Applied Psychology and Cognitive Science

Review
Volume 1(1), 2026, 1−13
https://doi.org/10.53623/apcs.v1i1.1336

Artificial Intelligence in Psychological Practice: A Systematic Review of Opportunities, Challenges, and Ethical Considerations

Rizky Aditya Pranawa*, Nabila Ayu Kurniasari

Universitas Muhammadiyah Palu, Jalan H. M. Rusdy Toana No. 1, Mantikulore, Palu 94118, Indonesia

*Correspondence: rizkyaditpranawa@gmail.com

SUBMITTED: 10 August 2026; REVISED: 29 September 2026; ACCEPTED: 2 October 2026

ABSTRACT: Artificial intelligence (AI) is increasingly reshaping psychological practice by extending the capabilities of clinical assessment, personalized intervention, decision support, and continuous behavioral monitoring. This systematic review critically synthesized current evidence on the clinical opportunities, implementation barriers, and ethical implications associated with the integration of AI into psychological care. Contemporary applications encompassed machine learning models for diagnostic screening and suicide risk prediction, conversational agents delivering cognitive behavioral therapy, clinician decision support systems, and digital phenotyping through smartphones and wearable sensors. Evidence from randomized controlled trials and systematic reviews suggested that AI assisted interventions could enhance accessibility, scalability, patient engagement, and outcomes related to depressive symptoms. However, evidence regarding anxiety, stress, and other psychological outcomes remained heterogeneous, limiting definitive conclusions regarding clinical effectiveness. Translation into routine psychological practice was further constrained by small and insufficiently representative datasets, limited external and prospective validation, algorithmic bias, inadequate model interpretability, uncertain clinical accountability, workforce readiness, and concerns regarding potential disruption of therapeutic relationships. Ethical and governance challenges were particularly prominent and included privacy, data security, informed consent, transparency, accountability, equitable access, and compliance with rapidly evolving regulatory frameworks. Collectively, the evidence indicated a substantial gap between technological capability and clinically validated implementation. Future research should prioritize representative and culturally diverse datasets, prospective multicenter validation, longitudinal evaluation of clinical outcomes, systematic assessment of algorithmic bias, and interdisciplinary collaboration among psychologists, AI developers, healthcare institutions, policymakers, and regulators. Rather than functioning as a substitute for psychological expertise, AI should be positioned as an augmentative clinical tool operating within clearly defined professional and ethical boundaries. Responsible integration therefore requires a human centered framework in which clinical judgment, therapeutic relationships, patient autonomy, transparent governance, and robust ethical safeguards remain central to psychological care.

KEYWORDS: Artificial intelligence; psychological practice; mental health; digital psychotherapy; ethical considerations.

1. Introduction

Artificial intelligence (AI) has increasingly emerged as a transformative force in the care of individuals with mental health conditions, offering new capabilities for phenotypic characterization, diagnostic assessment, personalized intervention, clinical decision support, and continuous behavioral monitoring that extend beyond the temporal and geographical limitations of conventional psychological services [1]. The growing interest in these technologies has coincided with a substantial unmet need for mental health care. Mental disorders account for a considerable and increasing proportion of the global burden of disease, while demand for psychological services frequently exceeds the availability of trained professionals [2]. Within this context, AI has been applied across three major domains of psychological and psychiatric practice, namely diagnosis, monitoring, and intervention, using technologies ranging from supervised machine learning classifiers and predictive algorithms to conversational agents capable of delivering structured therapeutic content [3].

The potential value of AI is particularly relevant because conventional psychological care continues to face persistent barriers related to cost, geographical accessibility, workforce shortages, treatment capacity, and stigma [4]. AI enabled technologies offer opportunities to expand access to psychological support, identify individuals at risk, personalize interventions, and extend monitoring beyond conventional clinical encounters. Reviews of AI enabled psychotherapy have indicated that conversational agents and machine learning based decision support systems may improve accessibility and engagement [5]. AI has also been incorporated into ongoing psychological interventions to support real time or near real time clinical decision making for emotional and behavioral problems, potentially extending therapeutic support beyond the consulting room [6]. These developments suggest that AI could contribute to a transition from episodic and reactive psychological care toward more continuous, personalized, and data informed models of care.

However, a critical problem remains between the rapid technological development of AI and its readiness for routine psychological practice. Demonstrating technical performance does not necessarily establish clinical effectiveness, safety, generalizability, or ethical acceptability. AI models may perform well within development datasets but demonstrate reduced reliability when applied to different populations, cultural settings, clinical environments, or real world conditions. Furthermore, AI based psychological interventions raise concerns regarding algorithmic bias, privacy, sensitive personal data, model transparency, accountability for incorrect recommendations, professional responsibility, and the potential alteration of therapeutic relationships. Consequently, the central challenge is no longer simply whether AI can be applied to psychological care, but whether existing AI applications have sufficient clinical evidence, ethical safeguards, and implementation readiness to support responsible integration into routine practice.

