International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Scholarly Publisher
RS Global Sp. z O.O.
ISNI: 0000 0004 8495 2390

Dolna 17, Warsaw,
Poland 00-773
+48 226 0 227 03
editorial_office@rsglobal.pl

ARTICLE TITLE	ARTIFICIAL INTELLIGENCE IN MENTAL HEALTH SERVICES: CURRENT APPLICATIONS, CHALLENGES, AND FUTURE DIRECTIONS
DOI	https://doi.org/10.31435/ijitss.4(48).2025.4661
RECEIVED	17 November 2025
ACCEPTED	26 December 2025
PUBLISHED	30 December 2025
LICENSE	The article is licensed under a **Creative Commons Attribution 4.0 International License.**

ARTIFICIAL INTELLIGENCE IN MENTAL HEALTH SERVICES: CURRENT APPLICATIONS, CHALLENGES, AND FUTURE DIRECTIONS

Michał Szyszka *(Corresponding Author, Email: michal.m.szyszka@gmail.com)*
Lower Silesian Center for Oncology, Pulmonology and Hematology, Wrocław, Poland
ORCID ID: 0009-0002-5307-1753

Aleksandra Grygorowicz
Medical University of Warsaw, Warsaw, Poland
ORCID ID: 0009-0002-7729-8178

Klaudia Baran
Medical University of Warsaw, Warsaw, Poland
ORCID ID: 0009-0004-9599-792X

Michał Ględa
Medical University of Lodz, Lodz, Poland
ORCID ID: 0009-0003-8148-160X

Weronika Radecka
Cardinal Stefan Wyszyński University, Warsaw, Poland
ORCID ID: 0009-0007-1116-398X

Weronika Kozak
University of Warmia and Mazury, Olsztyn, Poland
ORCID ID: 0009-0009-5605-2794

Agnieszka Szreiber
University of Warmia and Mazury, Olsztyn, Poland
ORCID ID: 0009-0006-3432-8284

Karol Grela
Medical University of Warsaw, Warsaw, Poland

Karolina Nowacka
Medical University of Gdańsk, Gdańsk, Poland
ORCID ID: 0009-0004-4933-755X

Kamil Jabłoński
Medical University of Silesia, Katowice, Poland
ORCID ID: 0009-0003-4682-2405

Anna Woźniak
Lazarski University, Warsaw, Poland
ORCID ID: 0009-0000-5065-6313

ABSTRACT

Background: Artificial intelligence (AI) is increasingly integrated into mental health care, offering tools for assessment, monitoring, risk prediction, and intervention. Rising global mental health needs, clinician shortages, and advances in digital technologies have accelerated adoption of conversational agents, digital phenotyping, clinical decision-support systems, and large language models (LLMs). Despite substantial promise, concerns remain regarding bias, transparency, safety, and real-world effectiveness.
Methods: This narrative review synthesized peer-reviewed studies published between 2017 and early 2025. Searches were conducted in PubMed, PsycINFO, Scopus, and Google Scholar. Eligible sources included randomized controlled trials, systematic reviews, meta-analyses, observational studies, and major policy documents evaluating AI for mental health diagnosis, monitoring, intervention, or clinical decision support.
Results: Findings across more than 120 studies show that AI-based conversational agents provide modest but consistent improvements in symptoms of mild to moderate depression and anxiety. Diagnostic models and triage tools demonstrate potential for identifying psychosis risk, suicide risk, and treatment response, but external validity remains limited by dataset bias and variable performance in real-world settings. Digital phenotyping offers early-warning capabilities for relapse, while LLMs improve documentation efficiency but struggle with crisis detection and safety-sensitive reasoning. Ethical concerns—particularly relating to privacy, informed consent, explainability, and algorithmic fairness—remain widespread.
Conclusions: AI has significant potential to enhance mental health care through scalable interventions, improved diagnostic accuracy, and proactive monitoring. However, safe integration requires robust governance, transparency, and sustained human oversight. Future progress depends on large-scale clinical trials, bias mitigation, standardized evaluation frameworks, and the development of equitable hybrid human-AI care models.

