentration problems, and emotional dysregulation can occur in depression, anxiety, bipolar disorder, post traumatic stress disorder and psychosis. The presence of comorbid conditions is very common and the presenting symptoms often change over time. The diagnosis therefore needs to take into account a comprehensive history, developmental issues, current stress and the patient's cultural context and chronological evolution. There may be disagreement even among senior psychiatrists during early or difficult or unusual presentations. Delay in diagnosis can be costly, in the case of psychosis or bipolar disorder the duration of untreated illness being shown to impact functioning and outcomes. The situation may be similar in risk assessment of suicide, where standard assessment tools usually show only moderate predictive ability and may not detect swift changes in risk.

These problems of unreliability have prompted interest in precision and computational psychiatry. The idea of precision psychiatry is to try to relate symptoms to biological, cognitive, behavioural and environmental data in order to make treatment and prevention more individually focused (Bzdok & Meyer-Lindenberg, 2021; Lin et al., 2023). Computational psychiatry uses formal modelling of behaviour, cognition, and neurobiology. These do not take the place of clinical judgment, but their use allows the incorporation of an objective, quantitative element into traditionally subjective aspects of psychiatric assessment.

One of the primary mechanisms by which this is being attempted is through Artificial intelligence. Applications of AI in medicine already exist across several specialties including radiology, pathology, cardiology, and oncology (Rajpurkar et al., 2022). The most salient field to study such methods, is psychiatry, as mental illnesses manifest in the form of language, behavior, social functioning, and longitudinal patterns. Machine learning may analyze structured and unstructured electronic health records. Deep learning, may analyze neuroimaging, speech and other signals. Natural language processing can read clinical notes and narratives. Digital phenotyping can analyze smartphone and wearable sensor data and turn them into correlates of sleep, activity, mobility, or communication.

However psychiatry is also arguably the most sensitive specialty in which to apply AI. Information about traumas, interpersonal relationships, substance use, suicidal behavior, psychosis, sexuality, domestic conflict, cognitive performance will likely exist in psychiatric data. The potential for the AI to mis-predict, thus resulting in stigma, unnecessary treatment or miss of a risk or treatment can be substantial. This also relies on relationships of trust and empathy, and the development of therapeutic relationships. As a result, psychiatric application of AI needs to be critically evaluated, and implemented via a human centered approach (Torous et al., 2021; Tavory, 2024).

This review provides a clinically oriented synthesis of current applications, opportunities and challenges in AI-assisted psychiatric diagnosis and assessment. It focuses on machine learning, depression, psychosis prediction, bipolar disorder, anxiety disorders, suicide risk, NLP, digital phenotyping, large language models (LLMs), multimodal modelling, explainability, ethics and implementation.

2. Materials and Methods

This article has been written in the format of a narrative review. The aim was to explore the use of AI in psychiatric diagnosis, assessment, prediction and monitoring rather than treatment effectiveness solely. Searching was conceptual using PubMed, Scopus, Web of Science and Google Scholar. Terms used included; artificial intelligence, machine learning, deep learning, psychiatry, psychiatric diagnosis, mental health, precision psychiatry, depression, schizophrenia, bipolar disorder, anxiety disorder, suicide risk, natural language processing, digital phenotyping, wearable sensors, large language models, explainable artificial intelligence and ethics.

Literature from 2020-2025 was prioritized although earlier studies are included if they are particularly fundamental, highly cited, or important methodologically. Systematic reviews, meta-analyses, scoping reviews, multicenter studies, longitudinal studies and highly cited conceptual papers are considered especially relevant. Studies that did not have diagnostic, predictive or monitoring relevance and solely looked at drug efficacy are excluded, as are studies where mental health was only a background issue, or methodologically unclear to allow synthesis.

A narrative review was considered to be most appropriate for the rapidly expanding, methodologically disparate literature; the intention is not to estimate an aggregated diagnostic accuracy for one condition or one type of model but to plot the themes important in the clinical field, compare general approaches and data types and explore recurring shortcomings and areas where implementation must be approached responsibly. Throughout this review the distinction is drawn between research promise and current clinical utility.

3. Results and Synthesis

The data indicate that AI systems are capable of identifying clinically important patterns in psychiatric data, but many are not yet close to routine autonomous use. A pervasive finding in the literature is that excellent performance in a structured research dataset does not necessarily translate into useful performance in everyday practice. A model trained to discriminate among existing diagnostic categories may simply replicate the biases of these categories; for example, a model able to distinguish moderate and severe depression from controls is not necessarily useful in the clinical task of distinguishing unipolar from bipolar depression, normal bereavement, adjustment disorder, or anxiety.

More plausible roles for AI will be to assist in dimensional and probabilistic assessment; for example, a system can generate a probability of risk or self-harm, signal atypical presentations, identify co-occurring symptoms or clusters, summarize change over time, or flag items warranting further clinical inquiry, rather than providing a single label. The more optimistic reviews suggest promising diagnostic and prognostic capabilities of psychiatric AI, while also pointing out areas for improvement in terms of validation, reporting, and clinical integration (Graham et al., 2020; Lee et al., 2021; Rony et al., 2025). The ultimate challenge therefore is no longer whether AI can find patterns in psychiatric data, but whether these patterns are valid, interpretable, fair and clinically beneficial to patients.

3.1 Machine Learning in Psychiatric Diagnosis

ML is the most researched AI approach to psychiatric diagnosis. Supervised algorithms (e.g., support vector machines, random forests, gradient boosting, neural networks) use labeled data to perform classifications or make predictions. Unsupervised ML approaches (e.g., clustering, dimensionality reduction) can be used to discover dimensions of symptoms, or sub-types of patients, where pre-determined labels may not be sufficient. Deep learning is a type of ML, which is well-suited to the types of inputs typically found in psychiatric assessment, such as imaging, speech and free text.

