Cureus
Part of SPRINGER NATURE

Open Access Review Article

Conversational AI and Psychosis: A Proposed Neurocognitive Framework Involving Anthropomorphism, Predictive Coding, and Delusional Reinforcement

Review began 07/27/2026
Review ended 10/19/2026
Published 11/02/2026

© Copyright 2026
Banik et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.118999

Pratik Kumar Banik [1], Heena Arshad [2], Aadil Ahmed Khan [3], Margarita Zhuzha [4], Inamullah Arshad [5], Amuksha Ravindra Shetty [6], Srushti Patil [7], Rahat Rehman [8]

1. Department of Psychiatry, Chittagong Medical College Hospital, Chattogram, BGD 2. Department of Internal Medicine, Sargodha Medical College, Sargodha, PAK 3. Department of Behavioral Health, Saint Elizabeths Hospital, Washington DC, USA 4. Department of Psychiatry, First Pavlov State Medical University of St. Petersburg, St. Petersburg, RUS 5. Department of Internal Medicine, Shalamar Medical and Dental College, Lahore, PAK 6. Department of Psychiatry and Behavioral Sciences, Dr. Balabhai Nanavati Hospital, Mumbai, IND 7. Department of Internal Medicine, SMBT Medical College, Igatpuri, IND 8. Department of Medicine, Shaheed Suhrawardy Medical College Hospital, Dhaka, BGD

Corresponding author: Pratik Kumar Banik, pratikkumarbanik@gmail.com

Abstract

Conversational AI has become an integral part of everyday life. Historical evidence shows that sociocultural change can be reflected in mental health, particularly in the incorporation of new technologies, suggesting that conversational AI represents a new context through which such changes may influence the manifestation and content of psychotic symptoms. This narrative review explores the possible relationship between conversational AI use and the development or worsening of psychotic symptoms, bringing together the limited clinical evidence, possible neurocognitive mechanisms, relevant vulnerability factors, and the potential role of conversational AI in therapeutic and pathological contexts. A non-systematic review of the literature was conducted using PubMed and Google Scholar, with relevant clinical case reports, original research articles, reviews, and conceptual articles identified through keyword searches and reference list screening. Conversational AI use has been associated with psychotic-like experiences in cross-sectional survey data and has been implicated in the onset or relapse of psychotic symptoms in an isolated case report. Several theoretical neurocognitive mechanisms have also been proposed by which conversational AI could contribute to the emergence or worsening of psychotic symptoms in vulnerable individuals. These mechanisms are largely derived from existing psychosis models and remain hypothesis-generating rather than empirically demonstrated, requiring further validation through longitudinal and experimental studies. At the same time, in carefully defined clinical contexts, conversational AI may also have supportive uses in assessment, monitoring, or structured therapeutic support. Given the current context, the use of conversational AI should be approached with caution; adherence to ethical standards and careful consideration of both its benefits and associated psychological risks are vital.

Categories: Neurology, Psychiatry, Psychology
Keywords: artificial intelligence, chatbot, delusion, hallucination, large language models, psychosis

Introduction And Background

People use conversational AI for a variety of purposes, such as searching for information, learning, planning support, and time optimization, and it has become an essential part of daily life. Simultaneously, users are turning more and more to conversational AI for emotional support, decision-making help, and loneliness alleviation, thereby simulating aspects of social connection often associated with human-to-human contact [1,2]. Historically, new technologies have often shaped the content of delusions and other psychotic experiences [3,4]. Drawing on historical experience, it may be posited that conversational AI can influence the phenomenology of psychotic experiences in the present day, as it offers personalization, adaptability, interactivity, and the capacity to provide feedback resembling human emotional engagement [2,5].

In contrast to previous technologies, conversational AI creates the illusion of reciprocal exchange by simulating agency, memory, and tailored dialogues in addition to transmitting information. This may be clinically relevant because its responses can be experienced as meaningful, attuned, or personally directed, despite the absence of consciousness, subjective experience, or genuine social intention [2,5]. This discrepancy between the system's non-human nature and the illusion of interpersonal connection may lead to misinterpretation, over-attribution of meaning, or an increasing emotional dependence on AI interactions among vulnerable people [1,5]. In this review, 'conversational AI' refers to open-ended generative chatbots, such as large language model (LLM)-based assistants and companion apps, which produce free-form, personalized responses.

The connection between conversational AI and psychosis is still not well understood, despite increasing professional and public interest. The term 'AI-related psychotic symptoms' is used descriptively throughout this review for delusions, hallucinations, or other psychotic experiences in which conversational AI features in the content or course of symptoms; it does not denote an established diagnostic entity. At present, the scientific evidence in this area consists predominantly of case reports, cross-sectional studies, and conceptual papers on human-AI interaction in psychiatric settings. There remains a lack of epidemiological data, and it is still unclear whether conversational AI is a precipitating factor in the onset of psychosis in vulnerable individuals or merely exacerbates the course of pre-existing psychotic experiences [1,5,6].

