Journal of Family and Society Research 4(2), December 2025
ISSN – 1668 (online)   1676 (print)

Artificial Intelligence in Child and Adolescent Mental Health: Prevention, Diagnosis, and Treatment in Hybrid Human–AI Care Models

Nnubia U.I[1] & Nwauzoije E.J[2]
[1]Department of Educational Research, Werklund School of Education,
University of Calgary, Alberta, Canada
[2]Department of Human Development and Family Science, College of Education,
University of Nevada, Reno, U.S.A
Corresponding email: uju.nnubia@ucalgary.ca

Submitted - November 29, 2025
Final revision – December 12, 2025
Accepted - December 12, 2025

Abstract

Mental health disorders among children and adolescents have become increasingly common and burdensome, with conditions such as anxiety, depression, suicidality, and trauma-related disorders contributing significantly to disability and death. While timely identification and intervention are vital, progress is often limited by the scarcity of trained providers, ongoing stigma, and dependence on subjective evaluation methods. Against this backdrop, artificial intelligence (AI) is being explored to improve mental healthcare through enhanced early detection, monitoring, individualized interventions, and clinical decision support. This narrative review synthesizes research and systematic reviews from 2015 to 2025, sourced from Google Scholar, Web of Science, PubMed Central, PsycINFO, Science Direct, and EBSCO. Articles included focused on AI applications in children and adolescents' mental health, highlighting advances in machine learning, natural language processing, multimodal data integration, and digital cognitive-behavioral therapy. Evidence suggests that AI can analyze behavioral, physiological, and linguistic data to predict mental health risks, detect emerging symptoms, and deliver personalized interventions within a hybrid human–AI care model, where AI complements clinician expertise to improve access, engagement, and treatment outcomes. However, challenges persist, including algorithmic bias, limited model interpretability, data quality, privacy concerns, and integration into clinical workflows. Ethical and practical governance are essential to ensure that AI supports, rather than replaces, human-centered care. Future priorities include expanding research on underrepresented populations and conditions, developing explainable and equitable models, validating tools in real-world settings, and building large, FAIR-compliant datasets. Responsible, human-centered integration of AI has the potential to improve early intervention, personalize treatment, and enhance equitable access to mental healthcare for young people globally.

Keywords: artificial intelligence, mental health, children, adolescents, digital phenotyping.

Introduction

Mental health disorders among children and adolescents have risen in prevalence and burden, with anxiety, depression, suicidal behaviors, and trauma-related conditions among the leading contributors to disability and mortality in youth populations (WHO, 2025). Early identification and intervention are vital, yet challenged by shortages of trained providers, stigma, and reliance on subjective measures (McGinty, 2023). Artificial intelligence (AI) offers new ways to improve mental health care.

In the past ten years, the research in this field has grown from small early studies to a broader body of research and includes AI technologies such as natural language processing (NLP), machine learning (ML), deep learning (DL), language models (LLMs), and multimodal integration (Poudel et al., 2025). Machine learning identifies patterns from structured inputs such as questionnaire scores to support early detection, while deep learning uses neural networks to analyze complex, unstructured data such as behavioral patterns, neuroimaging, and biometric information, to predict treatment outcomes and detect early mental health issues; natural language processing (NLP) analyzes speech, text, and social media content to identify markers of disorders like depression and anxiety, large language models (LLMs) are a form of deep learning designed specifically for understanding and generating natural language, enabling conversational tools such as mental health chatbots, and multimodal data integration of audio, facial expressions, physiological signals, and textual inputs provides a comprehensive understanding of an individual's mental state, especially in remote assessments (Ali et al., 2025; Poudel et al., 2025).

As these technologies have advanced, scholars have increasingly explored the potential of AI to enhance the diagnosis, prediction, monitoring, and treatment of mental illness (Cruz-Gonzalez et al., 2025). Generative artificial intelligence (GenAI) is rapidly changing how people obtain health information, with roughly 17% of adults in the United States currently accessing it monthly, rising to 25% among younger adults (Blease & Rodman, 2025). In clinical settings, approximately 40% of mental health practitioners use tools such as ChatGPT to make decisions and communicate with patients (Jones et al., 2025). In Canada, approximately 10% of people actively seek AI support for mental health (Canadian Mental Health Association, 2025). In Nigeria, its use in mental health, though limited, shows promise in early screening and accurate detection of conditions like depression, attention-deficit hyperactivity disorder (ADHD), and psychotic disorders using electroencephalogram (EEG) and speech-analysis techniques (Abiodun et al., 2025).

Shatte et al. (2019) identified four main domains of mental health applications of AI: detection and diagnosis; prognosis, treatment, and support; public health applications; and research and clinical administration. Use of AI for detection and diagnosis is focused on identifying or diagnosing mental health conditions in individuals; prognosis, treatment, and support, is useful for prediction of illness trajectories as well as treatment and support interventions; public health applications deals with the use large-scale epidemiological or public datasets such as social media to monitor mental health trends and estimate prevalence; and research and clinical administration is aimed to improve administrative processes within mental health research, clinical practice, and health-care organizations (Shatte et al., 2019). This implies that AI has the potential to personalize therapeutic information, simulate supportive interaction, and anticipate mental health disorders using linguistic and sentiment analysis (Jones et al., 2025).

Artificial Intelligence is reshaping mental health care by providing advanced tools for diagnosis, monitoring, and treatment. These innovations tend to provide insights and solutions that traditional methods often cannot, aiming to improve accessibility and effectiveness (Cruz-Gonzalez et al., 2025; Ni & Jia, 2025). A growing body of research suggests that AI may enable earlier detection of mental-health conditions (Ali et al., 2025), support scalable intervention delivery (Fitzpatrick et al., 2017), and expand access to underserved communities (Grenon et al., 2025). The rapid integration of AI into mental health research and clinical practice has generated enthusiasm alongside concerns about validity, bias, safety, transparency, and ethics. Although advances in computational psychiatry and digital mental health offer promise, they also pose challenges related to accuracy and risk management in sensitive settings (Arvai et al., 2025; Wang et al., 2025).