In psychiatric diagnosis ML has been used for electronic health records, psychometric scales, neuroimaging, speech, genetic markers, social media data, and wearable sensors. A key strength of ML lies in its capacity to evaluate the relationships, and non-linearities, between potentially thousands of different variables-a clinician may be considering a dozen during an assessment. This may be beneficial in psychiatry where the diagnoses, and risk, emerge from an interplay between vulnerability, development, symptoms, environment and behavior. In a systematic review of the field Shatte et al. (2019), Dwyer et al. (2021) and Le Glaz et al. (2021) show exponential growth in the application of ML and NLP for diagnosis, prognosis and clinical decision support, whilst simultaneously observing a large gap between algorithm development and clinical utility.

One of the biggest barriers is validation. Many psychiatric ML papers only train a model on part of one dataset and show good internal performance. A model trained this way could perform poorly when tested on data from another hospital, country, language group, or population, as the patterns of documentation, diagnostic tendencies, social determinants, comorbidities and data quality may differ. The external validation is therefore not a luxury for technical reasons; it is a necessity for clinical reasons. A model that performs well only in the set of data in which it was built is of limited practical use.

Overfitting is another common issue. It is typical for psychiatric datasets to be small relative to the number of features included, for example thousands of imaging-derived features from hundreds of participants in a neuroimaging study. Careful pre-processing, cross-validation and reporting are required to ensure a model learns a signal that reflects illness, rather than noise. This risk can be reduced by adhering to reporting standards such as TRIPOD and AI-specific clinical trial extensions (Collins et al., 2015; Liu et al., 2020).

Workflow also plays a role. A clinically valid model will be unusable if its output is delayed too much, disrupts a consultation, is not integrated into electronic records, or gives the clinician a suggestion that they cannot easily interpret. Tools in mental health are in the early days of being tested in a real world environment and more research needs to be done into their usability, use in shared decision-making and long term integration (Auf et al., 2025). Those tools most likely to be useful are those which are answering an explicit clinical question, provide reasoning for their outputs and seamlessly fit into the existing workflow.

3.2 Depression

One of the most frequent clinical issues investigated in AI is Major Depressive Disorder. One reason for this focus is the high prevalence and recurring nature of the condition, but also because depressive characteristics can be identified in speech, communication, movement, sleep, and psychomotor behaviour. These characteristics can be observed from clinical records, screening tests, voice samples, written statements, and data from smartphones, wearables, social media, and neuroimaging.

A key strand of the research revolves around language analysis. A key aspect of the depressed state involves negative emotional terms, the use of first person language, hopelessness, narrow vocabulary, and diminished consistency in the narrative. NLP can capture some of these aspects of language present in clinical records, interviews, therapy sessions, or written communications. A comprehensive survey of NLP approaches applied to depression screening by Teferra et al (2024) noted both the potential and the caveats: 'these tools show promise for screening or monitoring patients, but they should not be perceived as independent diagnostic systems', due to limitations in universality of language across cultures, education, ages and contexts.

The voice is another interesting area to monitor depression. Slower speech and pauses, lack of variability in pitch and reduced acoustic energy have all been related to depression. It may be most valuable to monitor changes in speech over time, perhaps signaling the need to follow up on the patient's mood or sleep, but language variations, medications, illness, the context of the recording and cultural factors may be confounders for a single, non-context specific signal.

Another aspect relevant to our digital behavior is smartphone/wearable data that can record behavior (e.g., movement, circadian regularity, communication, activity levels). If depression worsens, there may be decreased movement, more erratic circadian patterns, and more infrequent social interaction. However, these may not exclusively be associated with depression; staying home can be related to teleworking, caring responsibilities, physical illness, financial constraints or choice, and therefore the relevance of digital markers is to raise questions and invite dialogue, not establish a diagnosis in themselves.

Neuroimaging models are already employed for the classification of depression, or prediction of treatment response, and these are an area of active research (though neuroimaging is not always indicated for most depressive presentations given cost). In the short term, this research is valuable, and an extension of these studies is to investigate subtypes and response. In the long term, the most feasible clinical utility of AI in depression appears to be multimodal and longitudinal. Combined measures of symptom scales, prior episode history, current medication, activity, sleep, speech and clinical notes would allow for more frequent review and better relapse prevention.

3.3 Schizophrenia and Psychosis Prediction

AI researchers are focused heavily on schizophrenia and related psychotic disorders, since initial diagnosis impacts upon subsequent functioning. An extended period of untreated psychosis has been demonstrated to predict a poorer outcome, and intervention in clinical high-risk states may prevent further disability later on. Thus, AI algorithms have been created which aim to predict psychosis transition, classify schizophrenia from control groups, evaluate abnormalities in neuroimaging, and characterize language dysfunction.

The literature is particularly rich in its utilization of neuroimaging data. This may be structural magnetic resonance imaging, functional imaging, or diffusion measures which evaluate cortical thickness, gray matter volume, connectivity, or white matter structure. ML is capable of discerning patterns distributed across these variables that are undetectable with standard analytic techniques. Koutsouleris et al. (2021) demonstrated that multimodal ML workflow are capable of predicting transition to psychosis in high-risk patients; work in this area is critical for the identification and understanding of models built with disparate information integrated in a meaningful way.