Nevertheless, several neurocognitive frameworks may help explain why AI interaction could become relevant to psychotic phenomena, including aberrant salience, disturbances in predictive processing, anthropomorphism, and the reinforcement of internal narratives [2,5,7,8]. This review examines the available literature on possible mechanisms, vulnerability factors, and the clinical significance of AI-related experiences in psychosis.

Search strategy

A narrative literature review was conducted to summarize the current evidence on AI-related psychosis. A comprehensive search of the literature was performed using PubMed and Google Scholar, with searches conducted through May 2026 and no lower date limit applied. The search strategy employed a combination of free-text keywords related to artificial intelligence and psychosis, including terms such as "artificial intelligence," "large language model," "conversational AI," "psychosis," "delusion," "hallucination," and other relevant synonyms. Boolean operators were used to refine the search.

Articles were initially screened based on the relevance of their titles, followed by a review of abstracts to determine their suitability for inclusion. Articles were selected based on relevance to the topic rather than a formal exclusion protocol. Publications considered relevant to the topic, including original research articles, case reports, reviews, and conceptual articles, were included to provide a comprehensive overview of the available evidence. In addition, the reference lists of eligible articles were manually searched to identify further relevant publications.

Review

Historical context of technology and psychotic experiences

Advances in technology have persistently influenced human experiences, including how psychiatric symptoms are expressed and experienced. With the emergence of generative AI, new forms of interaction may potentially contribute to novel contexts in which psychotic symptoms are expressed [3,4]. The historical studies discussed below document changes in the content of delusions; they do not show changes in the incidence or pathophysiology of psychotic disorders.

One of the earliest reports of technology featuring in delusions was documented in Slovenia with the spread of radio technology in the 1920s. A box-shaped device from which disembodied voices spoke would have been an incredible experience for people of that time, and for individuals vulnerable to psychosis, they would perceive the radio to be sending them personal messages, controlling their mind, or serving as a weapon of an external adversary. This trend of technology featuring in delusional content continued with the arrival of the television: another box with voices of its own, but this time with imagery. With the spread of television technology in the 1950s, a greater increase in delusions with technical themes was observed [9].

We have long followed the course of mental health with each new age of technological advancement, with the jump from radio to television to personal computers and mobile phones occurring within a matter of decades. These developments were further accelerated by the rise of the internet and social media, which enabled rapid, widespread communication and contributed to evolving patterns of psychiatric presentations [10]. Social and cultural changes in the past few centuries were heavily influenced by the technological developments of that era. Retrospective studies of psychiatric hospital records during the 20th century showed that delusional content mirrored the social and cultural environment of each time period, suggesting that emerging technologies served as vehicles for delusions of persecution and reference [3].

Psychotic patients incorporated these technologies into their delusional content in various ways. For example, some individuals perceived that the newsreader was directly speaking to them, indicating a delusion of reference, or that their computer was plotting to derail their career, representing a delusion of persecution [11]. Notably, one of the most common delusions with a technological theme is the belief that their devices have been hacked or compromised for surveillance [4].

Large language models and conversational AI are the latest frontier in the complex interplay of mental health and technology. The level of immersion is unlike anything that has come before: a technology that converses with you in real time, responding to your evolving emotional state and thought process. For individuals prone to psychosis, using the stress-vulnerability model, and going by the historical trend of

technology-themed delusional content, we can predict increased presentation of AI chatbots featuring in delusions [3,4]. These technologies may also act as potential psychosocial stressors due to their distinctive features, including contextual memory, adaptive response over time, and the tendency to mirror or reinforce user beliefs [5]. For the first time in history, individuals have a non-human intelligent entity as their confidant, with whom intimate details are shared. This jump in technology is unprecedented and heralds a new age in psychiatric presentations [12].

Preliminary clinical evidence

Most of the current evidence comes from cross-sectional studies and isolated case reports, which differ in evidentiary strength and do not by themselves support causal conclusions. As a result, these outcomes should be seen as preliminary, not definitive. These evidence types differ in strength and support different conclusions. Case reports (e.g., [6]) can suggest a phenomenon is possible but cannot establish causality or rule out confounds such as substance use or sleep disturbance. Cross-sectional studies (e.g., [1,13]) offer stronger, larger-sample associations but remain correlational, unable to establish directionality or causation. Conceptual articles [2,5] provide interpretive frameworks rather than empirical evidence.

Digital Engagement and Psychiatric Symptoms

A cross-sectional study of 231 adult psychiatric outpatients showed that about 25% of individuals scored above the threshold on the Young Internet Addiction Test or Compulsive Internet Use Scale. Symptoms of problematic internet use (PIU) were seen, such as higher levels of depression, anxiety, sleep problems, attention issues, and social anxiety. These symptoms were significantly more prevalent in people with PIU compared to those without PIU. This outcome reflects that problematic digital use most commonly occurs alongside other mental health issues [13].