Methods

This paper adopts the narrative review approach and focuses on recent systematic, scoping, and narrative reviews, meta-analyses, conceptual papers, and primary studies (randomized controlled trials and pilot studies) published between 2015 and 2025. The articles were obtained from Google Scholar, Web of Science, PubMed Central, PsycINFO, Science Direct, and EBSCO. The search included different combinations of the following key terms: (artificial intelligence, AI, digital tools) AND (mental health, psychiatric disorders) AND (children, adolescents, youth, young people) AND (human-AI hybrid, AI-Clinician, human and AI). Studies or articles that focused only on the use of AI in adults' mental health care were excluded from the review. The review prioritized papers that emphasize the use of AI in the mental health care of children and adolescents (5 -18 years of age). However, articles that had both adolescents and young adults were also included. The synthesis was organized and presented under the following subheadings:

- The use of AI for prevention, monitoring, and diagnosis.
- AI for intervention and treatment.
- The human-AI hybrid approach.
- Ethical implications of AI use in mental health care and the way forward.

AI for Prevention, Monitoring, and Diagnosis of Mental Health Problems

Mental health prevention entails primary, secondary, and tertiary prevention. Primary prevention focuses on reducing incidence in the general population, secondary prevention is concerned with early detection and intervention to stop progression, while tertiary prevention involves helping people with sub-threshold symptoms before disorders develop (Stephan et al., 2025). AI-driven approaches for mental health prevention rely on several complementary methods. AI supports diagnosis through early detection, risk prediction, and analysis of complex data. Studies show these tools can utilize digital and behavioral data to identify and predict conditions such as depression, anxiety, post-traumatic stress disorder (PTSD), and bipolar disorder with good accuracy, but results depend on how the studies are designed and tested (Cruz-Gonzalez et al., 2025; Ni & Jia, 2025).

One major area is digital phenotyping and passive sensing, which refers to the continuous, real-time measurement of individual behaviors and experiences in everyday contexts using data collected from personal digital devices (Onnela & Rauch, 2016). In this process, AI models analyze data from smartphones and wearable devices, such as global positioning system (GPS) movement, sleep patterns, keystroke dynamics, and voice or text features, to detect early behavioral shifts linked to mental health risk. These technologies can identify subtle changes in mood, sleep disruption, or social withdrawal long before clinical symptoms become obvious. Research consistently shows promising classification performance and temporal prediction, suggesting that passive sensing can serve as an early-warning system for emerging mental health problems (Huang et al., 2025; Onnela & Rauch, 2016).

AI-enabled wearables provide continuous symptom monitoring and real-time feedback, improving patient outcomes and therapy effectiveness (American Psychological Association, 2024). These wearables can use body signals such as voice-analysis algorithms to assess tone, pitch, and rhythm in speech, identifying subtle changes associated with anxiety or depression (Wang et al., 2025) Another method involves facial expression analysis, where AI systems use computer vision technologies (e.g., Affectiva) to detect subtle micro-expressions and short-lived facial cues that often go unnoticed by humans, providing indicators of psychological strain or emerging mental health concerns (Ni & Jia, 2025).

Natural Language Processing (NLP) can also be used to examine language and text from social media posts, messages, or chat logs to detect emotional patterns and shifts that may signal growing distress (Phiri et al., 2025). This AI-based tool uses different kinds of data to identify early signs of mental health problems using emotion analysis. Kamdan et al. (2025) developed an NLP-based chatbot for the early detection of mental health issues and emotional states in teenagers in Indonesia. Utilizing datasets from social media sources, the system employs preprocessing techniques, sentiment analysis, and machine learning models like Long Short-Term Memory (LSTM) to identify signs of mental distress. The study emphasizes the importance of early intervention, data privacy, and the integration of AI technologies to improve adolescent mental health support through accessible and empathetic digital tools. The study by McNeilly et al. (2023) also utilized this AI feature to analyze large volumes of adolescents' digital social communication. Specifically, the researchers applied a novel computational tool, the Effortless Assessment Research System (EARS), to automatically classify and quantify linguistic features such as the use of first-person pronouns, emotion words, and temporal focus words in over 22,000 messages from social media, email, and texting. By examining how these linguistic markers correlated with daily mood reports and depression symptoms, the AI-driven analysis allowed for passive, real-time identification of language patterns associated with internalizing symptoms, offering a potential avenue for early detection of mental health risks in adolescents (McNeilly et al., 2023).

Another key method is predictive modelling using clinical and non-clinical data. In this approach, machine learning algorithms are applied to individuals' data to identify individuals at elevated risk for issues such as self-harm, relapse, or first-onset psychiatric disorder. Systematic reviews show that AI is being used to analyze electronic health records using machine learning to identify risk patterns in medical history and clinical notes, enabling early detection and timely intervention before mental health symptoms worsen (Cruz Gonzalez et al., 2025). A study by Wolf et al. (2019) evaluated a novel screening algorithm using the Strengths and Difficulties Questionnaire (SDQ) to identify children at risk of life-altering mental health problems (MHP). By combining parental SDQ responses with comprehensive cohort data, they found that the algorithm effectively identified children with MHP who subsequently demonstrated poor school performance, highlighting its potential for proactive identification and support. A systematic review and meta-analysis by Wang et al. (2025) evaluated the effectiveness of AI-assisted multimodal approaches in depression screening, using physiological and behavioral data such as electroencephalography (EEG), eye movement, and gait analysis. They found that multi-modal AI methods outperformed the traditional uni-modal methods, with deep learning models showing the highest accuracy. Crowley et al. (2025) also highlight the potential of AI-based multi-modal techniques to enhance depression detection. Their study explored the use of machine learning to predict childhood mental health problems among children in social care in Wales using linked data from the SAIL Databank. The gradient boosting classifier performed best, identifying risk factors such as age, substance misuse, and being in a social care setting.