This problem is compounded by fragmentation in the existing evidence. Studies have examined diagnostic algorithms, suicide risk prediction, conversational agents, digital phenotyping, wearable technologies, and clinician decision support systems, but these applications have frequently been evaluated separately and using heterogeneous methodological approaches and outcome measures. At the same time, discussions of clinical

effectiveness have often developed separately from debates concerning ethics, regulation, workforce readiness, and implementation [5,7]. Consequently, the literature does not yet provide a sufficiently integrated understanding of whether demonstrated technological opportunities are matched by clinical validity, practical feasibility, and appropriate governance. This fragmentation creates an important knowledge gap for psychologists, healthcare organizations, developers, and policymakers seeking to determine how AI can be responsibly incorporated into psychological practice.

Based on this gap, the present review was guided by three review hypotheses. First, it was hypothesized that AI applications provide measurable opportunities to enhance psychological practice through improved assessment, accessibility, personalization, intervention delivery, and continuous behavioral monitoring. Second, it was hypothesized that the translation of these technological capabilities into routine psychological practice is constrained by limitations in clinical validation, data representativeness, interpretability, workforce readiness, and implementation feasibility. Third, it was hypothesized that ethical and governance factors, particularly privacy, algorithmic bias, transparency, accountability, equity, and regulatory uncertainty, represent critical conditions determining the responsible adoption of AI in psychological care. Together, these hypotheses provided an analytical framework for evaluating not only what AI technologies can accomplish, but also the extent to which existing evidence supports their safe and sustainable implementation.

Accordingly, this systematic review aimed to critically synthesize current evidence on the opportunities, implementation challenges, and ethical considerations associated with the integration of AI into psychological practice. Specifically, the review aimed to (1) identify and characterize major AI applications in psychological assessment, intervention, clinical decision support, and behavioral monitoring; (2) evaluate the reported clinical and practical opportunities associated with these applications; (3) identify methodological, technical, clinical, and workforce barriers affecting their implementation; and (4) critically examine ethical and governance issues related to privacy, data security, algorithmic bias, transparency, accountability, equity, and regulation. By integrating these dimensions within a single analytical framework, this review sought to clarify the gap between technological capability and clinically responsible implementation and to identify priorities for future research, clinical practice, and governance of AI assisted psychological care.

2. Opportunities of Artificial Intelligence in Psychological Practice

2.1. Clinical Assessment, Screening, and Risk Prediction.

AI methods, particularly supervised machine learning classifiers such as support vector machines and random forest algorithms, have been widely applied to diagnostic, screening, and risk prediction tasks in mental health. These approaches have demonstrated promising performance in detecting, classifying, and predicting the risk of various psychiatric conditions [3]. Their potential clinical value lies in the ability to process large and multidimensional datasets, identify patterns that may not be readily detectable through conventional assessment, and support earlier identification of individuals who may require further psychological evaluation or intervention. Suicide risk prediction has emerged as a particularly important application. Machine learning models trained using longitudinal clinical and survey data have demonstrated improvements over traditional risk assessment approaches, with predictive

performance increasing as the time interval before a suicide attempt becomes shorter [8]. Similar models developed and externally validated using large national survey datasets have demonstrated the feasibility of identifying individuals at elevated risk based on routinely collected psychosocial and demographic indicators [9]. These findings suggest that AI based predictive models could complement professional assessment and potentially support population level screening and public mental health surveillance. However, their clinical utility remains dependent on the quality and representativeness of the underlying data and their validation across populations and clinical settings.

2.2. AI Assisted Psychotherapy and Conversational Agents.

Conversational agents, commonly referred to as chatbots, represent one of the most developed applications of AI in psychological intervention. Randomized controlled trials of rule based and natural language processing driven agents delivering cognitive behavioral therapy (CBT) content, including Woebot, Tess, and Wysa, have reported reductions in self reported symptoms of depression and anxiety among community and clinical populations, together with generally favorable levels of engagement and acceptability [10–12]. These systems offer the potential to provide psychological support beyond conventional clinical settings and may improve accessibility for individuals who face geographical, financial, or other barriers to traditional psychological services.

AI has also been incorporated as a clinician support tool rather than functioning exclusively as an automated therapeutic agent. Platforms capable of transcribing and analyzing therapy sessions and providing feedback regarding the application of evidence based therapeutic techniques have demonstrated feasibility and acceptability in community mental health settings, with promising secondary outcomes for symptoms of depression and anxiety [13]. Such applications illustrate an alternative role for AI in psychological practice in which technology augments clinicians' capabilities rather than replacing professional involvement.