The area of language analysis is quite pertinent to psychosis. The unusual associations, disorganization, tangentiality, and reduced semantic coherence seen in schizophrenia spectrum disorders are important features of the illness. NLP is capable of quantifying several aspects of language organization that a clinician uses qualitatively to determine thought processes in a mental status examination. This would be useful since language is easy, affordable, and clinically meaningful to collect. The risks associated with predicting psychosis are primarily ethical ones; the false positives may create unwarranted stigma or anxiety and treatment, whereas the false negatives may delay necessary intervention, meaning that such tools are not simply tested on the basis of accuracy.

It is important to remember the base rate problem too. The transition to psychosis is medically serious, and far from certain among the high risk. An adequate model even with excellent discrimination will still produce too many false positives where the event is rare. To demonstrate whether the model can facilitate better clinical care than current assessment processes it is necessary to evaluate its calibration and undertake a decision-curve analysis and ideally a prospective trial. Currently AI should be seen as an aid, at best, rather than a replacement to the specialist clinician.

3.4 Bipolar Disorder and Mood Instability

Bipolar disorder serves as a demonstration of the potential, and of the pitfalls, of AI in psychiatry. Most patients attend for assessment in a depressive episode and it can be challenging to differentiate unipolar from bipolar depression. This differentiation is important, as treatment approaches differ, and antidepressant monotherapy may be problematic for some patients with bipolar disorder. AI has been applied to attempt to differentiate bipolar illness from its unipolar equivalent using medical records, history, speech, sleep, activity and communications.

Mood episodes are unfolding, across days and weeks; mania, hypomania and depression develop with resultant changes in sleep, activity, socialization and communication. This has made bipolar disorder an obvious target for digital phenotyping. Smartphone or wearable monitoring could record diminished sleep duration, increased activity levels, change in communication or telephone use patterns and altered sleep-wake cycles. This could be used as an early warning sign in relapse prevention. An overview by Pan et al (2025) found a rapidly growing literature on using ML for bipolar diagnosis but pointed out weaknesses, many studies having small sample size, heterogeneous methods, poor external validation.

Safe clinical use is likely to be in a collaborative, opt-in manner. For instance a patient with diagnosed bipolar disorder could agree to share sleep and activity summaries with a clinic team. If the system detects a pattern of increased activity or decreased sleep suggestive of developing hypomania, this could trigger a discussion rather than an intervention. The patient can remain in control, the AI serving as an input to the clinical decision making process. It must be interpreted carefully however, increased activity may indicate work, exercise, travel or parenting, reduced sleep may be due to insomnia, stress or child care rather than mania.

3.5 Anxiety Disorders and Post-Traumatic Stress Disorder

AI research to date has addressed depression and psychosis to a greater degree than anxiety disorders despite their considerable prevalence. Physiological arousal, avoidance, attentional bias, poor sleep, speech and comorbidity with depression all represent different manifestations of anxiety. Wearables, ecological momentary assessment, speech, social media and electronic health records have all been employed to infer anxiety level or risk.

Wearables provide heart rate variability, sleep patterns and electrodermal activity but these are non-specific to anxiety. Autonomic arousal could represent a panic attack, a traumatic event, physical exertion, caffeine intake, medication side effect or an unrelated physical illness. The utility of AI in anxiety might therefore lie less in making diagnoses and more in monitoring changes, detecting triggers and contributing to treatment decisions.

Post traumatic stress disorder is a vital area of research where speech, facial expression, physiology and neuroimaging are studied using AI. It may provide potential support for screening in military, emergency, or primary care settings, but any such approach needs to be trauma informed. Automated querying or passive monitoring could prove distress-provoking if implemented inappropriately. More generally, transdiagnostic rather than category-specific classifications seem more achievable in the case of AI. Systems might identify patterns of distress, avoidance, sleep disorganization, cognition deficits, or functional loss across categories rather than attempting to force each presentation into one diagnostic box.

3.6 Natural Language Processing and Clinical Text

Language is fundamental to psychiatry. What patients tell us, how they tell us it, what they leave out and how that changes over time is key to diagnosis. Unusually for healthcare, mental healthcare has thus always used some form of NLP. Applications range from analyzing clinical notes to patient text messages to therapy transcripts, crisis hotline conversations, social media postings and speech-to-text interview transcripts. Le Glaz et al. (2021) conducted a review of ML and NLP in mental health including applications in diagnosis, prognosis, identification of risk factors, and treatment outcomes. Malgaroli et al. (2023) have proposed a conceptual framework for using NLP in mental health interventions, Scherbakov et al. (2025) have highlighted the importance of the social determinants of health for mental health NLP.

Clinical notes are rich with information but reflect clinician practice, templates and financial incentives to document. A model trained on clinical notes may identify correlations with patterns of service usage or the structure of documentation, rather than illness. For example, mentions of homelessness, legal problems or failure to take medication may be predictive of crisis, but can also be social determinants of health or simply how services record the presence of particular problems and resources, so not a purely biological signal. We need bias audits and subgroup evaluations.

In a routine clinical setting, summarization and retrieval are likely the most valuable uses. We might be able to summarize a patient's symptom history over years of treatment, medication changes and side effects, and relevant risks and social determinants of health. This might relieve cognitive burden and help with continuity of care. But poor summaries could miss key risk factors and the clinical implications can be great; a full human review will still be necessary. Transformer models have enhanced performance, but they do produce very convincing outputs that can be completely false; in psychiatry, where nuances of language are key to diagnosis and emotional states, this is a very significant limitation.