AI is also used in conversational agents and automated psychosocial interventions. Chatbots and automated cognitive-behavioral therapy (CBT) modules provide psychoeducation, stress management strategies, and brief coping exercises at scale (Feng et al., 2025; Nicol et al., 2022). These tools serve as preventive support and aim to assist individuals experiencing mild or early symptoms before full disorders develop. Recent randomized controlled trials and meta-analyses report small-to-moderate reductions in anxiety and depressive symptoms, indicating a meaningful but modest preventive impact (Fitzpatrick et al., 2017; Walder et al., 2025).

The various research presented above provides evidence that AI can enhance early detection of mental health risks by analyzing extensive data patterns to identify individuals needing early intervention. Studies indicate that these models often perform well retrospectively, demonstrating strong discrimination of high-risk groups. It is increasingly integrated into clinical support tools, including digital therapeutics, which are validated software programs that assist treatment. While promising for early intervention, the study highlights concern about algorithmic fairness and the need for equitable, validated models across diverse populations. They also highlight a gap in prospective validation, which limits how confidently such tools can be used in real-world prevention settings (Ghassemi et al., 2020; Shatte et al., 2019).

Treatment and Intervention

Studies have shown the use and effectiveness of AI technology in treating mental health problems. These include the use of Conversational Agents and Digital Cognitive-Behavioural Therapy **(**CBT), Digital Therapeutics and AI Decision-Support Tools, and Personalized Interventions and Adaptive Content.

Conversational Agents and Digital Cognitive-Behavioural Therapy

One of the most visible uses of AI in mental health treatment is the growth of conversational agents (CAs), ranging from simple rule-based programs to advanced chatbots powered by machine learning and natural language processing. Platforms like Woebot, Wysa, Tess, and Youper use chat-based interfaces to deliver core CBT strategies, providing psychoeducation, cognitive restructuring, behavioral activation, mood tracking, and coping support (Hawke et al., 2025). Evidence from randomized controlled trials and meta-analyses shows that conversational agents can produce small-to-moderate reductions in depression and anxiety symptoms in young people, especially in the short term (Hawke et al., 2025; He et al., 2023). More recent reviews continue to find short-term benefits of AI chatbots to modestly reduce mental distress and improve health behaviors in adolescents and young adults, especially for depression, anxiety, and stress, with effectiveness influenced by design, delivery, and user engagement (Feng et al., 2025; Li et al., 2025). However, Humayun et al. (2025) note that AI interventions in mental healthcare tend to mostly produce modest, short-term benefits, as most studies rely on brief intervention periods with limited follow-up, making it unclear whether early improvements are sustained over time, and some evidence suggests these gains may diminish without ongoing support.

Personalized Interventions and Adaptive Content

Another growing area uses AI to personalize treatment over time. These systems learn from a person's behavior and test different support options to see what works best and when. This makes it possible to deliver *just-in-time* support, where digital help is adjusted to fit the user's current needs, habits, and situation (Nahum-Shani et al., 2018). AI systems can create customized treatment plans by predicting how someone might respond to certain medications or by adjusting cognitive-behavioural therapy (CBT) strategies in real time (Cruz Gonzalez et al., 2025). This makes therapy more responsive and better suited to the individual. Virtual therapists, such as the chatbots Woebot and Wysa, offer personalized CBT techniques and coping tools anytime they are needed. These tools provide private, 24/7 support and can help people manage stress, anxiety, and low mood (Chiauzzi et al., 2023).

Emerging work also examines how large language models (LLMs) can help tailor intervention content to users' expressed preferences. Yoon (2024) designed a pilot study aimed at assessing whether an AI-based digital therapeutic system could enhance adolescent mental health by delivering personalized and adaptive interventions. Using machine learning, reinforcement learning, and multimodal data (such as text, voice, and behavioral inputs), the system continuously built and updated individual user profiles to tailor interventions like CBT exercises and mindfulness activities based on users' needs, engagement, and emotional states. Findings indicated that this personalized approach led to improved outcomes, including higher overall well-being, greater adherence to recommended interventions, and meaningful reductions in stress, anxiety, and depression compared with control groups (Yoon, 2024).

In addition, researchers are exploring new forms of AI-based interventions. For example, robotic personal companions like Sony's Aibo have been tested for their ability to reduce stress markers in the body and increase oxytocin, a hormone linked to bonding and well-being (Yamada et al., 2024). Wanniarachchi et al. (2025) noted that personalized digital mental health interventions are effective in improving engagement and adherence among adolescents and young people by tailoring content, delivery, and user experiences to individual needs. According to their research, strategies such as customized therapeutic content, user choice in intervention pathways, and adaptive features help make interventions more relevant and motivating, addressing common challenges with sustained participation in standardized programs. These studies provide evidence that AI helps make mental health support more personalized and easier to access. It can tailor treatment to each person's needs and provide support at a much larger scale than traditional care alone. However, studies on patient acceptability indicate that while users value the immediacy and accessibility of chatbots, they prefer human support in complex situations and express concerns about empathy, privacy, and the need for clear crisis-response protocols (Chaudhry & Debi, 2024; Hipgrave et al., 2025).

This underscores the need to safeguard young people, especially with the rising rates of mental health crises, self-harm, and suicidal ideation post-pandemic (Eapen et al., 2023). This entails identifying risks early and giving quick, organized support to keep them safe (Edwards et al., 2024). The process requires a coordinated team, including mental health specialists, crisis staff, trained counsellors, school and research safeguarding leads, community workers, peers, and families (Randhawa et al., 2024).