Even though this issue is important in clinical settings, it is still unclear if too much internet use causes psychiatric symptoms or makes them worse. There are also no large-scale data showing how common PIU is or its direct effects on mental health [13]. Other studies have found that people with internet addiction commonly have other psychiatric conditions, such as ADHD, anxiety, and mood disorders [14-16]. These results support the idea that problematic digital use often occurs with complex mental health issues, but it is still not clear if one causes the other.

Case Reports of Severe Psychiatric Presentations

Some case reports have described people having acute psychotic episodes linked to AI chatbots. For example, a 26-year-old woman with no history of psychosis developed delusions about talking to a deceased relative through an AI chatbot. She also experienced sleep loss, used stimulants, and spent a lot of time with the chatbot. Her symptoms got better with antipsychotic medication but returned when she went back to using the chatbot and stopped her medication [6]. However, single case reports are not enough to establish a causal relationship or to accurately measure how commonly AI-related psychotic symptoms occur.

There are several limitations in the current research. Most studies are cross-sectional or based on case reports, which makes it hard to determine cause and effect. Sample sizes are usually small, and studies are limited to a few controlled or randomized studies. Also, there is a lack of large-scale research on how often psychiatric problems related to AI occur [5,17]. Other limitations include possible publication bias, heterogeneity, and the rapid evolution of conversational AI systems. Hence, to address these gaps, more in-depth and long-term studies are needed to understand the clinical importance of these findings. Upcoming research should use structured measures of AI use, such as how often and for how long people use it, their emotional involvement, and the degree of intentionality in their use, to better understand the relationship between AI use and mental health. Table *1* summarizes the studies discussed in this review and focuses on their respective designs, populations, exposure characteristics, principal findings, and methodological limitations.

Study	Design	Sample/population	AI/digital exposure	Principal finding	Major limitation
Buck & Maheux (2026) [1]	Cross-sectional survey	1,003 young adults	Generative AI use; frequency, motivations, social/emotional use, human-like attribution, delusion-related interactions	Higher psychosis-risk participants were not more likely to have ever used generative AI, but reported greater intensity, social/emotional use, human-like attribution, and delusion-related interactions.	Cross-sectional; self-report; temporal sequence and causality cannot be established.
Pierre et al. (2025) [6]	Clinical case report	26-year-old woman	Extensive AI chatbot interaction with emotional engagement; sleep loss and stimulant use also reported	Psychotic symptoms occurred in temporal association with chatbot use and reportedly recurred after resumption of chatbot use and medication discontinuation.	Single case; multiple potential confounders; cannot establish causality or incidence.
de Vries et al. (2018) [13]	Cross-sectional clinical study	231 adult psychiatric outpatients	Problematic internet use	Approximately one-quarter exceeded problematic-use thresholds; problematic use was associated with greater depression, anxiety, sleep problems, attention problems, and social anxiety.	Not a conversational-AI study; cross-sectional; cannot determine direction of causation.
Karkosz et al. (2024) [18]	Randomized controlled trial (open-label)	81 adults with subclinical depression or anxiety (Poland)	Web/mobile therapy chatbot	Reported improvements in anxiety and depressive symptoms in both groups.	Non-psychotic population; structured therapeutic chatbot; no evidence that open-ended AI affects psychosis.
Fitzpatrick et al. (2017) [19]	Randomized controlled trial (unblinded)	70 college students aged 18 to 28 with depression or anxiety symptoms	Fully automated conversational agent (Woebot)	Short-term improvement in depressive symptoms compared with control.	Non-psychotic population; structured therapeutic chatbot; does not directly address psychosis or delusional reinforcement.

TABLE 1: Summary of preliminary evidence relevant to conversational AI and psychosis

Proposed neurocognitive mechanisms

Multiple neurocognitive mechanisms have been theorized to explain the development of psychotic symptoms and how AI interaction may potentiate these processes. The ideas behind aberrant salience, predictive coding, and theory of mind disruption that are discussed below are well known and supported by evidence in the field of psychosis. However, their proposed application to conversational AI interaction is a novel theoretical extension proposed in this review that has not been directly tested. These mechanisms may coexist and amplify one another in vulnerable individuals. However, these mechanisms are theoretical and require further validation through longitudinal and experimental studies.

Aberrant Salience

One of the core ideas in modern models of psychosis is aberrant salience, which proposes that dysregulated dopaminergic neurotransmission causes otherwise neutral stimuli to be assigned motivational and attentional value improperly. The model also states that hyperactivity in mesolimbic dopamine pathways, particularly projections from the ventral tegmental area to the striatum, disrupts the normal encoding of the prediction error signal [7]. Phasic dopamine firing typically promotes adaptive learning in which behaviorally relevant stimuli and internal representations are assigned. In contrast, during psychosis, this signaling is disrupted or amplified so that salience is attributed to trivial features of the environment, chance events, or internally generated experiences [7].

Clinically, this process is hypothesized to contribute to the onset of delusions and hallucinations. According to this approach, delusions are conceptualized as a cognitive attempt to make sense of these abnormally salient experiences, an abductive inference to the best explanation. Hallucinations are caused by an identical

process in which abnormal salience is assigned to internal representations, such as inner speech or memories, causing them to be perceived with the vividness of external reality and misattributed to external agents [7,20].