Evidence from meta analytic research has further indicated that AI based psychotherapeutic interventions can reduce depressive symptoms both immediately following intervention and during follow up periods of 6 to 12 months [14]. However, effects on anxiety and stress relative to active comparators have been less consistent, indicating that the effectiveness of AI assisted psychotherapy may vary according to the psychological outcome, intervention design, and comparison condition. Conceptual research has also proposed that AI based psychotherapy should be evaluated through established common psychotherapeutic factors, including empathy, therapeutic alliance, and expectancy, which contribute to treatment effectiveness across different therapeutic modalities [15]. This perspective is particularly important because technological sophistication alone may not ensure therapeutic effectiveness.

The emergence of large language models (LLMs) has further expanded the capabilities of conversational systems. Fine tuned LLM based chatbots combined with structured prompting have demonstrated improvements in conversational quality and clinical relevance compared with earlier or standalone systems, based on patient and clinician evaluations [16]. Nevertheless, increased conversational capability also introduces important questions concerning clinical accuracy, inappropriate responses, risk identification, and harm prevention. Therefore, improvements in conversational performance should be considered alongside evidence of clinical safety and effectiveness.

The characteristics and reported outcomes of selected conversational agents and clinician support platforms are summarized in Table 1. The evidence illustrates the diversity of AI applications, ranging from structured CBT chatbots to clinician facing platforms and LLM based conversational systems. Across these applications, the reported benefits primarily included reductions in psychological symptoms, increased engagement, improved conversational quality, and support for clinical practice. However, differences in study design, sample characteristics, technological architecture, and outcome measures indicate that direct comparisons among platforms should be interpreted cautiously.

Table 1. Selected AI Tools for Psychological Intervention.

Agent / Platform	Underlying Technique	Study Design and Sample	Key Reported Outcome	Ref.
Woebot	Rule based CBT chatbot with NLP	Unblinded RCT; N = 70 college students	Significant reduction in depressive symptoms (PHQ 9) compared with information only control	[10]
Tess	Psychological AI conversational agent	RCT; university students with depression or anxiety symptoms	Reduced self reported depression and anxiety symptoms	[11]
Wysa	Empathy driven conversational AI agent	Mixed methods, real world data evaluation	High engagement and improvements in self reported mood ratings	[12]
Eleos Health AI platform	Clinician facing session transcription and feedback	RCT; community mental health center patients	Feasible and acceptable, with reductions in PHQ 9 and GAD 7 scores	[13]
Fine tuned LLM chatbot (GPT 3.5 + DialoGPT)	Large language model with prompt engineering	Experimental comparative study	Improved conversational quality and clinical relevance compared with standalone LLM	[16]

2.3. Digital Phenotyping and Continuous Behavioral Monitoring.

Digital phenotyping, which involves the use of smartphone sensors, wearable devices, and passive digital data to characterize individuals' behavioral and physiological states, has emerged as another important opportunity for AI enabled psychological practice. Unlike conventional psychological assessment, which typically captures information at specific clinical encounters, digital phenotyping can generate longitudinal information regarding behavioral patterns in everyday environments. This capability offers the potential for more continuous and ecologically relevant monitoring of changes associated with psychological well being.

Scoping reviews of smartphone based sensing have documented applications across a broad range of mental health and behavioral domains [17]. Frequently employed data sources have included global positioning system information, accelerometry, screen usage patterns, and other indicators of digital behavior that may provide information related to mood, sleep, mobility, and social activity. When combined with AI based analytical methods, these longitudinal data streams may facilitate the identification of behavioral changes that precede clinically meaningful deterioration or improvement.

Applications among adolescent populations have further demonstrated the potential of digital phenotyping to capture moment to moment behavioral patterns associated with well being and social media use [18]. However, these applications have also highlighted methodological challenges in transforming highly individualized sensor data into reliable and clinically actionable indicators. Variability in device use, contextual influences, missing data, and individual behavioral patterns can complicate interpretation. Therefore, the value of digital phenotyping depends not merely on the quantity of data collected but also on the ability to

establish clinically meaningful relationships between digital signals and psychological outcomes.

Collectively, digital phenotyping and continuous behavioral monitoring complemented AI based diagnostic and intervention technologies by providing longitudinal behavioral information that could support more adaptive and personalized psychological care [3]. Rather than relying exclusively on retrospective self reports or occasional clinical observations, these approaches offered opportunities to identify behavioral changes over time and potentially inform the timing and personalization of psychological interventions.