Effective safeguarding must be child-centred, developmentally appropriate, and culturally safe, actively involving young people in service design, while providing trauma-informed responses that identify distress in real time, communicate confidentiality limits clearly, and ensure accountability through designated safeguarding leaders (Edwards et al., 2024). The protocol for safeguarding children and adolescents' mental health involves a systematic progression from proactive identification to immediate stabilization and integrated care**.** This pathway is to ensure that no young person "slips through the gaps" during a mental health crisis (Edwards et al., 2024; Randhawa et al., 2024). Emerging evidence shows that through digital interventions, psychoeducation, and coping strategies, AI can reach adolescents who may face barriers to traditional care, improving accessibility while supporting professionals in monitoring and responding to youth at risk. Safeguarding is further strengthened when AI is integrated into the existing support systems to provide calm, private, and specialist-led care, ensuring young people receive timely, contextually appropriate, and effective protection from harm (Randhawa et al., 2024; Sharma et al., 2025).

The summary of the use of AI in mental healthcare for children and adults is presented in Table 1.

Table 1. Youth and Children-Focused Applications of AI in Mental Health Care

AI Task	Primary Use	Setting(s)	Outcome Type	Evidence Level	Age Range of Participants	Sources
Risk prediction & early warning (ML on administrative/social care data)	Prevention / Early detection	School, social care, clinic	Clinical (risk identification, service use)	Retrospective cohort studies; external validation studies	5–18 years	Crowley et al. (2025); Wolf et al. (2019)
Symptom screening (text, speech, multimodal data)	Diagnosis / Screening	School, clinic, telehealth	Clinical (screening accuracy, sensitivity/specificity)	Systematic reviews & meta-analyses	6–18 years	Phiri et al. (2025); Wang et al. (2025); Shatte et al. (2019)
Neuroimaging-based classifiers in children & adolescents	Diagnosis	Clinic	Clinical (classification accuracy)	Methodological validation studies (limited generalizability)	8–17 years	Chen et al. (2023)
Digital phenotyping via smartphones	Monitoring	Home, telehealth	Clinical (symptom trajectories, relapse signals)	Feasibility studies; longitudinal cohort studies	12–18 years	Onnela & Rauch (2016); Huang et al. (2025); McNeilly et al. (2023)
Conversational agents (supportive chatbots)	Prevention / Early support	Home, telehealth	Engagement; Clinical (distress reduction)	Systematic reviews & meta-analyses of RCTs	10–18 years	He et al. (2023); Feng et al. (2025); Li et al. (2025); Chaudhry & Debi (2024)
Chatbot-delivered CBT (children & adolescents)	Treatment / Intervention	Home, telehealth	Clinical (depression/anxiety outcomes)	Randomized controlled trials; feasibility studies	13–18 years	Fitzpatrick et al. (2017); Nicol et al. (2022)
Relational agents for children & adolescents	Treatment augmentation	Clinic, telehealth	Engagement, adherence	RCT protocols; early-phase trials	12–17 years	Chiauzzi et al. (2023)
AI-assisted online social therapy	Treatment / Recovery support	Home, online platforms	Engagement, social functioning	Pilot studies; mixed-methods evaluations	14–18 years	D'Alfonso et al. (2017)
Just-in-time adaptive interventions (JITAIs) for youth	Prevention / Relapse prevention	Home, mobile	Clinical & behavioral outcomes	Conceptual frameworks; emerging youth trials	13–18 years	Nahum-Shani et al. (2018); Pratap et al. (2022)

Social media-based depression detection (children & adolescents)	Prevention / Surveillance	Online environments	Population-level risk signals	Systematic reviews & meta-analyses	12–17 years	Phiri et al. (2025)
AI-enabled crisis detection & triage	Monitoring / Acute intervention	Clinic, emergency & telehealth	Clinical (stabilization, referral)	Service protocols; observational evaluations	5–18 years	Eapen et al. (2023); Edwards et al. (2024)
Personalization algorithms in youth DMHIs	Treatment optimization	Home, telehealth	Engagement, treatment response	Scoping reviews of youth interventions	12–18 years	Wanniarachchi et al. (2025)
AI-based digital therapeutics for children & adolescents	Treatment / Ongoing management	Home, telehealth	Clinical outcomes; adherence	Narrative reviews; selected controlled trials	10–18 years	Yoon (2024); Walder et al. (2025)
Explainable AI for youth mental health tools	Cross-cutting (trust & safety)	Clinic, school	Engagement, clinician trust	Conceptual papers; systematic reviews	5–18 years	Amann et al. (2020); Rosenbacke et al. (2024)

Evidence level definitions:
- **Systematic review & meta-analysis** = highest level of synthesized evidence
- **Randomized controlled trial (RCT)** = causal evidence of effectiveness
- **Feasibility / pilot study** = implementation readiness, not effectiveness
- **Protocol / conceptual framework** = design and theory-building stage

Hybrid Human–AI Care Models

Many health systems are moving toward blended care, where AI handles low-intensity tasks such as psychoeducation, automated monitoring, or reinforcement scheduling, while clinicians address more complex or high-risk needs. In the mental health care context, AI and clinicians work together in a complementary care pathway in which AI systems screen and continuously monitor mental health indicators, flag individuals at potential risk, and clinicians then review these insights to make informed decisions about escalation, diagnosis, and appropriate treatment. For instance, clinicians can pair traditional therapy with digital interventions to support cognitive-behavioral therapy practices at home. This was demonstrated by D'Alfonso et al. (2017), who evaluated the moderated online social therapy (MOST) system, an online platform combining peer support, evidence-based interventions, and clinician/consumer-centered delivery, demonstrated viability in clinical trials for preventing relapses in young people recovering from psychosis or depression. The study emphasizes the need to integrate AI and automated content delivery to scale the platform, personalize interventions in real time, and complement human moderators for improved therapeutic reach and effectiveness.