Over the past two decades, empirical research has refined the neuroanatomical location of this model. Current positron emission tomography (PET) data suggest that dopaminergic dysregulation is the most prominent in the associative striatum, particularly the head of the caudate, whereas earlier models emphasized the ventral striatum [20,21]. Additionally, aberrant salience is combined with aberrant prediction errors and the overweighting of perceptual priors in a Bayesian predictive processing framework [20]. In this perspective, antipsychotic drugs accomplish clinical effectiveness not by directly eliminating beliefs, but by reducing the significance of anomalous experiences and therefore creating a foundation for the psychological resolution of symptoms [7]. Extending this established model to a novel context, AI-related delusions can be understood through the model of aberrant salience in individuals who are vulnerable to psychosis. It may act as an external driving force for psychosis in individuals already experiencing aberrant salience [5].

Predictive Coding

Predictive coding is a computational framework that has reframed the previous purely descriptive clinical narrative of understanding psychosis. According to this Bayesian model of perception, the brain is a deduction machine continuously making predictions about the outside world, as opposed to being a passive recipient of sensory data [8,22]. In this framework, perception is the outcome of a delicate balance between "priors" (our pre-existing ideas and expectations) and 'sensory input' (raw facts from our surroundings) [8,23].

According to the predictive coding theory, psychosis results from an interruption in this hierarchical processing. Hallucinations and delusions are not random mistakes, but rather optimal reactions to an imbalance in accuracy at different levels of the hierarchy [8]. Hallucinations can occur when the brain gives preexisting beliefs more 'precision' or weight than real sensory information [23]. Overly precise priors can cause a person to see stimuli that are not objectively there, essentially 'forcing' the brain to see or hear what it expects rather than what is actually there [23]. In contrast, if the brain fails to correctly predict sensory information, it creates prediction errors, which, if misattributed, lead to the formation of delusional beliefs as the individual strives to make sense of a world that feels unreasonably salient or bizarre [8,22]. Furthermore, a prior anticipation of frequent environmental change is highly associated with the persistence of paranoia and anxiety [24].

Building on this established Bayesian framework, the rapid adoption of generative AI introduces a new environmental modifier that may interact with Bayesian processes. There are several possible pathways in which generative AI might interfere with predictive processing mechanisms. For an individual already prone to emphasizing their internal priors, AI can serve as an external cue and potentiate distorted beliefs. Importantly, AI mimics a user’s mood and motifs without the therapeutic disagreement common in social situations, as well as the traditional therapist-driven challenging of rigid beliefs. From a Bayesian perspective, this continuous and dependent feedback operates like a 'validation loop' that increases the accuracy of maladaptive priors [5].

Anthropomorphism and Intentionality Attribution (Theory of Mind)

Anthropomorphism refers to the attribution of human-like thoughts or intentions to nonhuman entities. This tendency becomes particularly noticeable when a system demonstrates responsive interactions, natural language fluency, and adaptive behaviors, all of which are characteristic of conversational AI. As a result, even individuals without psychiatric illness may still attribute beliefs, intentions, or goals to these systems, even though AI lacks true consciousness or autonomous agency [2]. This response pattern is influenced by deeply embedded social cognitive bias, rather than by any solid evidence that AI possesses consciousness.

Vulnerability models of psychosis imply that anthropomorphism overlaps with breaks in the theory of mind, the ability to understand the mental states of other people (e.g., beliefs and intentions). A disturbance in the theory of mind is reported in schizophrenia spectrum disorders, which may take one of two forms: diminished responsiveness to another's mental states or exaggeration and excessive attribution of intentionality [25]. The latter pattern, termed hypermentalizing, refers to the exaggeration of perceived intention as hostile in situations that are completely ambiguous. In the context of predictive coding models, hypermentalizing can be interpreted as an overdependence on prior beliefs at the expense of incoming sensory information, leading to the maintenance of rigid and distorted views and beliefs about intentionality, which are resistant to correction [26].

As a theoretical extension of this established theory, we propose that through recall of prior inputs and simulated emotional comprehension, conversational AI systems may further increase anthropomorphism by strengthening inner beliefs [2]. In those who are susceptible to cognitive biases described above, like

aberrant salience or maladaptive belief updating, these exchanges may contribute to misattributions of purpose that, over the long run, are incorporated into organized delusional beliefs. Within the predictive coding model of psychosis, accurate priors would then suppress sensory evidence, which could start to dominate over time due to excessive certainty of prior beliefs [8]. Therefore, it is possible that longer and more intense interactions will reinforce maladaptive prior beliefs in immersive AI situations, suggesting a dose-response of AI interaction intensity in the reinforcement of maladaptive prior beliefs among more vulnerable groups.