3.7 Large Language Models and Conversational Artificial Intelligence

Large language models have shaped public perceptions about AI in mental health. As psychiatry is a talking-based discipline, fluent conversational agents may appear clinically adept. Patients may utilize them for psychoeducation, emotional regulation, self-exploration, and support with deciding to seek professional help. Clinicians may use LLMs to draft progress notes, summarize charts, and supplement their teaching. However, fluency is not a proxy for clinical competence.

Systematic reviews have demonstrated LLMs as a potential complement for screening, digital interventions, and clinical tasks, while still noting the risks of unsupervised clinical application (Guo et al., 2024; Omar et al., 2024). LLMs may confabulate clinically dangerous content, provide unsafe advice, fail to detect suicidal crises, and reinforce patients' beliefs about maladaptive cognitions. However, conversational agents have already existed in mental health care prior to the advent of LLMs, with prior reviews having pointed to both advantages (access) and disadvantages of chatbot interventions (Vaidyam et al., 2021). While more novel LLM-based agents may be more appealing due to their more human-like responses, this naturalness could lead to increased dependency and over-reliance on the agent.

Responsible applications should be strictly limited to tasks such as psychoeducation, preparation for sessions, symptom monitoring, assistance with administrative tasks, and clinician-reviewed summaries of interactions, rather than to independent diagnoses, management of suicidality, recommendations about medication adjustment, or substitution of psychotherapy. Any clinical LLMs will need clear limitations, privacy guarantees, robust emergency escalation mechanisms, and ongoing evaluation. Regulation is also difficult as most general-purpose LLMs were not intended to function as medical devices even as they are employed by patients in this capacity.

3.8 Digital Phenotyping and Wearable Technologies

Digital phenotyping involves capturing repeated or continuous measurements of behavior and physiology via personal digital devices. Within psychiatry it could potentially address the disconnect between intermittent appointments and the dynamic nature of symptomatology in everyday life. Clinicians and researchers would move away from just asking patients about symptoms retrospectively and instead be able to assess their sleep, mobility, communication, typing, voice, app usage and physiological patterns over time.

The available reviews indicate that digital phenotyping is feasible, though still methodologically uneven. Bufano et al. (2023) performed a review of digital phenotyping for the purpose of relapse or worsening of symptomatology and found it promising yet preliminary, and Chia & Zhang (2022) describe the field as broad and in early development. More recent reviews emphasize the potential of smartphone and wearable data such as accelerometry, step counts, heart rate and sleep, for psychiatric monitoring (Jung et al., 2025; Zhang et al., 2025).

The attraction is clear: mental health symptoms change in real-life situations. While a client might present with what appears to be good adjustment in session, the reality might be sleep deprivation, isolation or panic attacks at home. Passive sensors could catch decline prior to a scheduled clinic visit. Loss of sleep and increased activity can signal mania in bipolar disorder; a decrease in activity and sleep cycle disruption can predict depression relapse; a change in communication patterns and social withdrawal can forewarn psychosis crisis.

Interpretation is complicated: reduced mobile phone usage may indicate withdrawal or it might mean the phone is broken, that an individual is trying to reduce distractions from their own device usage, or simply has an adjusted routine. Location data might indicate reduced mobility, but an individual's work, financial responsibilities or role as caregiver may be more explanatory. Wearable device usage is dependent on adherence, but physical illness, medication and skin color can also influence signal output and device usability is known to vary by maker. There are considerable ethical issues, and even with permission, constant monitoring may feel intrusive and patients worry about access to their data by insurers, employers, courts, or relatives. The most appropriate use seems to be patient driven opt-in for relapse prevention and collaborative self-management, whereby the patient controls their sharing of data and response protocols for any alarms triggered.

3.9 Suicide Risk Prediction

Suicide risk assessment is one of the most pressing priorities of psychiatry. The traditional approach, guided by the combination of clinical interview, prior self-harm, present suicidal ideation, risk factors, protective factors, and professional judgment is indispensable, but it does not make prediction feasible. Suicide is an infrequent event even in high-risk groups, and risk can shift rapidly. AI has been proposed for predicting suicide based on a wide array of data, including electronic health records, emergency room visits, psychiatric records, prescriptions, scales, clinical notes, social media posts and speech characteristics. Bernert et al. (2020) concluded after an extensive literature review on AI and suicide prevention research that although performance in some areas is encouraging, important limitations persist. Similar concerns were noted regarding ethics and implementation of AI for this purpose (D'Hotman and Loh, 2020).

Pigoni et al. (2024) have highlighted the need for models to focus on psychiatric populations, since general models might be strongly influenced by diagnosis as a proxy for risk, rather than more specific indicators. The attraction for ML is that suicidal risk is an interaction of past attempt, mood disorders, substance use, trauma, loneliness, pain, access to means, recent stresses, sleeplessness, and fluctuating affective states; interactions that ML excels at identifying. NLP could potentially capture risky language and digital phenotyping could track behavioral changes. However, this utility cannot be determined on the basis of discriminatory accuracy alone; a high area under the curve may translate into many false positives. If all high-risk alerts necessitate crisis intervention, resources may become overstretched and patients may be involuntarily detained. False negatives in the form of reassuring low-risk scores also pose a threat. For the time being, the most ethically viable use of AI for suicide prevention likely lies in triaging individuals in high-volume areas such as ERs, crisis hotlines and large medical facilities. However, these systems must be subject to external validation, calibrated properly, continuously evaluated for biases, and integrated within a clear escalation framework. AI might raise awareness about risk, but the moral burden of patient care must always be borne by human beings.