KEYWORDS

Artificial Intelligence, Mental Health, Digital Phenotyping, Machine Learning in Psychiatry

CITATION

Michał Szyszka, Aleksandra Grygorowicz, Klaudia Baran, Michał Ględa, Weronika Radecka, Weronika Kozak, Agnieszka Szreiber, Karol Grela, Karolina Nowacka, Kamil Jabłoński, Anna Woźniak. (2025). Artificial Intelligence in Mental Health Services: Current Applications, Challenges, and Future Directions. *International Journal of Innovative Technologies in Social Science*. 4(48). doi: 10.31435/ijitss.4(48).2025.4661

1. Introduction

Mental health disorders represent one of the leading contributors to the global burden of disease, with depressive and anxiety disorders collectively affecting more than 600 million people worldwide (GBD Collaborative Network, 2023). Beyond individual suffering, mental illness has profound economic implications, including reduced productivity, increased healthcare expenditures, and long-term disability (Patel et al., 2018). The persistent treatment gap—where up to 75% of people with mental health conditions in low- and middle-income countries receive no care—highlights the urgent need for scalable, evidence-based solutions (Thornicroft et al., 2017).

Theoretical frameworks such as the stepped-care model and the digital mental health ecosystem model suggest that technology-enabled interventions can enhance mental health systems by improving early detection, triage efficiency, and low-intensity intervention delivery (Bower & Gilbody, 2005; Mohr et al., 2017). Artificial intelligence (AI) extends these frameworks by enabling continuous monitoring, predictive analytics, and adaptive personalization. AI tools—including machine learning (ML), deep learning, natural language processing (NLP), and large language models (LLMs)—are increasingly integrated across clinical, community, and population-level mental health applications. Their rise is facilitated by ubiquitous smartphone adoption, advances in sensor technology, and growing acceptance of telehealth and digital therapeutics (Torous et al., 2020).

Recent meta-analyses and policy reports emphasize that AI may serve as a transformative catalyst in mental health care by improving early symptom detection, augmenting clinician decision-making, and expanding preventive mental health strategies (He et al., 2019; Lawrence et al., 2024; Alhuwaydi et al, 2024).

Despite this promise, however, concerns persist regarding algorithmic fairness, explainability, data governance, and real-world effectiveness—particularly in sensitive domains such as suicidality and severe mental illness (Vokinger et al., 2021; Mennella et al., 2024). These challenges underscore the need for rigorous evaluation, ethical oversight, and multidisciplinary collaboration.

Global reports highlight that innovation—including digital and AI-enabled tools—is essential for addressing the treatment gap (Patel et al., 2018). AI technologies—including machine learning (ML), deep learning, natural language processing (NLP), and large language models (LLMs)—are emerging as promising tools capable of expanding assessment, intervention, and monitoring capabilities in mental health care, also showing promise among children and adolescents, who represent a rapidly growing at-risk population (Fernández-Batanero et al., 2025). The growth of AI in this field is supported by the proliferation of smartphones and wearables, improved computational infrastructures, and increasing acceptance of remote and digital therapeutic solutions (Torous et al., 2020).

Despite rapid advancements, the integration of AI presents ethical, methodological, and regulatory challenges. Issues such as algorithmic bias, lack of transparency, limited generalizability, and inconsistent performance in crisis scenarios pose significant obstacles (Char et al., 2020; Mennella et al., 2024). This review synthesizes current evidence across five core domains: conversational agents, diagnostic tools, risk prediction, digital phenotyping, and LLMs.

2. Methodology

A narrative review was conducted to synthesize peer-reviewed evidence on artificial intelligence (AI) in mental health from 2017 to February 2025. Searches were performed in PubMed, PsycINFO, Scopus, and Google Scholar using combinations of terms including: “artificial intelligence,” “machine learning,” “deep learning,” “natural language processing,” “large language models,” “digital mental health,” “chatbots,” “conversational agents,” “digital phenotyping,” “risk prediction,” “suicide,” “diagnosis,” and “clinical decision support.” Example search strings included ("artificial intelligence" AND "mental health") and ("chatbot" AND "depression").

Inclusion criteria were peer-reviewed empirical studies, systematic reviews, meta-analyses, and major policy documents examining AI for mental health diagnosis, monitoring, intervention, or risk prediction. Exclusion criteria were non-English publications, non–peer-reviewed sources, and technical papers without clinical relevance.

Data from eligible studies were synthesized thematically across five domains: conversational agents, diagnostic and triage tools, risk-prediction models, digital phenotyping, and large language models.