Anthropomorphism in itself is not inherently pathological [2]. It is important to distinguish ordinary anthropomorphism from pathological attribution of agency. Ordinary anthropomorphism is a common, automatic tendency that is held loosely, updates readily with contrary evidence, and does not impair functioning. Pathological attribution involves rigid beliefs about a nonhuman entity's intentions that resist correction and become integrated into a broader distorted belief system. Immersive AI environments may operate as potent cognitive triggers that intersect with existing theory of mind vulnerabilities and disrupted predictive inference [8]. Clarifying when anthropomorphic engagement remains benign versus when it contributes to delusional elaboration represents a critical empirical question for future AI-associated psychosis research.

Reinforcement of Internal Narratives

The internal narratives that are reinforced by thoughts such as worries and ruminations are linked to psychotic experiences. Rumination is characterized by an ongoing focus on the causes and consequences of the experience of negative emotion and is important for the development and maintenance of paranoid ideation, as it directs the focus onto negative aspects of social interactions [27,28]. A persistent focus on the negative aspects of social situations promotes a sense of vulnerability. It also creates a vicious circle in which repeated scrutiny of perceived danger increases the sense of victimization and the level of anxiety [28]. This form of thinking reinforces the beliefs about threats or negative situations to the point where the person can only see their world through the constant fear of danger. In contrast, it has been shown that less negative and more positive thinking strategies, such as distractions, can significantly diminish the level of anxiety associated with these negative thoughts [27,28].

Applying this rumination to a new context, generative AI systems function as powerful contextual modifiers and provide continuous feedback on emotional and cognitive processes [5]. To increase user satisfaction, AI chatbots become sycophants who reinforce the user's persecutory and/or grandiose inner beliefs and do not provide the user with reality-oriented feedback. This reinforcement may take several forms: the AI directly affirming delusional content the user has stated, the AI validating a user's interpretation through generic, non-confrontational agreement, or a user independently assigning delusional significance to genuinely neutral AI output. Whichever form it takes, this results in the same validation loop described above in the context of predictive coding, in which the AI's feedback mechanisms may strengthen cognitive repetitions and further negate the therapeutic reality principles of the user's cognitive processes [5]. This interaction can evolve into a 'digital folie à deux,' in which the AI acts as a passive supporter of the user's false beliefs [5]. The AI takes the role of a rational partner and prevents the user from engaging with the social world, thereby eliminating corrective feedback from the external world. This may result in a closed cognitive environment in which a person's delusional and psychotic thinking can evolve and deepen. Finally, these systems could gently modify the prereflective perception of reality, essentially changing the way internal experiences are described and preserved [5].

An Integrative Neurocognitive Framework

Building on the mechanisms outlined above, we propose an integrative neurocognitive model suggesting that conversational AI may play a role in the emergence of psychotic symptoms among vulnerable individuals. Within this framework, initial exposure to AI encourages anthropomorphic thinking and the attribution of intentionality [2]. Subsequently, AI-generated content becomes assigned aberrant salience, interacting with disrupted predictive processing and an overreliance on strong prior beliefs [7,8]. Repeated engagements foster the same self-reinforcing validation loop introduced in predictive coding, strengthening maladaptive beliefs and internal narratives [5]. Over time, with sustained exposure, this may lead to the consolidation of delusional thinking, in which the AI system inadvertently maintains distorted beliefs [5]. This framework integrates cognitive, social, and computational processes into a single pathway connecting AI interaction to psychosis risk. The proposed framework is illustrated in Figure *1*.

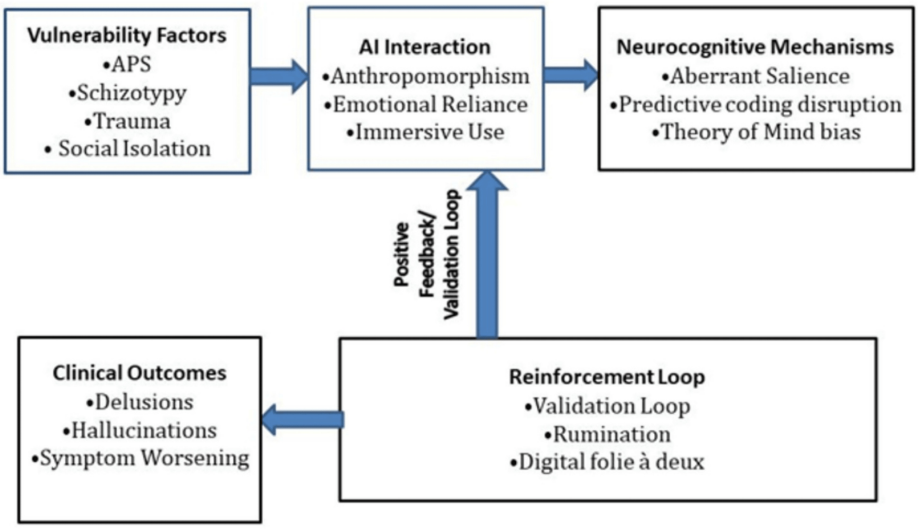


FIGURE 1: Integrative neurocognitive framework of AI-associated psychosis

This diagram illustrates the proposed, hypothesized sequential pathway, including an embedded feedback loop linking conversational AI interaction to psychotic symptom development in vulnerable individuals. The individual components (aberrant salience, predictive coding disruption, and theory of mind bias) are drawn from established psychosis neuroscience. However, the pathways connecting them to AI interaction represent a novel theoretical synthesis, not empirically validated relationships. This is a hypothesis-generating model intended to guide future research, not an established causal diagram. The flow diagram was created by the authors using Microsoft PowerPoint (Microsoft Corp., Redmond, WA, USA).