3.10 Multimodal Data Integration

Because psychiatric disorders are multi-dimensional, multimodal AI has become an important area. Typically, no single modality will provide a full representation. For example, neuroimaging may reveal aspects of brain structure or function, speech the patients' affect or cognition, wearables sleep and physical activity, records historical clinical data, and social data provide context. Multimodal AI attempts to provide a more complete representation of a patient by combining various sources.

There is a dual appeal: the clinical, and the statistical. Psychiatry has used a biopsychosocial model for decades but lack mechanisms to combine biological, psychological, and social information quantitatively; AI can facilitate this operationalization. Psychosis prediction, for example, might involve combining family history, cognitive performance, social function, speech coherence, and neuroimaging, and depression relapse the model of previous episodes, adherence, sleep, mobility and language patterns. Some literature suggest that multimodal or hybrid models are more effective than single modality ones at specific tasks (Rony et al., 2025), and that clinically relevant predictions can be made with a psychosis model, as demonstrated by Koutsouleris et al. (2021).

However, multimodal models also raise additional issues. Missing data are common; various modalities will be acquired at different points in time; preprocessing is a complex endeavor, and sensitive data sources combined will also carry increased privacy risks. Interpretability becomes more challenging as complexity increase, such that a model that integrates imaging, notes, symptoms, and continuous wearable data is difficult to be refuted or explained by clinicians. The ideal clinical model will likely be defined not by how much data it uses, but by whether it makes a better decision within the constraints of acceptable cost, burden, and risk.

3.11 Explainability, Trust and Accountability

Psychiatric AI is difficult without explainability. A clinician will require an explanation for a high-risk prediction or diagnostic uncertainty/ relapse, while patients will likely inquire as to how a conclusion was reached. Predictions that lack interpretability and opacity are highly problematic in terms of both ethical and legal responsibility, and explainable AI seeks to make the output of a model more readily comprehensible to humans. Typical explanation techniques are feature importance, LIME (Ribeiro et al., 2016), SHAP values (Lundberg and Lee, 2017), attention visualization and intrinsically interpretable models. In the domain of healthcare, explainability is predominantly viewed as a requirement for safety, trust and adoption (Sadeghi et al., 2024).

In psychiatry, an explanation must be clinically relevant: stating that a model gave weight to token embedding or imaging features does not add clinical value unless it can be related to clinical reasoning. Instead, an explanation is useful if it can tell the clinician that the risk associated with a patient is due to increased emergency room visits, poor sleep, medication discontinuation, recent suicide attempt and increasingly negative sentiment within notes. This can serve as a launching point for discussion. However, the explanation techniques can themselves be misleading, as post-hoc explanations merely attempt to model a black-box and feature importance can be unstable.

Explainability must thus be an inherent feature of design, validation, and clinical assessment rather than an superficial application of a post-hoc explanation. Accountability also needs clarification. Should a clinician heed the advice of a model and that advice leads to injury, who bears the responsibility-the clinician, the institution, the vendor or the regulator? If a clinician overrides an alert, the rationale must be recorded and justified based on clinical principles. Clinicians are not expected to defer to the algorithm; instead the purpose of transparent decision support is to enhance clinical care without abnegating professional judgment.

3.12 Ethics, Bias, Privacy and Regulation

Mental Health Data Privacy and AI A lot of things have been said about psychiatric AI – its potential benefits, its risks, and the ethical challenges it raises – but a crucial piece missing in much of this discussion is the aspect of psychiatric data privacy. This is the more so since, when dealing with psychiatric data, the stakes can become quite high. Suicidal intentions, traumatic memories, history of substance misuse, instances of psychosis, familial disputes, experiences of violence, issues relating to sex and sexuality and cases of social marginalization may form part of the information stored in patient data, a leak or mismanagement of which can have dire consequences for their safety, cause social stigma and discrimination, and potentially lead to legal action against patients and caregivers. As such, psychiatric data privacy measures should aim to do better than simply meeting bare legal minimums.

'Algorithmic bias', another well-documented problem with any AI systems is no less relevant here. Any AI system uses previous or training data, and data typically includes disparities that reflect inequalities. For example, Obermeyer et al (2019) had established that 'health algorithm encoding racial bias' occurs through cost indicators for identifying the poor, with similar risks being present if the 'frequency of service use', 'hospitalization', 'frequency of contact during crisis' or 'documented clinical information intensity' are taken as unbiased markers for the severity of illness. 'Language models' might be 'biased in dialect, education, culture, platform and other cultural indicators,' and digital phenotyping could discriminate against patients lacking access to 'smartphones or wearable technology' or the like, and most 'neuroimaging databases feature overrepresented samples of participants from high-income countries and from the research volunteer population.'

3.13 Tables

The following tables summarize major applications, data sources and implementation challenges discussed in this review.

Table 1. Major clinical applications of artificial intelligence in psychiatric diagnosis and assessment.

Clinical domain	Typical data sources	AI methods	Potential contribution	Main caution
Depression	Clinical scales, speech, text, wearables, smartphone behaviour	ML, DL, NLP, digital phenotyping	Screening, relapse monitoring, treatment-response prediction	Low specificity of behavioural markers
Psychosis and schizophrenia	MRI/fMRI, cognition, speech, EHR, functional data	ML, DL, NLP, multimodal fusion	Risk prediction and early detection	False positives and stigma
Bipolar disorder	Sleep, activity, communication, EHR, mood diaries	Digital phenotyping, time-series ML	Early warning signs of mania/depression	Context-dependent behavioural changes
Anxiety and PTSD	Physiological signals, text, ecological momentary assessment, wearables	ML, NLP, sensor fusion	Severity monitoring and trigger detection	Physiological signals are nonspecific
Suicide risk	EHR, crisis notes, prior attempts, social determinants, language	Predictive analytics, NLP	Triage and risk stratification	Base-rate problem and clinical accountability

Note. EHR = electronic health records; ML = machine learning; DL = deep learning; NLP = natural language processing.