AI increasingly appears in clinical decision-support systems, which assist clinicians in choosing treatments, matching patients to therapies, assessing risk, and identifying early signs of deterioration. Systematic reviews suggest these machine-learning models perform well on internal validation tasks but also stress that generalizability and external validation remain major challenges, limiting routine clinical use (Cruz-Gonzalez et al., 2025). The Food and Drug Administration has authorized digital treatments that work like medical devices. One example is reSET, a prescription app used with 12-week regular treatment to help people with substance use disorder by giving them cognitive-behavioral therapy on their phones, to be used in conjunction with outpatient clinician-delivered care (Novartis, 2018). Furthermore, studies show that AI should primarily complement clinicians by supporting, not replacing, them; by helping to personalize treatment plans, aiding clinical decision-making, and providing additional services alongside human-led care, they reduce clinicians' workload and allow them to spend more time on direct patient care (Ni & Jia, 2025). In summary, AI-assisted tools like conversational agents and app-based CBT consistently produce short-term symptom reductions and broaden access to care, with evidence suggesting their effectiveness is maximized when paired with personalization and explicit human oversight.

Navigating Ethical, Technical, and Clinical Barriers to Responsible AI Adoption in Mental Health Care and Way Forward

The integration of Artificial Intelligence into mental health care shows strong potential for expanding access, enhancing personalization, and improving decision-making. However, the adoption of AI is hindered by interconnected technological, human, and ethical–legal issues that reduce trust, slow implementation, and constrain the safe and equitable use of AI (Abd-Alrazaq et al., 2025). Technical weaknesses include poor explainability, bias, limited generalizability, validation difficulties, and integration with existing clinical systems; human factors include resistance to change, insufficient training, weak stakeholder engagement, and lack of resources; and ethical–legal challenges are related to issues of privacy, consent, liability, equity, and unclear or fragmented regulation (Abd-Alrazaq et al., 2025; Moreno-Sánchez et al., 2026).

Mental health data, including therapy notes, behavioral patterns, and personal disclosures, are extremely sensitive and vulnerable to misuse. Hence, the use of AI in mental health care raises privacy concerns, especially with ongoing uncertainty around consent, data ownership, and acceptable use (Löchner et al., 2025). Bias is another key ethical concern. AI systems, often trained on datasets dominated by high-income country populations, show lower diagnostic accuracy for depression in some ethnic minority groups, which can lead to unfair treatment recommendations for marginalized communities (Narimani & Naeim, 2025). On the other hand, Föyen et al. (2025) observed that clinicians are also biased against the use of AI for providing health guidance, even when they perceive that AI can provide health advice comparable to experts in empathy and quality, highlighting the need to address human biases for broader AI acceptance in practice. Clinician acceptance is found to be higher when referral procedures are well-defined, operated under human supervision, and workflows are co-designed with end-users (Abd-Alrazaq et al., 2025). Similarly, patients trust AI tools more when a clinician is involved, and new policies and payment rules help make this kind of AI-plus-human care safer and more sustainable (Foresman et al., 2025).

Model explainability also remains a concern, as AI systems often identify patterns and correlations in data rather than true causal relationships, which can limit their clinical usefulness (Rosenbacke et al., 2024). As models become more complex, particularly deep learning systems and large language models, it becomes increasingly difficult to understand how specific inputs lead to specific outputs. This lack of transparency can undermine clinician trust, hinder safeguarding and accountability, and make it harder to identify bias or errors in decisions affecting young people (Amann et al., 2020). In addition, many advanced AI models, particularly deep learning systems, function as "black boxes," producing predictions without offering clear explanations of how specific outputs are derived, thereby reducing clinician trust and approval (Ali et al., 2025). In addition, mental health datasets often have small sample sizes, inconsistent labeling, and subjective assessments, which weaken model performance and reliability, and many systems also fail to demonstrate generalizability because they are not rigorously tested on diverse populations or real-world conditions (Chen et al., 2023; Pratap et al., 2022). Furthermore, AI chatbots struggle to interpret emotional nuance because they rely primarily on text and lack access to non-verbal cues such as facial expressions, tone, and body language, making their responses less emotionally attuned than human interactions (Sharon, 2025). Even AI-simulated empathy in healthcare can create the appearance of caring and improve patient satisfaction, but it risks undermining genuine empathy, deceiving patients, and diverting attention from systemic barriers that prevent real human-centered care, while introducing new burdens and safety risks from AI errors (Sharon, 2025).

Many AI systems do not fit smoothly into existing clinical workflows, and practitioners frequently view clinicians' judgment as more valuable than algorithmic recommendations. This misalignment can lead to low adoption, even when the technology is promising. Hong and Tartaglia (2025) observed that clinicians report a mix of enthusiasm and caution regarding AI adoption. While clinicians recognize potential benefits, including improved access to patient data for personalized treatment and potential cost savings, they also identify significant challenges, which include the added pressure to understand complex algorithms, increased risk of burnout, concerns that AI could compromise clinical competency, privacy and data security issues, and the need for stronger regulation to address ethical considerations and prevent patient harm (Hong & Tartaglia, 2025).

Moving forward, these limitations shape the priorities for future research, design, and regulation. So, there is a critical need to ensure that AI complements rather than replaces these human elements. Meta-analyses and systematic reviews repeatedly call for larger, preregistered trials, external validation, robust safety monitoring, and extended follow-up to determine whether AI-supported treatments yield long-term improvements in mental health (Feng et al., 2025). Researchers (Ali et al., 2025; Pratap et al., 2022; Sharon, 2025; Shatte et al., 2019) insist that the development and use of AI in mental healthcare should emphasize responsible, supportive, and human-centered integration. Key priorities include using AI as a decision-support tool that augments clinicians rather than replacing core human values like empathy, ensuring explainable AI (XAI) for transparency and trust, and addressing algorithmic bias through diverse, inclusive datasets. Advancing digital mental health also requires expanding applications beyond traditional diagnoses to under-researched conditions, leveraging multiple data sources (e.g., surveys, social media, longitudinal behavioral data), and improving real-world validation of AI models. Ethical and practical governance, covering privacy, accountability, and mindful problem definition, is essential to ensure AI addresses the structural barriers preventing empathetic care rather than creating "orphan problems. Finally, building large, accessible, findable, accessible, interoperable, and reusable (FAIR)-compliant datasets is critical to enable robust, representative, and reproducible research, paving the way for personalized, effective, and culturally sensitive AI-supported mental healthcare (Ali et al., 2025; Pratap et al., 2022; Sharon, 2025; Shatte et al., 2019).