Beyond individual technologies, evidence from systematic reviews and meta analyses provided a broader perspective on the opportunities for AI across psychological and psychiatric practice. As summarized in Table 2, existing reviews covered diagnostic prediction, psychological monitoring, conversational interventions, psychotherapy, and real time AI supported interventions. Across these domains, the literature generally indicated potential benefits for early detection, symptom reduction, patient engagement, risk prediction, and augmentation of psychological interventions. At the same time, variation in the number and design of included studies demonstrated that the maturity of the evidence differed considerably across AI applications. Thus, the available literature supported the potential of AI across multiple areas of psychological practice but also indicated the need for cautious interpretation of the strength and generalizability of current evidence.

Table 2. Evidence from Reviews of AI in Psychological Practice.

Scope / Domain	Studies Included	Principal Finding	Ref.
Diagnosis, monitoring, and intervention	85 studies	SVM and random forest were frequently used for diagnosis; AI chatbots were commonly used for intervention; promising risk and treatment response prediction was reported	[3]
General applications of AI in mental health	15 studies	AI supported early detection and patient engagement through chatbots and predictive modeling	[2]
AI chatbots for depression and anxiety	10 studies, mostly RCTs	Conversational AI agents were associated with improvements in psychological symptoms	[4]
AI for psychotherapy	Narrative synthesis	Cautious, patient centered integration was recommended, with emphasis on privacy and trust	[5]
AI based psychotherapeutic interventions	30 RCTs, 31 publications, N = 6,100	Significant reductions in depressive symptoms were reported, while effects on anxiety and stress were comparable with active comparators	[14]
Real time or near real time AI enhanced interventions	Systematic review	AI showed potential to augment ongoing psychological interventions for emotional problems	[6]

3. Challenges in Implementing Artificial Intelligence in Psychological Practice.

3.1. Technical and Methodological Limitations.

Despite encouraging developments, the translation of AI applications into routine psychological practice remained constrained by substantial technical and methodological limitations. Systematic reviews consistently identified small and single site samples, reliance on internal rather than external validation, heterogeneous outcome measures, and considerable variation in methodological quality across studies [2, 3]. These limitations reduced confidence in the generalizability of reported findings and raised uncertainty regarding whether models developed under controlled research conditions would maintain their performance across different populations, institutions, and real world clinical environments.

These limitations were particularly important for high risk applications such as suicide prediction. Machine learning models demonstrated improved predictive accuracy as the time

interval before a suicide attempt decreased, indicating their potential value for identifying individuals at elevated short term risk [8]. However, comprehensive evaluation across diverse populations and healthcare settings remained limited, and relatively few models had undergone prospective validation within routine clinical workflows [8,9]. Consequently, high predictive performance within a development dataset did not necessarily demonstrate clinical effectiveness or reliability when implemented in different populations or healthcare systems.

Broader reviews of AI and machine learning in healthcare similarly identified inadequate prospective clinical validation as a major barrier to safe and reliable implementation [19]. Another methodological concern involved the transparency and interpretability of complex predictive models. Clinicians need to understand, at least at an operational level, how an AI system generates recommendations and which factors contribute to its predictions, particularly when outputs influence high consequence clinical decisions. Limited interpretability may therefore reduce professional trust, complicate clinical accountability, and restrict the effective integration of AI generated information into psychological assessment and treatment planning [19].

The practical characteristics of selected machine learning and passive sensing approaches are summarized in Table 3. These studies demonstrated the diversity of data sources and analytical strategies used for suicide risk prediction and digital phenotyping, ranging from electronic health records and national survey datasets to smartphone sensor data. Although the reported findings demonstrated promising predictive and monitoring capabilities, differences in datasets, populations, analytical methods, and outcome measures highlighted the continuing difficulty of establishing standardized evidence regarding clinical performance and generalizability.

Table 3. AI Approaches for Risk Prediction and Digital Phenotyping.

Approach	Data Source	Algorithm(s) / Method	Performance / Outcome	Ref.
Longitudinal suicide attempt risk prediction	Electronic health records and administrative or clinical data	Machine learning classifiers and ensemble methods	Predictive accuracy improved as the time to an attempt narrowed from 720 to 7 days	[8]
National survey suicide risk model	National survey data from Japan	Machine learning predictive model	Identified important psychosocial and demographic risk indicators for population level screening	[9]
Smartphone based digital phenotyping across conditions	Smartphone sensors, including GPS, usage logs, and accelerometry	Passive sensing and behavioral feature extraction	Identified consistent behavioral associations across mental and physical health conditions	[17]
Adolescent mobile sensing pilot	Smartphone passive monitoring application (AWARE)	Idiographic mobile sensing	Demonstrated feasibility for capturing digital behaviors associated with social media use and well being	[18]