Table 2. Data modalities used in AI-assisted psychiatric assessment.

Data modality	Examples	Strengths	Limitations
Electronic health records	Diagnoses, medications, visits, clinical notes	Longitudinal and routinely available	Documentation bias and missing context
Speech and language	Prosody, semantic coherence, sentiment	Clinically relevant and relatively low cost	Language and culture dependent
Neuroimaging	MRI, fMRI, DTI	Biological insight and subgroup discovery	Cost, small samples and limited routine use
Wearables	Sleep, steps, heart rate variability	Continuous real-world monitoring	Adherence, device variability and no specificity
Smartphone behaviour	Mobility, app use, communication patterns	Ecological and longitudinal	Privacy concerns and contextual ambiguity
Social media	Posts, engagement, network signals	Large-scale naturalistic language	Consent, representativeness and surveillance risks

Table 3. Implementation challenges and mitigation strategies.

Challenge	Why it matters in psychiatry	Mitigation strategy
Algorithmic bias	May worsen inequities in already stigmatized groups	Representative datasets, subgroup reporting, fairness audits
Limited explainability	Clinicians and patients need understandable reasons	Clinically meaningful explanations and calibrated outputs
Weak external validation	Models may fail outside the original dataset	Multicenter prospective validation
Privacy and consent	Mental health data reveal intimate information	Specific consent, data minimization, strong governance
Workflow disruption	Poorly integrated tools increase burden	Human-centered design and usability testing
Regulatory uncertainty	Consumer tools may be used clinically without oversight	Clear standards for risk, evidence and crisis pathways

4. Discussion

The literature reviewed indicated that AI does hold promise in improving psychiatric assessment, but not in replacing human clinicians. Its most current applications and likely future applications focus on decision support, risk stratification, symptom monitoring and augmentation of documentation. There seems to be evidence that for depression, psychosis, bipolar disorder, anxiety and risk for suicide, that AI can detect patterns of clinical significance, but a tougher question is if the use of these patterns can improve outcome, decrease harm and be incorporated into the current practice settings. One significant benefit of AI is its ability to integrate information across time and modality. Psychiatry has always been based on a longitudinal course, and with diagnosis relies on changes over time, context and development. AI might be able to capture this by using data from records, repeated symptom inventories, speech, sleep and behavior in between appointments.

This is particularly appealing for overstretched psychiatric services, but simply processing more data doesn't guarantee improved care and may instead produce an accumulation of bad data, non-representative datasets and indecipherable output, thus resulting in new forms of risk. There is a significant translational gap; many studies show good internal validity, yet lack prospect prospective external validation. Samples are small and non-representative; and many focus on the development of algorithms based on factors easy to extract over more complex diagnostic judgments clinicians actually make (for example, a tool separating a psychiatrically healthy individual from a psychiatric patient may not be helpful for differentiating from among patients that have depression, bipolar illness, trauma, or a combination thereof, from those with symptoms related to substance use).

Such clinical AI will need to be measured against realistic clinical choices, patient-centered outcome data will need to be the standard for its use. The most effective tools may be more mundane than transformative, assisting with following up patients lost to care, with summarizing lengthy, complex charts, with highlighting risk factors, or with suggesting bipolar disorder when a recurrent depression is present. This would be less impressive, yet more useful. The psychiatric profession should continue to champion the indispensable importance of human beings and human interaction in treatment, ensuring that human empathy, cultural understanding, trust and meaning cannot be ceded to the algorithm.

That AI must be a tool to allow humans to see what they may otherwise miss and to take on burdens currently faced by clinicians to reduce paperwork, while keeping the moral and clinical responsibility solely in human hands. The field needs to progress from proofs of concept to prospect prospective multi-center studies in diverse patient populations with an emphasis on patient-centered outcomes, clinician time saved, equity issues, patient and clinician experience of the AI and patient and community buy-in and acceptance in order to effectively implement this technology.

5. Practical Implications for Clinical Psychiatry

Clinicians should treat AI not as a "one size fits all" system but as a portfolio of dynamic tools. AI may be able to aid with diagnosis, but others might assist with monitoring, note-taking, patient triage, or education. Basic AI knowledge, including understanding how models are trained, populations in which models are tested, what uncertainty means and how bias can manifest, will need to be incorporated into psychiatric training.

Hospitals shouldn't implement AI tools purely on the basis that they are new. Purchasing should mandate: clinical utility evidence, external validation of models, governance of data used, examination of bias, and usability tests, along with post-implementation surveillance. Clinicians should participate in the design and testing of AI systems early on. Patients need to know when AI is utilized in the decision of what is clinically important, and particularly in contexts of potential harm to the patient such as suicide risk, diagnosis, or involuntary care.

More effort should be spent by researchers on generating more useful evidence instead of simply more models. Studies should explicitly mention: dataset properties, data missingness, data pre-processing, validation methodology, calibration, analysis of performance of certain subgroups, and clinical decision thresholds. Where possible, studies should also compare the AI to existing practice, and report net benefits. Policy makers must take measures for safety, privacy, crisis care, advertising of products, and clinical oversight, particularly for devices that bridge the gap between medical devices and well-being products.

6. Limitations of This Review

The limitations of this review: This is a narrative synthesis, not a formal systematic review or meta-analysis. Although a broad literature search was undertaken which prioritized up-to-date, clinically pertinent publications, there was not a reproducible PRISMA flow diagram, scoring of risk of bias or pooled statistical analysis. Thus, the conclusions should be regarded as a structured clinical summary rather than a firm quantitative estimate of treatment effects.