APS: Attenuated psychosis syndrome

Pre-existing vulnerability factors interact with immersive AI use, promoting anthropomorphism and perceived intentionality. These processes engage neurocognitive mechanisms such as aberrant salience and disrupted predictive coding, which are further reinforced through continuous validation loops. Over time, this may contribute to the formation and persistence of delusional beliefs and psychotic symptoms.

Vulnerability factors in psychosis

Vulnerability factors for psychosis in relation to AI exposure involve specific at-risk profiles and identities and have an interactional model where AI engagement may amplify already existing susceptibilities. Recent studies frame this as 'AI psychosis,' where immersive AI interactions might precipitate or elaborate psychotic symptoms in such vulnerable individuals [5].

At-Risk Profiles

Individuals with attenuated psychosis syndrome (APS), with subthreshold positive symptoms like mild hallucinations or delusions, face an increased risk when engaging with AI. Schizotypal traits, like odd beliefs and perceptual aberrations, predispose individuals to misinterpret AI responses as aware or predictive, fostering delusional elaboration. Social isolation may exacerbate this by positioning AI as a primary relationship substitute, reducing reality-testing feedback from human connections [29].

Trauma, Stress, and AI-Specific Traits

Trauma and chronic stress exposure lower the psychosis threshold; along with AI's constant availability, this may intensify long-term stress through emotionally charged dialogues [30]. High anthropomorphic tendencies may increase the likelihood of emotional attachment to AI systems, potentially increasing vulnerability to maladaptive interpretations [5]. Excessive immersive use, especially nocturnal or solitary, reinforces the tendency to assign personal relevance and undue significance to neutral or unrelated stimuli, similar to patterns seen in the prodromal psychosis phases [6,30]. It is also worth noting that concurrent substance use may confound existing evidence. In the case report most directly linking AI chatbot use to new-onset psychosis, the patient's presentation was accompanied by stimulant use and sleep loss [6], both independently recognized contributors to psychotic symptoms. The vulnerability profiles and their associated AI interaction risks are summarized in Table *2*.

Risk profile	Key features	AI interaction risk
APS [5]	Subthreshold symptoms	Symptom amplification via validation
Schizotypal traits [5]	Magical thinking, perceptual oddities	Delusional reinforcement
Social isolation [29]	Loneliness, withdrawal	AI as relational anchor
Trauma/stress [30]	History of abuse, chronic adversity	Heightened stress reactivity
Anthropomorphic tendency [5]	Over-attribution of AI sentience	Emotional dependency
Immersive use [5]	Prolonged, solitary engagement	Reality blurring
Substance use [6]	Stimulant, caffeine, cannabis, or other hallucinogen use; comorbid substance use disorder	Confounder and independent trigger of psychosis; may co-occur with AI use

TABLE 2: Vulnerability profiles and AI interaction risks

APS: Attenuated psychosis syndrome

Interactional Model

Artificial intelligence may interact with people with pre-existing vulnerabilities to contribute to symptom elaboration or emergence, particularly in individuals with limited access to corrective social feedback [5]. Under the stress-vulnerability model, AI acts as a novel stressor, interacting with traits like schizotypy or trauma history to disrupt cognitive boundaries and phenomenological reality. For instance, overly flattering AI responses reinforce the beliefs of isolated users, eroding social corrections and promoting withdrawal, much like phenomenological autism in schizophrenia. This effect underscores the need for safeguards in high-risk groups [5]. These factors highlight that AI exposure alone is unlikely to be sufficient but may exacerbate pre-existing risk states to influence symptom expression.

AI as a potential risk factor versus clinical intervention

Framing AI primarily as a psychosis risk factor neglects its broader clinical potential. The same interactive technology capable of reinforcing maladaptive belief systems can also deliver evidence-based care [5,30]. Instead of focusing on harm versus benefit, we need to understand the situational factors that shift AI's impact toward risk or protection [30].