Conclusion

Artificial intelligence (AI) offers transformative potential for the prevention, detection, monitoring, and treatment of mental health disorders among children and adolescents. Evidence demonstrates that AI-driven approaches, including machine learning, natural language processing, multimodal data integration, conversational agents, and digital therapeutics, can enhance early identification of risk, personalize interventions, and expand access to care, particularly for underserved populations. Hybrid human-AI care models further highlight the importance of pairing technological tools with clinician oversight to optimize outcomes while maintaining empathy, trust, and safety. However, significant challenges remain. Algorithmic bias, limited interpretability, small or non-representative datasets, ethical concerns around privacy and consent, and integration with existing clinical workflows all constrain the real-world adoption and reliability of AI in mental health care. Patient and clinician acceptance is highest when AI tools complement rather than replace human judgment, incorporate transparent decision-making, and are designed with end-user input. Moving forward, responsible and human-centered AI integration is essential. This includes prioritizing explainable and fair AI, robust external validation, diverse and representative datasets, and ethical governance that safeguards privacy, equity, and trust. By addressing these challenges, AI has the potential to enhance early intervention, improve treatment personalization, and ultimately support equitable, scalable, and culturally sensitive mental healthcare for children and adolescents worldwide.

Recommendations

Based on this review, it is recommended that AI tools developers, researchers, and regulators:

1. Ensure AI tools are tested on diverse youth populations particularly, the underrepresented populations and conditions, before clinical or educational deployment.
2. Collaborate with clinicians, schools, and families to develop user-centered and context-appropriate AI workflows.
3. Integrate privacy-by-design principles and establish clear crisis-response protocols in AI-powered youth mental health tools.

4. Develop explainable and equitable models, validating tools in real-world settings, and building large, FAIR-compliant datasets.

Ethical Declaration: The article is a narrative review, involving no human participants; therefore, no ethical approval was required.
Conflicts of Interest and Data Availability: There was no conflict of interest in this work, and no data was generated.

References

Abd-Alrazaq, A., Solaiman, B., Mekki, Y. M., Al-Thani, D., Farooq, F., Alkubeyyer, M., Abubacker, M. Z., AlSaad, R., Aziz, S., Serag, A., Thomas, R., Sheikh, J., & Ahmed, A. (2025). Hype vs reality in the integration of artificial intelligence in clinical workflows. *JMIR Formative Research*, *9*(1), e70921. https://doi.org/10.2196/70921

Abiodun, O. A., Ajiboye, P. O., Salihu, M. O., Sulyman, D., Akinsulore, A., Obayi, O., & Salihu, H. B. (2025). Psychiatrists' and trainees' knowledge, perception, and readiness for integration of artificial intelligence in mental health care in Nigeria. *BMC Psychiatry*, *25*(1), 1–13. https://doi.org/10.1186/s12888-025-07135-1

Ali, M., Ali, S., Abbas, Q., Abbas, Z., & Lee, S. W. (2025b). Artificial intelligence for mental health: A narrative review of applications, challenges, and future directions in digital health. *Digital Health*, *11*. 1-12; https://doi.org/10.1177/20552076251395548

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*(1), 1–9. https://doi.org/10.1186/s12911-020-01332-6

American Psychological Association. (2024, March 12). *Artificial intelligence in mental health care*. https://www.apa.org/practice/artificial-intelligence-mental-health-care

Arvai, N., Katonai, G., & Mesko, B. (2025). Health care professionals' concerns about medical AI and psychological barriers and strategies for successful implementation: Scoping review. *Journal of Medical Internet Research*, *27*(1). https://doi.org/10.2196/66986

Blease, C., & Rodman, A. (2025). Generative artificial intelligence in mental healthcare: An ethical evaluation. *Current Treatment Options in Psychiatry*, *12*(1). https://doi.org/10.1007/s40501-024-00340-x

Canadian Mental Health Association. (2025, November 3). *More people are using AI for mental health care, but are we ready?* https://cmha.ca/news/ai-mental-health/

Chaudhry, B. M., & Debi, H. R. (2024). User perceptions and experiences of an AI-driven conversational agent for mental health support. *Mhealth*, *10*, 22. https://doi.org/10.21037/MHEALTH-23-55/PRF

Chen, Z., Hu, B., Liu, X., Becker, B., Eickhoff, S. B., Miao, K., Gu, X., Tang, Y., Dai, X., Li, C., Leonov, A., Xiao, Z., Feng, Z., Chen, J., & Chuan-Peng, H. (2023). Sampling inequalities affect generalization of neuroimaging-based diagnostic classifiers in psychiatry. *BMC Medicine*, *21*(1), 241. https://doi.org/ 10.1186/S12916-023-02941-4

Chiauzzi, E., Robinson, A., Martin, K., Petersen, C., Wells, N., Williams, A., & Gleason, M. M. (2023). A Relational agent intervention for adolescents seeking mental health treatment: protocol for a randomized controlled trial. *JMIR Research Protocols*, *12*, 1–11. 12:e44940. https://doi.org/10.2196/44940

Crowley, R., Parkin, K., Rocheteau, E., Massou, E., Friedmann, Y., John, A., Sippy, R., Liò, P., & Moore, A. (2025). Machine learning for prediction of childhood mental health problems in social care. *BJ Psych Open*, *11*(3), 1–9. https://doi.org/10.1192/bjo.2025.32

Cruz-Gonzalez, P., He, A. W. J., Lam, E. P. 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. In *Psychological Medicine* (Vol. 55). https://doi.org/10.1017/S0033291724003295