The landscape is changing rapidly: Findings about LLMs, digital phenotyping and regulation can be superseded quickly. There is limited up-to-date, robust clinical research and many studies to date are exploratory or rely on small participant samples. The vast majority of papers are written in English, often sourced from high-income countries. These findings may not be representative in terms of different cultures, systems and languages. The focus of this review has been on the diagnosis, assessment and monitoring applications of digital technologies. The aspects of selection of treatment, augmentation of psychotherapy, or digital therapeutics have been covered briefly.

7. Future Directions

Future AI in psychiatry is unlikely to be the realm of radical novelty and instead will focus on its appropriate implementation. To be clinical models need to be prospectively validated in practice; datasets need to be less Eurocentric, and instead language-, culture-, age- and class-, as well as healthcare system specific; and outputs need to convey uncertainty effectively in interpretable language.

Privacy-preserving methods (e.g. Federated learning) could allow institutes to work together without pooling sensitive patient data, yet still cannot eliminate privacy concerns and the risk of bias. The potential for LLMs to engender anthropomorphic relationships means they will require particular supervision, whilst systems based upon them should be task-specific, observable and with safety pathways built-in. Chatbots which are more general should not be promoted as a substitute for a psychiatrist.

An algorithmic psychiatrist is not a desirable endpoint. An intelligent, evolving mental health system where data is harnessed in the public good, where clinicians are appropriately supported, where patients maintain control, and where the therapeutic relationship remains paramount; seems achievable, and far more ethically desirable. AI has the potential to make psychiatry a more precise and reliable science; this however will depend on development guided by humility and an awareness of the intricacies of human suffering.

8. Conclusions

Artificial intelligence is one of the most promising areas in the current psychiatric literature. Through machine learning, natural language processing, digital phenotyping, multimodal modeling and large language models, there is now scope for predicting symptoms, estimating risk and assisting in clinical decisions. Early diagnosis, monitoring for relapse, suicide risk assessment, documentation and precision psychiatry research represent the areas where AI is most likely to make the greatest contribution to psychiatric care, at least at this early stage.

The current psychiatric use of AI is still relatively primitive and there are many caveats to the models produced. Models generally require further independent validation, clearer reporting, further work on interpretability and careful monitoring for bias. Issues such as privacy, autonomy, stigma, dependence and responsibility in mental health contexts are significant and the use of AI in psychiatry should primarily be thought of as an adjunct to psychiatrists, psychologists and therapeutic relationships, rather than a substitute for them. Applied thoughtfully, AI may increase the timeliness, individualization and accessibility of psychiatric care whilst still valuing clinical acumen and human compassion.

REFERENCES

1. American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). American Psychiatric Association Publishing.
2. Auf, H., Svedberg, P., Nygren, J., Nair, M., & Lundgren, L. (2025). The use of artificial intelligence in mental health services to support decision-making: Scoping review. *Journal of Medical Internet Research, 27*, e63548. https://doi.org/10.2196/63548
3. Bernert, R. A., Hilberg, A. M., Melia, R., Kim, J. P., Shah, N. H., & Abnousi, F. (2020). Artificial intelligence and suicide prevention: A systematic review of machine learning investigations. *International Journal of Environmental Research and Public Health, 17*(16), 5929. https://doi.org/10.3390/ijerph17165929
4. Bzdok, D., & Meyer-Lindenberg, A. (2021). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 6*(3), 223–230. https://doi.org/10.1016/j.bpsc.2020.09.003
5. Bufano, P., Iacob, G., Banfi, G., & Gualtieri, P. (2023). Digital phenotyping for monitoring mental disorders: Systematic review. *Journal of Medical Internet Research, 25*, e46778. https://doi.org/10.2196/46778
6. Chia, A. Z. R., & Zhang, M. W. B. (2022). Digital phenotyping in psychiatry: A scoping review. *Technology and Health Care, 30*(6), 1331–1342. https://doi.org/10.3233/THC-213648
7. Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). *Annals of Internal Medicine, 162*(1), 55–63. https://doi.org/10.7326/M14-0697
8. D'Hotman, D., & Loh, E. (2020). AI enabled suicide prediction tools: A qualitative narrative review. *BMJ Health & Care Informatics, 27*(3), e100175. https://doi.org/10.1136/bmjhci-2020-100175
9. Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2021). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology, 17*, 91–118. https://doi.org/10.1146/annurev-clinpsy-081219-092940
10. Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., & Jeste, D. V. (2020). Artificial intelligence for mental health and mental illnesses: An overview. *Current Psychiatry Reports, 22*, 116. https://doi.org/10.1007/s11920-020-01192-8
11. Guo, Z., Lai, A., Thygesen, J. H., Farrington, J., Keen, T., & Li, K. (2024). Large language models for mental health applications: Systematic review. *JMIR Mental Health, 11*, e57400. https://doi.org/10.2196/57400
12. Jung, H. W., Kim, D. Y., Lee, I., Kim, O., Lee, S., Lee, S., Chung, U. S., Kim, J.-H., Kim, S., Kim, J. W., Shin, A. L., & Lee, J. J. (2025). Key features of digital phenotyping for monitoring mental disorders: Systematic review. *Journal of Medical Internet Research, 27*, e77331. https://doi.org/10.2196/77331
13. Koutsouleris, N., Dwyer, D. B., Degenhardt, F., Maj, C., Urquijo-Castro, M. F., Sanfelici, R., Popovic, D., Oeztuerk, O., Haas, S. S., Weiske, J., Ruef, A., Kambeitz-Ilankovic, L., Antonucci, L. A., Neufang, S., Schmidt-Kraepelin, C., Ruhrmann, S., Penzel, N., Kambeitz, J., Haidl, T. K., ... Meisenzahl, E. (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
14. Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., Lenca, P., Billot, R., Ryan, T. C., Marsh, J., DeVylder, J., Walter, M., Berrouiguet, S., & Lemey, C. (2021). Machine learning and natural language processing in mental health: Systematic review. *Journal of Medical Internet Research, 23*(5), e15708. https://doi.org/10.2196/15708
15. Lee, E. E., Torous, J., De Choudhury, M., Depp, C. A., Graham, S. A., Kim, H.-C., Paulus, M. P., Krystal, J. H., & Jeste, D. V. (2021). Artificial intelligence for mental healthcare: Clinical applications, barriers, facilitators, and artificial wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 6*(9), 856–864. https://doi.org/10.1016/j.bpsc.2021.02.001