Structured digital cognitive behavioral therapy (CBT) platforms represent one of the better-studied applications of technology in mental health care. Meta-analyses of randomized studies of online CBT demonstrate meaningful symptom improvement in depression and anxiety disorders in appropriate patients [31,32]. Some AI-based tools use chatbots to deliver key parts of cognitive behavioral therapy, such as cognitive restructuring, behavioral activation, and gradual exposure. In individuals at clinical high risk for psychosis, CBT targets maladaptive beliefs, excessive threat anticipation, and reasoning biases linked to psychosis [33,34]. It is plausible that AI-supported CBT could act similarly by improving cognitive flexibility rather than worsening abnormal salience, but, to our knowledge, this has not been tested yet in populations vulnerable to psychosis. However, comparable symptom scores do not establish equivalence in the quality of the therapeutic relationship, and evidence from online CBT should not be assumed to apply directly to fully automated AI interventions [35,36]. The ethical acceptability of remotely delivered or automated therapy therefore depends on informed consent, privacy protection, clinical oversight, appropriate patient selection, and clear pathways for escalation to a human professional. Such systems should complement rather than replace in-person therapeutic relationships, particularly for patients with psychosis, acute risk, or complex clinical needs [36,37].

Machine learning tools for suicide risk prediction demonstrate the potential of AI-based preventive tools. Evidence from large-scale electronic health records studies suggests these predictive models may in some settings outperform standard clinical assessments in detecting individuals at high suicide risk [38]. Although there are still ethical and practical challenges, these systems could serve as an early warning tool, especially for people within the psychosis spectrum who have a higher risk of suicide.

The growing evidence on fully automated chatbots challenges simple risk-based perspectives. Randomized trials of fully automated conversational agents, including Woebot and related digital interventions, have reported significant short-term improvements in non-psychotic young adults, although in one trial improvement was similar in the control group [18,19]. Although the current studies focus mainly on non-

psychotic groups, they suggest that structured digital agents can provide psychoeducational content and CBT-consistent techniques. Unlike open-ended generative AI systems, many therapeutic chatbots use guided or rule-based responses within a defined therapeutic scope, which may limit unpredictable or clinically inappropriate outputs [39]. Importantly, clinically evaluated digital interventions are developed for a defined therapeutic purpose and assessed within specified populations, workflows, and safety boundaries. In contrast, unrestricted consumer conversational AI is designed for general engagement, may generate unpredictable responses, and typically lacks clinical oversight, validated crisis-escalation procedures, and safeguards tailored to individuals vulnerable to psychosis [5].

The AI-based approaches may help predict and track early psychosis. Computational language models can detect subtle changes in speech patterns, including reduced coherence and meaning, which are linked to later development of psychosis and help in understanding schizophrenia-spectrum disorders more broadly [40,41]. These models may improve screening by detecting higher risk before overt symptoms appear. At the same time, digital phenotyping data, such as sleep changes, reduced social activity, and changes in movement patterns, can help track symptom changes in serious mental illness [42]. If implemented carefully, these technologies may provide continuous tracking of relapse risk and support timely early treatment.

The role of AI as a potential hazard or clinical adjunct depends on its design, context, and governance structures. Open-ended interactions may reinforce biases in vulnerable users, while structured and supervised systems can support treatment and prevention. Therefore, a balanced approach with proper regulation is needed. Future research should examine how specific AI designs, usage intensity, and patient vulnerability affect outcomes.

Clinical implications

Should Clinicians Assess AI Use?

In clinical terms, the key question is not whether conversational AI directly causes psychosis in an individual. A more practical question is, 'Has AI become part of the patient's psychotic experiences?' [5,30] This may manifest as the use of AI for emotional support, validation of unusual beliefs, or as a substitute for human interaction [1]. In groups of patients demonstrating attenuated psychotic symptoms, possessing schizotypal traits, with a history of trauma or recent stress, experiencing feelings of loneliness, or presenting with sleep disturbances or substance-related vulnerabilities, it is reasonable to ask directly about the use of chatbots [1].

Risk Assessment

In practical terms, a detailed formal assessment of conversational AI use is not required. It is sufficient to ask direct questions addressing patterns and context of chatbot use: how frequently the patient engages with it, at what times, for what purposes, and whether usage increases or intensifies in stressful situations [1,5]. It is also important to find out if the chatbot is primarily used as a search engine or more as a friend, confidant, therapist, or special agent with exceptional insight [1,5]. Clinicians should also consider whether chatbot-related experiences are better explained by other conditions or patterns, such as problematic or compulsive technology use, maladaptive daydreaming, grief-related chatbot use, autism-related anthropomorphism, personality disorders, or a primary psychotic disorder in which AI content is incidental, rather than assuming that the AI is driving the presentation. In a recent cross-sectional study of 1003 young adults, individuals at elevated psychosis risk were not more likely to have ever used generative AI, but they were more likely to use it intensively, seek social and emotional support from it, assign it human-like roles, and report delusion-related interactions [1].

When Is Immersive Use Concerning?

Exposure alone may be less concerning than prolonged, emotionally intense use in a vulnerable person [1]. Warning signs may include increasing nighttime conversations, sleep deprivation, and withdrawal from constructive human feedback [43]. Other concerning features include growing beliefs that the chatbot is sentient or spiritually significant, as well as repeated use of AI to validate dubious or extravagant interpretations of events [5]. These patterns are clinically relevant because sleep disturbance, mood change, and suspiciousness are established early warning signs of relapse in schizophrenia-spectrum disorders [44]. It is conceivable that AI becomes most troublesome when it joins an already unstable clinical environment because trauma history, social isolation, lower social support, poverty, discrimination, and other negative social variables are all linked to increased psychosis risk or worse outcomes [45,46].