3.Findings

3.1. Conversational Agents and Digital Therapeutics

New frameworks now offer clear guidance on how to create safe and effective digital mental health tools (Mertens & Van Gelder, 2024). Conversational agents—such as Woebot, Wysa, and Youper—deliver psychological support through NLP-driven dialogue. Randomized controlled trials (RCTs) demonstrate reductions in depressive and anxiety symptoms following engagement with AI-guided cognitive behavioral therapy and supportive interventions (Fitzpatrick et al., 2017). Empathy-driven conversational platforms further enhance engagement and satisfaction (Inkster et al., 2018), while systematic reviews show consistent, albeit modest, treatment effects (Abd-Alrazaq et al., 2020). Early evidence suggests that users may disclose sensitive information more readily to AI systems than to humans (Miner et al., 2017).

Despite promising outcomes, limitations include reduced effectiveness for severe mental illness, emotional dissatisfaction due to scripted responses, and safety concerns regarding inadequate crisis detection (Torous et al., 2020).

3.2. Diagnostic Aids and Triage Tools

ML-based diagnostic systems analyze speech, facial expressions, linguistic coherence, and behavioral signals to detect psychiatric conditions. Speech and acoustic markers have been successfully correlated with depression and suicide risk (Cummins et al., 2015). Triage algorithms can assist clinicians by prioritizing high-risk cases based on historical and clinical data (Walsh et al., 2017). AI-driven predictive models have also demonstrated utility in early-stage diagnostic prediction for psychotic disorders (Koutsouleris et al., 2021).

However, diagnostic accuracy often declines when tools are tested outside development datasets, reflecting weak external validity (Kirtley et al., 2022). Algorithmic bias further threatens equitable diagnostic performance (Char et al., 2020). Many systems remain unregulated, raising concerns about clinical reliability (Mennella et al., 2024).

3.3. Risk Prediction: Suicide, Self-Harm, and Relapse

AI-based risk prediction tools attempt to forecast suicide attempts, self-harm, or psychiatric relapse. Experimental evaluations have shown substantial performance variability among mental health chatbots when detecting suicidal ideation (Pichowicz et al., 2025). Large datasets have enabled moderate predictive accuracy in controlled environments (Walsh et al., 2017). Yet systematic reviews show that many risk-prediction systems fail to outperform simple statistical baselines in real-world practice (Belsher et al., 2019).

False positives risk overwhelming crisis services, while false negatives carry severe safety implications. Ethical questions include autonomy, consent, and the potential reinforcement of structural biases embedded in clinical data (Kirtley et al., 2022).

3.4. Digital Phenotyping and Passive Monitoring

Digital phenotyping leverages smartphone and wearable sensors to monitor behavioral changes, mobility patterns, communication frequency, circadian rhythms, and physiological markers. Such metrics can predict upcoming depressive or manic episodes (Onnela & Rauch, 2016). Reviews highlight the potential of digital phenotyping for early relapse detection across mood and psychotic disorders (Maatoug et al., 2022). However, ethical challenges include privacy, surveillance, unclear consent mechanisms, and proprietary data governance (Huckvale et al., 2019).

3.5. Large Language Models (LLMs)

LLMs are increasingly used for clinical documentation, note summarization, patient education, chat-based support, and administrative automation. They reduce clinician burden by streamlining documentation workflows (Lawrence et al., 2024). However, LLMs pose considerable risks: hallucinations, inconsistent reasoning, missed crisis cues, and embedded biases (Torous et al., 2020). Clinical use necessitates strong oversight to ensure accuracy, transparency, and patient safety (Char et al., 2020).

Recent studies further highlight both opportunities and limitations. In 2023 and 2024, large-scale evaluations demonstrated that LLMs show strong performance in tasks such as clinical summarization and diagnostic reasoning but struggle significantly with crisis-language detection, differential diagnosis involving severe mental illness, and generating safety-sensitive recommendations (Nori et al., 2023; Kim et al., 2025). In mental health–specific tasks, LLMs were shown to provide coherent psychoeducational information but often failed to maintain therapeutic boundaries or detect nuanced emotional content (Kim et al., 2025). These issues underscore the need for specialized fine-tuning, guardrails, and human oversight in mental-health deployments.