D'Alfonso, S., Santesteban-Echarri, O., Rice, S., Wadley, G., Lederman, R., Miles, C., Gleeson, J., & Alvarez-Jimenez, M. (2017). Artificial intelligence-assisted online social therapy for youth mental health. *Frontiers in Psychology*, *8*(JUN), 1–13. https://doi.org/10.3389/fpsyg.2017.00796

Eapen, V., Gerstl, B., Winata, T., Jairam, R., Barton, G., & Bowden, M. (2023). A study protocol for safeguards child and adolescent mental health rapid response teams ('safeguards teams') service. *International Journal of Integrated Care*, *23*(3), 1–10. https://doi.org/10.5334/ijic.7004

Edwards, D., Carrier, J., Csontos, J., Evans, N., Elliott, M., Gillen, E., Hannigan, B., Lane, R., & Williams, L. (2024). Crisis responses for children and young people: A systematic review of effectiveness, experiences and service organisation (CAMH-Crisis). *Child and Adolescent Mental Health*, *29*(1), 70–83. https://doi.org/10.1111/camh.12639

Feng, X., Tian, L., Ho, G. W. K., Yorke, J., & Hui, V. (2025). The effectiveness of AI chatbots in alleviating mental distress and promoting health behaviors among adolescents and young adults: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *27*(1), e79850. https://doi.org/10.2196/79850

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*(2), 1–11. https://doi.org/10.2196/mental.7785

Foresman, G., Biro, J., Tran, A., MacRae, K., Kazi, S., Schubel, L., Visconti, A., Gallagher, W., Smith, K. M., Giardina, T., Haskell, H., & Miller, K. (2025). Patient perspectives on artificial intelligence in health care: Focus group study for diagnostic communication and tool implementation. *Journal of Participatory Medicine*, *17*(1), e69564. https://doi.org/10.2196/69564

Föyen, L. F., Zapel, E., Lekander, M., Hedman-Lagerlöf, E., & Lindsäter, E. (2025). Artificial intelligence vs. human expert: Licensed mental health clinicians' blinded evaluation of AI-generated and expert psychological advice on quality, empathy, and perceived authorship. *Internet Interventions*, *41*, 100841. https://doi.org/10.1016/J.INVENT.2025.100841

Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., Chen, I. Y., & Ranganath, R. (2020). A review of challenges and opportunities in machine learning for health. *AMIA Joint Summits on Translational Science Proceedings. AMIA Joint Summits on Translational Science*, *2020*, 191–200. http://www.ncbi.nlm.nih.gov/pubmed/32477638%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7233077

Grenon, S., Hoang, K. N., Luo, S., & Koch, A. (2025). Adolescent mental health in rural settings: the role of artificial intelligence and community engagement. *Frontiers in Public Health*, *13*(September), 1–9. https://doi.org/10.3389/fpubh.2025.1643466

Hawke, L. D., Hou, J., Nguyen, A. T. P., Phi, T., Gibson, J., Ritchie, B., Strudwick, G., Rodak, T., & Gallagher, L. (2025). Digital conversational agents for the mental health of treatment-seeking youth: Scoping review. *Journal of Medical Internet Research*, *12*. e77098. https://doi.org/10.2196/77098

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

Hipgrave, L., Goldie, J., Dennis, S., & Coleman, A. (2025). Balancing risks and benefits: clinicians' perspectives on the use of generative AI chatbots in mental healthcare. *Frontiers in Digital Health*, *7:1606291*. https://doi.org/10.3389/fdgth.2025.1606291

Hong, S. K., & Tartaglia, J. (2025). Exploring clinician perspectives on the acceptability and feasibility of AI-mediated tools in child and adolescent psychiatric care: A scoping review. *Journal of the American Academy of Child & Adolescent Psychiatry*, *64*(10), S308–S309. https://doi.org/10.1016/j.jaac.2025.08.489

Huang, D., Emedom-Nnamdi, P., Onnela, J. P., & Van Meter, A. (2025). Design and feasibility of smartphone-based digital phenotyping for long-term mental health monitoring in adolescents. *PLOS Digital Health*, *4*(7 July), 1–11. https://doi.org/10.1371/journal.pdig.0000883

Jones, A., Torous, J., & Amir, T. (2025). Balancing promise and risk: Ethical considerations for GenAI in mental health care. *JMIR Mental Health*, *12*, 12:e70439. https://doi.org/10.2196/ 70439

Kamdan, K., Fauziyah, N. G., Fadlullah, M. A., Hanif, D. A., & Kharisma, I. L. (2025). Early mental health detection and emotional states in teenagers through chatbot systems using natural language processing (NLP). *Engineering Proceedings*, *107*(1), 64. https://doi.org/10.3390/engproc2025107064

Li, J., Li, Y., Hu, Y., Ma, D. C. F., Mei, X., Chan, E. A., & Yorke, J. (2025). Chatbot-delivered interventions for improving mental health among young people: A systematic review and meta-analysis. *Worldviews on Evidence-Based Nursing*, *22*(4), e70059. https://doi.org/ 10.1111/WVN.70059

Löchner, J., Carlbring, P., Schuller, B., Torous, J., & Sander, L. B. (2025). Digital interventions in mental health: An overview and future perspectives. *Internet Interventions*, *40*; 100824. https://doi.org/10.1016/j.invent.2025.100824

McGinty, B. (2023). The future of public mental health: Challenges and opportunities. *Milbank Quarterly*, *101*(1), 532–551. https://doi.org/10.1111/1468-0009.12622

McNeilly, E. A., Mills, K. L., Kahn, L. E., Crowley, R., Pfeifer, J. H., & Allen, N. B. (2023). Adolescent social communication through smartphones: Linguistic features of internalizing symptoms and daily mood. *Clinical Psychological Science*, *11*(6), 1090–1107. https://doi.org/10.1177/21677026221125180