3.2. Workforce, Implementation, and the Therapeutic Relationship.

The introduction of AI into psychological practice also created important implementation and relational challenges for the clinical workforce. One central concern involved the allocation of professional responsibility when AI systems operated independently or contributed directly to clinical decisions. The deployment of AI based therapeutic tools without direct involvement of a human therapist created a novel therapeutic context in which the duty of care toward users might not be clearly defined [21]. This ambiguity complicated questions of responsibility when an AI system provided inappropriate guidance, failed to identify psychological deterioration, or generated responses that adversely affected users. Clinical implementation also required psychologists and other mental health professionals to develop competencies beyond

conventional therapeutic skills. Effective use of AI required clinicians to interpret algorithmic outputs, understand their limitations, recognize potential errors, and determine when automated recommendations should be questioned or disregarded. Commentary within psychiatric and mental health nursing similarly emphasized that AI technologies introduced potential risks to patient safety and quality of care that required active monitoring and mitigation rather than reliance on technological performance alone [20]. These challenges were closely connected to concerns regarding the therapeutic relationship. Psychological care depends substantially on empathy, trust, contextual understanding, professional judgment, and interpersonal communication. Although AI systems could support specific therapeutic tasks and increase access to psychological resources, excessive reliance on automated systems could alter the relational characteristics of care. Consequently, previous research emphasized that AI integration into psychotherapy should proceed cautiously and preserve human judgment and the therapeutic relationship as central components of psychological practice [5]. AI therefore appeared most appropriate as a complementary clinical resource rather than an autonomous substitute for professional psychological care.

3.3. Algorithmic Bias and Equity Gaps.

A further challenge concerned the possibility that AI systems could reproduce or amplify existing inequalities in mental health care. Algorithmic performance depends strongly on the quality, diversity, and representativeness of the data used for model development. When particular demographic, socioeconomic, cultural, or clinical populations are insufficiently represented in training datasets, predictive models may perform less reliably for these groups. Such disparities are particularly consequential in psychological practice because inaccurate predictions may influence diagnosis, risk classification, access to intervention, or treatment decisions. Reviews have therefore called for systematic identification and mitigation of bias throughout the AI lifecycle, extending from data collection and model development to validation, implementation, and post deployment monitoring [22]. Failure to address these issues could reinforce existing disparities in mental health diagnosis and access to appropriate treatment, particularly among populations that have historically been underrepresented in clinical datasets. Algorithmic bias was also closely connected to the methodological limitations discussed previously. Nonrepresentative training and validation datasets simultaneously reduced model generalizability and increased the possibility of inequitable performance [19,22]. Therefore, improving equity in AI enabled psychological practice required more than technical adjustments to algorithms. It also required representative datasets, subgroup specific performance evaluation, transparent reporting, diverse validation populations, and continued monitoring after clinical implementation.

4. Ethical Considerations

4.1. Privacy, Data Security, and Confidentiality.

The application of AI in psychological practice depended on the collection, processing, storage, and interpretation of highly sensitive psychological, behavioral, and personal information. This created substantial concerns regarding privacy, confidentiality, and data security. Ethical analyses of AI in mental health emphasized the need for clear and broadly applicable guidelines to protect the privacy, autonomy, and dignity of individuals with mental health conditions while

allowing potential benefits from AI enabled care to be realized [23]. These concerns became particularly significant when AI applications continuously collected behavioral information through conversational platforms, smartphones, wearable devices, or digital phenotyping systems. Such data could reveal sensitive information about individuals' emotional states, daily activities, social relationships, mobility patterns, and psychological vulnerabilities. Consequently, inappropriate access, secondary use, data breaches, or unauthorized transfer of these data could produce consequences extending beyond conventional concerns regarding digital privacy. Commercial AI based mental health applications presented an additional challenge because organizations processing sensitive therapeutic information might operate primarily under general data protection requirements rather than the professional confidentiality obligations applicable to licensed psychological practitioners [21]. This situation created a potential structural gap between users' expectations of therapeutic confidentiality and the legal responsibilities of technology providers. Broader reviews of AI ethics in healthcare similarly identified privacy and data security as persistent concerns across clinical applications, with particular significance in psychological practice because of the sensitivity of mental health information [19].