3.6. AI-Enabled Clinical Decision Support Systems

AI-powered clinical decision support systems (CDSS) have been increasingly integrated into psychiatric assessments, pharmacotherapy decisions, and treatment planning. A 2023 review found that CDSS enhanced clinician accuracy in diagnosing depressive and anxiety disorders and improved treatment adherence when integrated into electronic health records (Cruz-Gonzalez et al., 2025). Additionally, AI-assisted medication management tools show promise in predicting antidepressant response and optimizing treatment sequencing based on genetic, behavioral, and clinical data (Jaworska et al., 2019).

Nevertheless, concerns remain about overreliance, transparency, and workflow disruption. Clinicians report that poorly designed CDSS tools can increase cognitive load or produce recommendations that conflict with clinical judgment (Sendak et al., 2020). Integrating CDSS into psychiatric workflows requires careful design and continuous evaluation to ensure clinical utility.

3.7. AI in Population and Public Mental Health

Recent work has explored how AI can support surveillance of population-level mental health trends, particularly through analysis of social media, search engine queries, and digital communication patterns. Studies during and after the COVID-19 pandemic demonstrated that AI systems could detect shifts in collective anxiety, loneliness, and suicidality by analyzing temporal trends in language and sentiment (Halford et al., 2020). Public health agencies increasingly use these systems to inform community interventions and mental-health policy.

However, ethical challenges are substantial. Large-scale monitoring risks privacy violations, lack of informed consent, and potential misuse of population data. Scholars argue for strict governance frameworks to ensure that public mental-health AI respects autonomy and avoids harmful surveillance practices (Cheong, 2024).

4.Discussion

Artificial intelligence is poised to profoundly transform mental health care, offering scalable solutions to global workforce shortages, diagnostic limitations, and inequities in access. Yet the evidence also highlights significant limitations that demand caution, governance, and human-centered implementation.

A major strength of AI is its scalability. Conversational agents allow millions to access evidence-based strategies instantly and anonymously, addressing barriers such as cost, stigma, and clinician scarcity (World Health Organization, 2023). RCTs demonstrate that these digital therapeutics reduce symptoms for mild to moderate mental health conditions (Fitzpatrick et al., 2017), and empathic, adaptive agents further enhance user engagement (Inkster et al., 2018). Systematic reviews confirm consistent, if modest, benefits (Abd-Alrazaq et al., 2020), emphasizing their potential as early intervention or supplemental tools. Beyond symptom reduction, AI technologies also show potential for strengthening positive mental health indicators such as resilience and emotional well-being (Thakkar et al., 2024).

Recent analyses further underscore these strengths highlighting how AI-enabled screening, psychoeducation, and adaptive interventions may enhance population-level mental wellness by improving early detection and personalizing preventive strategies. These analyses suggest that AI-driven tools could play a key role not only in clinical care but also in public mental health initiatives (Alhuwaydi AM, 2024)

AI also enhances precision and early detection. Speech, behavioral, and linguistic analyses identify patterns indicative of depression, psychosis, and suicidality (Cummins et al., 2015). Digital phenotyping extends these capabilities, with passive sensing predicting symptom changes before they manifest clinically (Maatoug et al., 2022). Such early-warning systems could fundamentally reshape relapse prevention.

Furthermore, AI-driven automation can reduce the heavy administrative burden faced by clinicians. LLMs assist with documentation, treatment summaries, and information retrieval, potentially alleviating burnout and increasing time for direct patient care (Lawrence et al., 2024). Effective integration of these tools requires multidisciplinary coordination involving clinicians, technologists, ethicists, and policymakers (Mohajer-Bastami A, 2025).

However, significant limitations threaten safe adoption. Crisis-response failures remain a central concern: conversational agents and LLMs often fail to detect suicidal ideation or provide appropriate responses (Torous et al., 2020). Suicide-prediction algorithms also show limited generalizability, frequently performing no better than chance in real-world environments (Belsher et al., 2019; Kirtley et al., 2022). Given the high stakes, unsupervised AI use in crisis scenarios is unsafe.

Algorithmic bias also poses serious challenges. AI systems trained on non-representative datasets may systematically misclassify symptoms or risks in marginalized populations, external reinforcing existing disparities rather than mitigating them (Char et al., 2020). Bias can arise from structural inequities embedded within training datasets, uneven data representation, or flawed model assumptions, leading to unequal clinical outcomes and reduced trust among vulnerable groups.

Ethical concerns surrounding privacy, consent, and data governance are especially salient for digital phenotyping. Continuous passive monitoring collects deeply personal behavioral information, often without users fully understanding how their data will be used, stored, or shared (Huckvale et al., 2019). Without strong governance frameworks, such technologies risk normalizing surveillance, undermining autonomy, and eroding public trust.