Moreno-Sánchez, P. A., Del Ser, J., van Gils, M., & Hernesniemi, J. (2026). A design framework for operationalizing trustworthy artificial intelligence in healthcare: Requirements, tradeoffs and challenges for its clinical adoption. *Information Fusion*, *127*, 103812. https://doi.org/10.1016/j.inffus.2025.103812

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, *52*(6), 446–462. https://doi.org/10.1007/s12160-016-9830-8

Narimani, M., & Naeim, M. (2025). Artificial intelligence in mental health: integrating opportunities and challenges of multimodal deep learning for mental disorder prevention and treatment. *Annals of Medicine and Surgery*, *87*(9), 5757. https://doi.org/10.1097/MS9.0000000000003624

Ni, Y., & Jia, F. (2025). A scoping review of AI-driven digital interventions in mental health care: Mapping applications across screening, support, monitoring, prevention, and clinical education. *Healthcare (Switzerland)*, *13*(10), 1–23. https://doi.org/10.3390/healthcare13101205

Nicol, G., Wang, R., Graham, S., Dodd, S., & Garbutt, J. (2022). Chatbot-delivered cognitive behavioral therapy in adolescents with depression and anxiety during the COVID-19 pandemic: feasibility and acceptability study. *JMIR Formative Research*, *6*(11): e40242 https://doi.org/10.2196/40242

Novartis. (2018, November 20). *Sandoz and Pear Therapeutics announce launch of reSET® for treatment of patients with substance use disorder*|. https://www.novartis.com/news/media-releases/sandoz-and-pear-therapeutics-announce-launch-reset-treatment-patients-substance-use-disorder?utm_source=chatgpt.com

Onnela, J., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology, 41,* 1691–1696. https://doi.org/10.1038/npp.2016.7

Phiri, D., Makowa, F., Amelia, V. L., Phiri, Y. V. A., Dlamini, L. P., & Chung, M. H. (2025). Text-based depression prediction on social media using machine learning: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *27*(1), e59002. https://doi.org/10.2196/59002

Poudel, U., Jakhar, S., Mohan, P., & Nepal, A. (2025). AI in mental health: A review of technological advancements and ethical issues in psychiatry. *Issues in Mental Health Nursing*, *46*(7), 693–701. https://doi.org/10.1080/01612840.2025.2502943

Pratap, A., Homiar, A., Waninger, L., Herd, C., Suver, C., Volponi, J., Anguera, J. A., & Areán, P. (2022). Real-world behavioral dataset from two fully remote smartphone-based randomized clinical trials for depression. *Scientific Data*, *9*(1), 1–9. https://doi.org/10.1038/s41597-022-01633-7

Randhawa, A., Wood, G., Michail, M., Pallan, M., Patterson, P., & Goodyear, V. (2024). Safeguarding in adolescent mental health research: navigating dilemmas and developing procedures. *BMJ Open*, *14*(2), 1–7. https://doi.org/10.1136/bmjopen-2023-076700

Rosenbacke, R., Melhus, Å., McKee, M., & Stuckler, D. (2024). How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: Systematic review. *JMIR AI, 3,* 1–10. https://doi.org/10.2196/53207

Sharon, T. (2025). Technosolutionism and the empathetic medical chatbot. *AI and Society*, *0123456789*. https://doi.org/10.1007/s00146-025-02441-4

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

Stephan, J., Gehrmann, J., Sinha, M., Stullich, A., Gabel, F., & Richter, M. (2025). A scoping review of prevention classification in mental health: Examining the application of Caplan's and Gordon's prevention frameworks (2018–2024). *Journal of Prevention*, *46*(3), 427–454. https://doi.org/10.1007/ s10935-025-00834-1

Walder, N., Frey, A., Berger, T., & Schmidt, S. J. (2025). Digital mental health interventions for the prevention and treatment of social anxiety disorder in children, adolescents, and young adults: Systematic review and meta-analysis of randomized controlled trials. *Journal of Medical Internet Research, 27(1): e67067.* https://doi.org/10.2196/67067

Wang, L., Wang, C., Li, C., Murai, T., Bai, Y., Song, Z., Zhang, S., Zhang, Q., Huang, Y., Bi, X., & Jiang, J. (2025). AI-assisted multi-modal information for the screening of depression: A systematic review and meta-analysis. *Npj Digital Medicine*, *8*(1), 1–14. https://doi.org/10.1038/s41746-025-01933-3

Wang, X., Zhou, Y., & Zhou, G. (2025). The application and ethical implications of generative AI in mental health: Systematic review. *JMIR Mental Health, 12*(1), e70610. https://doi.org/ 10.2196/70610

Wanniarachchi, V. U., Greenhalgh, C., Choi, A., & Warren, J. R. (2025). Personalization variables in digital mental health interventions for depression and anxiety in adolescents and youth: a scoping review. *Frontiers in Digital Health*, 7:1500220. https://doi.org/10.3389/fdgth.2025.1500220

WHO. (2025, September 1). *Mental health of adolescents*. Fact Sheets. https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health

Wolf, R. T., Jeppesen, P., Gyrd-Hansen, D., & Oxholm, A. S. (2019). Evaluation of a screening algorithm using the strengths and difficulties questionnaire to identify children with mental health problems: A five-year register-based follow-up on school performance and healthcare use. *PLoS ONE*, *14*(10), 1–18. https://doi.org/10.1371/journal.pone.0223314

Yamada, A., Akahane, D., Takeuchi, S., Miyata, K., Sato, T., & Gotoh, A. (2024). Robot therapy aids mental health in patients with hematological malignancy during hematopoietic stem cell transplantation in a protective isolation unit. *Scientific Reports*, *14*(1), 1–13. https://doi.org/10.1038/s41598-024-54286-4

Yoon, S. (2024). AI-based digital therapeutics for adolescent mental health management and disaster response. *Information*, *15(10), 620*. https://doi.org/10.3390/info15100620