4.2. Accountability, Transparency, and the Therapeutic Relationship.

A second major ethical issue concerned accountability for AI generated recommendations and their consequences. When an AI system influenced psychological assessment, intervention, or risk evaluation, responsibility could become distributed among clinicians, healthcare institutions, technology developers, and platform providers. This fragmentation created uncertainty regarding who should be accountable when automated recommendations were inaccurate, inappropriate, or harmful. Analyses based on an ethics of care perspective argued that conventional responsible AI principles alone might be insufficient because they could overlook the relational dimensions of psychological treatment. Such analyses proposed that developers of AI based therapeutic systems should assume responsibilities analogous to a duty of care, including safeguards against inappropriate emotional influence or manipulation [21]. This perspective was particularly relevant for conversational systems that simulated empathic interaction and might therefore be perceived by users as possessing therapeutic authority. Transparency and interpretability were equally important for maintaining meaningful human oversight. Clinicians and patients needed sufficient information to understand the role, limitations, and uncertainty associated with AI generated outputs. Transparent systems could support informed clinical judgment, whereas opaque models could make it difficult to determine why a particular prediction or recommendation had been generated. Previous literature therefore identified interpretability, clinician oversight, and clear communication regarding system capabilities and limitations as important conditions for maintaining trust in AI assisted psychological care [19, 23]. These considerations reinforced the principle that AI should support rather than displace professional judgment and human centered therapeutic relationships [20].

4.3. Regulatory and Governance Frameworks.

Responsible integration of AI into psychological practice also required regulatory and governance frameworks capable of responding to rapidly evolving technologies. The speed of technological development created challenges for conventional regulatory systems because

new AI capabilities could emerge more rapidly than clinical standards, professional guidance, and legal requirements could be established. Existing regulatory initiatives, including comprehensive legislative frameworks such as the European Union Artificial Intelligence Act, represented efforts to translate general ethical principles into binding requirements for AI systems according to their levels of risk [21]. However, the application of these frameworks to mental health specific technologies remained an evolving area. Psychological AI systems introduced distinctive concerns because they could interact directly with individuals experiencing psychological distress, process highly sensitive information, and potentially influence decisions related to treatment or personal well being. Broader scientific commentary emphasized that safe and sustainable clinical translation required rigorous validation, active stakeholder participation, clearly defined safeguards, and appropriate education for clinicians [1]. Governance should therefore not be viewed solely as a restriction on technological innovation but as a necessary condition for establishing safety, professional accountability, public trust, and clinical legitimacy. Mental health specific ethical guidelines addressing privacy, impartiality, transparency, responsibility, and professional oversight have consequently been proposed as important complements to broader AI regulation [23].

The principal ethical concerns and corresponding safeguards identified across the reviewed literature are summarized in Table 4. The evidence demonstrated that ethical challenges extended beyond data protection to encompass accountability, transparency, algorithmic equity, therapeutic relationships, and regulatory oversight. Importantly, these domains were interconnected. For example, inadequate transparency could weaken accountability, while nonrepresentative data could simultaneously create technical limitations and ethical concerns related to equity. Responsible implementation therefore required an integrated governance approach rather than isolated responses to individual ethical risks.

Table 4. Ethical Concerns and Safeguards for AI in Psychological Practice.

Ethical Domain	Core Concern	Proposed Safeguard	Ref.
Privacy and data security	Sensitive therapeutic data handled under general rather than medical confidentiality requirements	Mental health specific privacy standards and restrictions on third party data transfer	[23]
Accountability and therapeutic relationship	Absence of clearly defined duty of care when an AI therapeutic agent operates without a human clinician	Ethics of care obligations for developers and safeguards against emotional manipulation	[21]
Transparency and interpretability	Complex or opaque models limit clinician and patient understanding of AI outputs	Interpretable model design, clinician education, and professional oversight	[19]
Algorithmic bias and equity	Nonrepresentative training and validation data may reinforce disparities in psychological care	Systematic bias assessment and mitigation throughout the AI lifecycle	[22]
Regulation and governance	Fragmented regulatory frameworks may be outpaced by rapidly developing AI technologies	Risk based legislation combined with mental health specific ethical and professional guidelines	[21, 23]

The interconnected relationship among the major findings of this review is illustrated in Figure 1. The framework organizes AI in psychological practice into three central domains: opportunities, implementation challenges, and ethical considerations. The opportunities encompass clinical assessment and risk prediction, AI assisted psychotherapy, and continuous behavioral monitoring, whereas implementation challenges include methodological limitations, workforce and therapeutic relationship concerns, and algorithmic bias. Ethical considerations encompass privacy and data security, accountability and transparency, and regulatory governance. Together, these domains demonstrate that the clinical potential of AI

cannot be evaluated independently of the methodological, professional, and ethical conditions under which these technologies are developed and implemented.