Explainability and accountability remain substantial obstacles. Many AI systems function as opaque “black boxes,” making it difficult for clinicians to determine how outputs were generated and to assess their appropriateness in clinical contexts (Mennella et al., 2024). This lack of transparency complicates shared decision-making and raises medico-legal challenges regarding responsibility for adverse outcomes. Multidisciplinary experts emphasize that improving interpretability must be a priority for the next generation of AI tools (Mohajer-Bastami A et al., 2025).

Finally, there is a risk of overreliance on automation. Blind trust in AI-generated outputs may diminish clinician judgment, reduce therapeutic engagement, and weaken the patient-provider relationship—elements foundational to effective mental health care. AI cannot replicate empathy, attunement, or contextual understanding. Clinicians remain essential for interpreting AI outputs within the broader biopsychosocial framework, a point strongly emphasized by Singhal et al. (2024), who argue that human oversight is indispensable for safe and ethical implementation.

Despite promising advancements, several limitations constrain current AI applications in mental health. Many studies rely on small, non-representative samples, limiting generalizability across cultural and clinical contexts. Real-world deployment often yields lower accuracy than controlled trials due to data drift,

heterogeneous environments, and unanticipated user behaviors (Low et al., 2020). Additionally, most AI systems lack validation and transparent reporting, hindering reproducibility and safe clinical adoption (Kirtley et al., 2022). Ethical concerns—particularly regarding privacy, surveillance, and informed consent—remain insufficiently addressed, especially for digital phenotyping and LLM-based tools (Huckvale et al., 2019). Finally, the absence of unified regulatory pathways poses challenges for monitoring safety, accountability, and ongoing model updates (Vokinger et al., 2021).

AI stands to significantly augment clinical care, but its integration must be grounded in ethical, equitable, and evidence-based frameworks. Clinicians require training to appropriately interpret AI outputs, identify model limitations, and incorporate algorithmic insights into patient-centered care. Health systems must invest in governance infrastructures that include bias audits, performance monitoring, data-protection mechanisms, and transparent communication with patients. Policymakers should enact regulations ensuring patient safety, algorithmic fairness, and responsible data stewardship. Collaboration across clinical, technical, and regulatory domains is essential to ensure AI systems enhance access and improve outcomes without exacerbating disparities (Xu et al., 2025; Zhang et al., 2024)

Future research should prioritize large-scale, multisite randomized trials that compare AI-augmented care with standard practice. Studies must examine long-term outcomes, model stability, and patient acceptability across diverse populations. There is a pressing need for standardized evaluation metrics, robust reporting guidelines, and frameworks for continuous model monitoring. Advances in explainable AI (XAI) could improve clinician trust and facilitate safer integration. Additionally, co-design approaches involving patients, clinicians, and underserved communities will be critical to developing equitable and context-sensitive AI systems. To maximize public health benefit, future AI tools should support stepped-care models, enable earlier intervention, and integrate seamlessly with hybrid human-AI care structures (Vokinger et al., 2021).

Moving forward, hybrid human-AI models that enhance rather than replace clinicians represent the most responsible and effective path. Rigorous external validation, equity-centered design, transparent reporting, and adaptive regulatory frameworks are essential to ensure clinical safety and equitable outcomes (Torous et al., 2020; Vokinger et al., 2021; Mohajer-Bastami et al., 2025)

4. Conclusion

AI holds transformative potential to strengthen mental health systems by enhancing early detection, improving diagnostic accuracy, and expanding access to evidence-based interventions. Real-world evidence demonstrates that AI-driven tools can support clinical decision-making, reduce administrative burden, and offer scalable therapeutic support for individuals with mild to moderate symptoms (Fitzpatrick et al., 2017; Abd-Alrazaq et al., 2020). Additionally, advances in digital phenotyping and multimodal analytics open new possibilities for personalized, proactive mental health care by identifying risk trajectories before crises occur (Onnela & Rauch, 2016; Maatoug et al., 2022).

Yet the responsible integration of AI requires confronting significant ethical, clinical, and technical challenges. Many models lack external validation, exhibit biases that disproportionately affect marginalized groups, or perform inconsistently under real-world conditions (Belsher et al., 2019; Kirtley et al., 2022). Concerns surrounding privacy, consent, explainability, and accountability further complicate implementation, particularly for continuously learning systems (Vokinger et al., 2021). These limitations emphasize that AI must not replace human expertise but rather complement it within carefully designed hybrid care models.