16. Lin, E., Tsai, S. J., & Kuo, P. H. (2023). Precision psychiatry applications of artificial intelligence and machine learning. *Journal of Affective Disorders, 324*, 300–312. https://doi.org/10.1016/j.jad.2022.12.041
17. Liu, X., Rivera, S. C., Moher, D., Calvert, M. J., & Denniston, A. K. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *Nature Medicine, 26*, 1364–1374. https://doi.org/10.1038/s41591-020-1034-x
18. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
19. Malgaroli, M., Hull, T. D., Zech, J. M., & Althoff, T. (2023). Natural language processing for mental health interventions: A systematic review and research framework. *Translational Psychiatry, 13*, 309. https://doi.org/10.1038/s41398-023-02592-2
20. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342
21. Omar, M., Soffer, S., Charney, A. W., Landi, I., Nadkarni, G. N., & Klang, E. (2024). Applications of large language models in psychiatry: A systematic review. *Frontiers in Psychiatry, 15*, 1422807. https://doi.org/10.3389/fpsyt.2024.1422807
22. Pan, Y., Wang, P., Xue, B., Liu, Y., Shen, X., Wang, S., & Wang, X. (2025). Machine learning for the diagnosis accuracy of bipolar disorder: A systematic review and meta-analysis. *Frontiers in Psychiatry, 15*, 1515549. https://doi.org/10.3389/fpsyt.2024.1515549
23. Pigoni, A., Delvecchio, G., Turtulici, N., Madonna, D., Pietrini, P., Cecchetti, L., & Brambilla, P. (2024). Machine learning and the prediction of suicide in psychiatric populations: A systematic review. *Translational Psychiatry, 14*, 140. https://doi.org/10.1038/s41398-024-02852-9
24. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine, 28*(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0
25. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778
26. Rony, M. K. K., Das, D. C., Khatun, M. T., Ferdousi, S., Akter, M. R., Khatun, M. A., Begum, M. H., Khalil, M. I., Parvin, M. R., Alrazeeni, D. M., & Akter, F. (2025). Artificial intelligence in psychiatry: A systematic review and meta-analysis of diagnostic and therapeutic efficacy. *Digital Health, 11*, 20552076251330528. https://doi.org/10.1177/20552076251330528
27. Sadeghi, Z., Alizadehsani, R., Cifci, M. A., Kausar, S., Rehman, R., Mahanta, P., Bora, P. K., Almasri, A., Alkhawaldeh, R. S., Hussain, S., Alatas, B., Shoeibi, A., Moosaei, H., Hladík, M., Nahavandi, S., & Pardalos, P. M. (2024). A review of explainable artificial intelligence in healthcare. *Computers and Electrical Engineering, 118*, 109370. https://doi.org/10.1016/j.compeleceng.2024.109370
28. Scherbakov, D. A., Hubig, N. C., Lenert, L. A., Alekseyenko, A. V., & Obeid, J. S. (2025). Natural language processing and social determinants of health in mental health research: AI-assisted scoping review. *JMIR Mental Health, 12*, e67192. https://doi.org/10.2196/67192
29. Shatte, A. B. R., 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
30. Tavory, T. (2024). Regulating AI in mental health: Ethics of care perspective. *JMIR Mental Health, 11*, e58493. https://doi.org/10.2196/58493
31. Teferra, B. G., Rueda, A., Pang, H., Valenzano, R., Samavi, R., Krishnan, S., & Bhat, V. (2024). Screening for depression using natural language processing: Literature review. *Interactive Journal of Medical Research, 13*, e55067. https://doi.org/10.2196/55067
32. Torous, J., Roberts, L. W., & Hsin, H. (2021). Ethical use of digital data and artificial intelligence in mental health. *Current Psychiatry Reports, 23*, 71. https://doi.org/10.1007/s11920-021-01275-5
33. Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2021). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *Canadian Journal of Psychiatry, 66*(1), 16–24. https://doi.org/10.1177/0706743720959636
34. World Health Organization. (2019). *International classification of diseases 11th revision (ICD-11).* https://icd.who.int/
35. World Health Organization. (2022). *World mental health report: Transforming mental health for all.* https://www.who.int/publications/i/item/9789240049338
36. Zhang, Y., Wang, J., Zong, H., Singla, R. K., Ullah, A., Liu, X., Wu, R., Ren, S., & Shen, B. (2025). The comprehensive clinical benefits of digital phenotyping: From broad adoption to full impact. *npj Digital Medicine, 8*, 196. https://doi.org/10.1038/s41746-025-01602-5