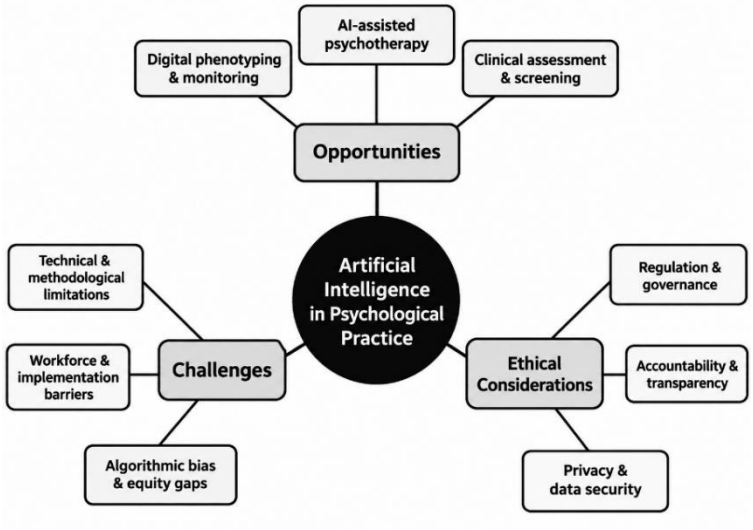


Figure 1. Opportunities, Challenges, And Ethical Considerations of AI In Psychological Practice.

5. Conclusion

Artificial intelligence is increasingly reshaping psychological practice by extending the capabilities of clinical assessment, screening, risk prediction, psychotherapy, and continuous behavioral monitoring. This review demonstrated that machine learning models can support the identification of psychiatric conditions and suicide risk, while conversational agents and AI assisted psychotherapy can improve accessibility, engagement, and depressive symptoms. Digital phenotyping through smartphones and wearable sensors further provides opportunities for continuous monitoring of behavioral and psychological changes outside conventional clinical settings. Nevertheless, these opportunities remain accompanied by substantial limitations. Small and nonrepresentative datasets, insufficient external and prospective validation, inconsistent outcome measures, limited model interpretability, and uncertain generalizability continue to restrict translation into routine psychological practice. The evidence for anxiety and stress outcomes also remains less consistent than that for depressive symptoms. In addition, workforce readiness, unclear professional responsibility, and the potential effects of automation on therapeutic relationships present important implementation challenges. Algorithmic bias may further reproduce existing inequalities when training and validation datasets inadequately represent diverse populations. These technical and clinical limitations are closely connected to ethical concerns surrounding privacy, data security, transparency, accountability, equity, informed oversight, and rapidly evolving regulatory requirements. Future development should therefore prioritize representative and culturally diverse datasets, prospective multicenter validation, long term evaluation, systematic bias assessment, transparent model development, and stronger collaboration among psychologists, healthcare institutions, AI developers, policymakers, and regulators. Overall, the evidence suggests that the greatest value of AI lies not in replacing psychologists but in augmenting professional capabilities and extending the reach and continuity of psychological care. Responsible implementation should consequently remain human centered, with professional judgment, patient autonomy, therapeutic relationships, clinical safety, equitable access, and robust governance forming the foundation of AI integration into psychological practice.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Author Contributions

Both authors contributed equally to the conceptualization, literature search, analysis, and writing of this manuscript. Both authors read and approved the final version of the manuscript.

Data Availability Statement

No new data were generated or analyzed in this review. All data supporting the findings of this study are available within the cited published articles referenced in this manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] Opel, N.; Breakspear, M. (2026). Transforming mental health research and care through artificial intelligence. *Science*, *391*, 249–258. https://doi.org/10.1126/science.adz9193.

[2] Dehbozorgi, R.; Zangeneh, S.; Khooshab, E.; Nia, D.H.; Hanif, H.R.; Samian, P.; Yousefi, M.; Hashemi, F.H.; Vakili, M.; Jamalimoghadam, N.; Lohrasebi, F. (2025). The application of artificial intelligence in the field of mental health: A systematic review. *BMC Psychiatry*, *25*, 132. https://doi.org/10.1186/s12888-025-06483-2.

[3] Cruz-Gonzalez, P.; He, A.W.-J.; Lam, E.P.; Ng, I.M.C.; Li, M.W.; Hou, R.; Chan, J.N.-M.; Sahni, Y.; Vinas Guasch, N.; Miller, T.; Lau, B.W.-M.; Sánchez Vidaña, D.I. (2025). Artificial intelligence in mental health care: A systematic review of diagnosis, monitoring, and intervention applications. *Psychological Medicine*, *55*, e18. https://doi.org/10.1017/S0033291724003295.

[4] Joshi, A.C.; Ghogare, A.S.; Madavi, P.B. (2025). Systematic review of artificial intelligence enabled psychological interventions for depression and anxiety: A comprehensive analysis. *Industrial Psychiatry Journal*, *34*, 158–166. https://doi.org/10.4103/ipj.ipj_363_24.

[5] Beg, M.J.; Verma, M.; Chanthar, V.K.M.M.; Verma, M.K. (2025). Artificial intelligence for psychotherapy: A review of the current state and future directions. *Indian Journal of Psychological Medicine*, *47*, 314–325. https://doi.org/10.1177/02537176241260819.