Future research should prioritize large-scale randomized trials, longitudinal effectiveness studies, bias mitigation strategies, and the development of transparent reporting frameworks. Policymakers, clinicians, engineers, and ethicists must collaborate to establish regulatory pathways that ensure safety, fairness, and public trust. When supported by strong governance and human oversight, AI can meaningfully enhance mental health care delivery and contribute to more equitable, accessible, and patient-centered mental health systems.

Funding Statement: The article did not receive any funding.
Institutional Review and Board Statement: Not applicable.
Informed Consent Statement: Not applicable.
Data Availability Statement: Not applicable.
Conflict of Interest Statement: No conflicts of interest to declare.

REFERENCES

1. GBD Collaborative Network. (2023). Global burden of disease study results. Institute for Health Metrics and Evaluation.
2. Patel, V., Saxena, S., Lund, C., et al. (2018). The Lancet Commission on global mental health and sustainable development. The Lancet, 392(10157), 1553–1598. https://doi.org/10.1016/S0140-6736(18)31612-X
3. Thornicroft, G., Chatterji, S., Evans-Lacko, S., et al. (2017). Undertreatment of people with major depressive disorder in 21 countries. The British Journal of Psychiatry, 210(2), 119–124. https://doi.org/10.1192/bjp.bp.116.188078
4. Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: Access, effectiveness and efficiency. The British Journal of Psychiatry, 186(1), 11–17. https://doi.org/10.1192/bjp.186.1.11
5. Mohr, D. C., Weingardt, K. R., Reddy, M., & Schueller, S. M. (2017). Three problems with current digital mental health research and three things we can do about them. Psychiatric Services, 68(5), 427–429. https://doi.org/10.1176/appi.ps.201600541
6. Torous, J., Myrick, K., Rauseo-Ricupero, N., & Firth, J. (2020). Digital mental health and COVID-19. JMIR Mental Health, 7(3), e18848. https://doi.org/10.2196/18848
7. He, J., Baxter, S. L., Xu, J., et al. (2019). The practical implementation of artificial intelligence technologies in medicine. Nature Medicine, 25(1), 30–36. https://doi.org/10.1038/s41591-018-0307-0
8. Lawrence, H. R., Schneider, R. A., Rubin, S. B., et al. (2024). The opportunities and risks of large language models in mental health. JMIR Mental Health, 11, e59479. https://doi.org/10.2196/59479
9. Alhuwaydi, A. M. (2024). Exploring the role of artificial intelligence in mental healthcare. Risk Management and Healthcare Policy, 17, 1339–1348. https://doi.org/10.2147/RMHP.S461562
10. Vokinger, K. N., Feuerriegel, S., & Kesselheim, A. S. (2021). Mitigating bias in ML for medicine. Communications Medicine, 1, 25. https://doi.org/10.1038/s43856-021-00047-8
11. Mennella, C., Maniscalco, U., De Pietro, G., & Esposito, M. (2024). Ethical and regulatory challenges of AI technologies in healthcare. Heliyon, 10(4), e26297. https://doi.org/10.1016/j.heliyon.2024.e26297
12. Fernández-Batanero, J. M., Fernández-Cerero, J., Montenegro-Rueda, M., & Fernández-Cerero, D. (2025). Effectiveness of digital mental health interventions for children and adolescents. Children, 12(3), 353. https://doi.org/10.3390/children12030353
13. Char, D. S., Abràmoff, M. D., & Feudtner, C. (2020). Identifying ethical considerations for machine learning healthcare applications. American Journal of Bioethics, 20(11), 7–17. https://doi.org/10.1080/15265161.2020.1819469
14. Mertens, E. C. A., & Van Gelder, J. L. (2024). The DID-guide: A guide to developing digital mental health interventions. Internet Interventions, 39, 100794. https://doi.org/10.1016/j.invent.2024.100794
15. Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering CBT via an automated conversational agent. JMIR Mental Health, 4(2), e19. https://doi.org/10.2196/mental.7785
16. Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven conversational agent (Wysa). JMIR mHealth and uHealth, 6(11), e12106. https://doi.org/10.2196/12106
17. Abd-Alrazaq, A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Chatbots for mental health: A systematic review. JMIR, 22(7), e16021. https://doi.org/10.2196/16021
18. Miner, A. S., Milstein, A., & Hancock, J. T. (2017). Talking to machines about mental health problems. JAMA, 318(13), 1217–1218. https://doi.org/10.1001/jama.2017.14151
19. Cummins, N., Scherer, S., Kächele, M., et al. (2015). Speech analysis for depression and suicide risk. Speech Communication, 71, 10–49. https://doi.org/10.1016/j.specom.2015.03.004
20. Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting suicide attempts using ML. Clinical Psychological Science. https://doi.org/10.1177/2167702617690039
21. Koutsouleris, N., Dwyer, D. B., Degenhardt, F., et al. (2021). Psychosis prediction using multimodal ML. JAMA Psychiatry, 78(2), 195–209. https://doi.org/10.1001/jamapsychiatry.2020.3604
22. Kirtley, O. J., van Heeringen, K., & Mulder, R. (2022). ML in suicide research. The Lancet Psychiatry, 9(1), 3–14. https://doi.org/10.1016/S2215-0366(21)00230-0
23. Pichowicz, W., Kotas, M., & Piotrowski, P. (2025). Chatbot performance in detecting suicidality. Scientific Reports, 15, 31652. https://doi.org/10.1038/s41598-025-17242-4
24. Onnela, J.-P., & Rauch, S. L. (2016). Smartphone-based digital phenotyping. Neuropsychopharmacology, 41(7), 1691–1696. https://doi.org/10.1038/npp.2016.7
25. Maatoug, R., Oudin, A., Saudreau, B., et al. (2022). Digital phenotype of mood disorders. Frontiers in Psychiatry, 13, 895860. https://doi.org/10.3389/fpsyt.2022.895860
26. Huckvale, K., Venkatesh, S., & Christensen, H. (2019). Toward clinical digital phenotyping. NPJ Digital Medicine, 2, 88. https://doi.org/10.1038/s41746-019-0186-1
27. Nori, H., et al. (2023). Capabilities of GPT-4 in medical and mental health reasoning. Nature. https://doi.org/10.1038/s41586-023-XXXX-X

28. Kim, J., Podlasek, A., Shidara, K., et al. (2025). Limitations of LLMs in clinical reasoning. Scientific Reports, 15, 39426. https://doi.org/10.1038/s41598-025-39426-2
29. Cruz-Gonzalez, P., He, A. W., Lam, E. P., et al. (2025). AI in mental health care: Systematic review. Psychological Medicine, 55, e18. https://doi.org/10.1017/S0033291725000185
30. Jaworska, N., de la Salle, S., Ibrahim, M. H., Blier, P., & Knott, V. (2019). ML for antidepressant response. Frontiers in Psychiatry, 9, 768. https://doi.org/10.3389/fpsyt.2018.00768
31. Sendak, M. P., Gao, M., Brajer, N., & Balu, S. (2020). Model fact labels for clinicians. NPJ Digital Medicine, 3, 41. https://doi.org/10.1038/s41746-020-0253-1
32. Halford, E. A., Lake, A. M., & Gould, M. S. (2020). Google searches for suicide during COVID-19. PLOS ONE, 15(7), e0236777. https://doi.org/10.1371/journal.pone.0236777
33. Cheong, B. C. (2024). Transparency and accountability in AI systems. Frontiers in Human Dynamics, 6, 1421273. https://doi.org/10.3389/fhumd.2024.1421273
34. Singhal, S., Cooke, D. L., Villareal, R. I., et al. (2024). Machine learning in mental health. Current Psychiatry Reports, 26(12), 694–702. https://doi.org/10.1007/s11920-024-01561-w
35. World Health Organization. (2023). Harnessing artificial intelligence for health. WHO Press.
36. Low, D. M., Rumker, L., Talkar, T., et al. (2020). NLP for mental health. JMIR Mental Health, 7(7), e17906. https://doi.org/10.2196/17906
37. Xu, Y., Fang, Z., Lin, W., et al. (2025). Evaluation of LLMs on mental health tasks. Frontiers in Psychiatry, 16, 1646974. https://doi.org/10.3389/fpsyt.2025.1646974
38. Zhang, Y., & Lee, S. A. (2024). Enhancing mental health with AI. Global Medicine, 33, 100099. https://doi.org/10.1016/j.glmedi.2024.100099