[6] Gual-Montolio, P.; Jaén, I.; Martínez-Borba, V.; Castilla, D.; Suso-Ribera, C. (2022). Using artificial intelligence to enhance ongoing psychological interventions for emotional problems in real- or close to real-time: A systematic review. *International Journal of Environmental Research and Public Health*, *19*, 7737. https://doi.org/10.3390/ijerph19137737.

[7] Simões, J.P.; ten Klooster, P.; Neff, P.K.; Niemann, U.; Kraiss, J. (2024). Editorial: Artificial intelligence and mental health care. *Frontiers in Public Health*, *12*, 1461446. https://doi.org/10.3389/fpubh.2024.1461446.

[8] Walsh, C.G.; Ribeiro, J.D.; Franklin, J.C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, *5*, 457–469. https://doi.org/10.1177/2167702617691560.

[9] Chou, P.-H.; Wang, S.-C.; Wu, C.-S.; Horikoshi, M.; Ito, M. (2022). A machine-learning model to predict suicide risk in Japan based on national survey data. *Frontiers in Psychiatry*, *13*, 918667. https://doi.org/10.3389/fpsyt.2022.918667.

[10] Fitzpatrick, K.K.; Darcy, A.; Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, *4*, e19. https://doi.org/10.2196/mental.7785.

[11] Fulmer, R.; Joerin, A.; Gentile, B.; Lakerink, L.; Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. *JMIR Mental Health*, *5*, e64. https://doi.org/10.2196/mental.9782.

[12] Inkster, B.; Sarda, S.; Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, *6*, e12106. https://doi.org/10.2196/12106.

[13] Sadeh-Sharvit, S.; Del Camp, T.; Horton, S.E.; Hefner, J.D.; Berry, J.M.; Grossman, E.; Hollon, S.D. (2023). Effects of an artificial intelligence platform for behavioral interventions on depression and anxiety symptoms: Randomized clinical trial. *JMIR Mental Health*, *10*, e46781. https://doi.org/10.2196/46781.

[14] Lau, Y.; Ang, W.H.D.; Ang, W.W.; Pang, P.C.I.; Wong, S.H.; Chan, K.S. (2025). Artificial intelligence–based psychotherapeutic intervention on psychological outcomes: A meta-analysis and meta-regression. *Depression and Anxiety*, *2025*, 8930012. https://doi.org/10.1155/da/8930012.

[15] Giotakos, O. (2025). Artificial intelligence-based psychotherapy: Focusing on common psychotherapeutic factors. *Frontiers in Psychiatry*, *16*, 1710715. https://doi.org/10.3389/fpsyt.2025.1710715.

[16] Yu, H.Q.; McGuinness, S. (2024). An experimental study of integrating fine-tuned large language models and prompts for enhancing mental health support chatbot system. *Journal of Medical Artificial Intelligence*, *7*, 16. https://doi.org/10.21037/jmai-23-136.

[17] Dumas, A.; Hokayem, J.; Goodman, G.; Venkatasubramanian, K.; Chai, P. (2026). Smartphone-based digital phenotyping across health conditions: Scoping review. *Journal of Medical Internet Research*, *28*, e84146. https://doi.org/10.2196/84146.

[18] Hamilton, J.L.; Dreier, M.J.; Durica, K.C.; Low, C.A. (2024). Improving the science of adolescent social media and mental health: Challenges and opportunities of smartphone-based mobile sensing and digital phenotyping. *Journal of Technology in Behavioral Science*. https://doi.org/10.1007/s41347-024-00443-5.

[19] Tilala, M.H.; Chenchala, P.K.; Choppadandi, A.; Kaur, J.; Naguri, S.; Saoji, R.; Devaguptapu, B. (2024). Ethical considerations in the use of artificial intelligence and machine learning in health care: A comprehensive review. *Cureus*, *16*, e62443. https://doi.org/10.7759/cureus.62443.

[20] Dell, C. (2025). The risks of artificial intelligence in mental health care. *Journal of Psychiatric and Mental Health Nursing*, *32*, 382–383. https://doi.org/10.1111/jpm.13119.

[21] Tavory, T. (2024). Regulating AI in mental health: Ethics of care perspective. *JMIR Mental Health*, *11*, e58493. https://doi.org/10.2196/58493.

[22] Timmons, A.C.; Duong, J.B.; Simo Fiallo, N.; et al. (2023). A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. *Perspectives on Psychological Science*, *18*, 1062–1096. https://doi.org/10.1177/17456916221134490.

[23] Warrier, U.; Warrier, A.; Khandelwal, K. (2023). Ethical considerations in the use of artificial intelligence in mental health. *Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, *59*, 139. https://doi.org/10.1186/s41983-023-00735-2.



©2026 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).