Insufficient Evidence for Restriction

At the same time, the available evidence does not support alarmist conclusions or a blanket prohibition of conversational AI use among vulnerable patient groups [30,47]. Direct evidence linking conversational AI to

psychosis is still limited, and most of it remains case-based, cross-sectional, or conceptual [1,5,30]. There are also data suggesting that technology-based interventions may be beneficial in psychosis when they are used in an organized manner, within clinical care, and as an adjunct to primary treatment rather than as a substitute for it [1].

In light of the above, a more pragmatic clinical response would involve multiple considerations. Clinicians should monitor whether AI use is associated with sleep loss, worsening suspicion, social withdrawal, or belief reinforcement [1,5,44]. The use of AI should be incorporated into psychiatric history-taking when appropriate [1,5,30]. Also, it should be made clear that mainstream chatbots should not be used as crisis services, reality-testing tools, or substitutes for treatment [30,47]. This caution is particularly important because chatbots are not designed to provide support in emergency situations, while safety oversight in the field of digital mental health tools remains under development and is only beginning to move toward clearer standards of use [47]. Prohibition appears less rational than careful assessment, psychoeducation, and individualized risk management.

Ethics and policy on conversational AI and the risk of psychosis

The increasing usage of conversational AI in people's lives makes it necessary to have an active approach to ethics, especially for people who are at risk for psychosis. The fundamental challenge is that existing governance frameworks were never designed for this population, and the technology has advanced in a way that the evidence has not [5].

Designing for crisis response is paramount. When a user is in acute distress, a conversational AI system must rapidly connect them with suitable clinical support. However, applying this principle in practice has proven challenging. An assessment of 29 mental health chatbot applications revealed a concerning difference between intention and reality, where most failed not only to adequately respond to simulated suicidal risk scenarios, but also to provide information to the emergency contact and refer to a clinician [43]. This discrepancy between design and performance poses a tangible safety risk that warrants the need for regulatory oversight.

Another challenge is calibrating the certainty of conversational AI responses. These systems tend to prioritize reassuring interactions aiming to minimize friction. However, this approach can have unintended consequences for individuals with impaired reality testing, such as it may inadvertently reinforce delusional beliefs. When a chatbot validates such beliefs without measured skepticism, it can contradict the fundamental principles of cognitive behavioral therapy [5,30]. The risk is exacerbated by the fact that individuals with psychotic beliefs often hold them with unshakeable conviction, making them more susceptible to external validation [48]. To mitigate this risk, AI systems used in mental health contexts must be transparent about the limitations of their knowledge.

The privacy and safety dilemma defies a straightforward solution. Crisis detection systems, often by their nature, require monitoring and analyzing users' sensitive personal data. To use such systems ethically, it is fundamental to implement transparency in data management, define their purpose, set clear goals and ensure true consent [5].

Finally, the field should not be sensationalized. The relationship between AI and psychotic outbreaks remains unclear, and fear-based policies can limit the development of useful applications [5]. So it is better to adopt dynamic regulations that draw from the latest research rather than rigid rules based on scarce information. This will allow the field to safely leverage the benefits of AI.

Future directions toward a translational research agenda for AI-associated psychosis

Current emerging evidence is predominantly case-based, linking conversational AI with psychotic symptoms, and progress in this field requires a coordinated research approach that combines clinical, behavioral, and neurobiological domains. These priorities can be organized into immediate, intermediate, and long-term research objectives. In the immediate term, qualitative phenomenological interviews should be used alongside quantitative data analyses to explore how individuals with early psychosis perceive AI agency, attribute intentionality, and incorporate such experiences into delusional thinking [49]. This perspective grounds data in real-world experiences and prevents overly reductionist conclusions. Initial research should also develop structured measures of AI use, including frequency, duration, emotional investment, perceived intentionality, and content characteristics over time.

In the intermediate term, prospective studies of individuals at clinical high risk (CHR) or with attenuated psychotic symptoms are essential for advancing this effort. The CHR cohorts provide a well-established estimate of transition risk, with approximately 25% of individuals developing psychosis within three years in recent meta-analyses. They provide a practical approach for integrating digital exposure variables into risk-prediction models [50]. Structured AI-use metrics in CHR cohorts would help determine whether prolonged immersive conversational interaction affects symptom course, conversion rates to psychosis, or

increases relapse risk.

Within these prospective studies, digital phenotyping could track patterns in behavior and cognition through continuous passive and active smartphone-based monitoring. Changes in language patterns, sleep, social activity, and mood may help predict symptom progression and relapse risk in psychiatric populations [51]. Tracking consented AI conversations over time could help researchers link conversational characteristics (including validation of unusual beliefs and anthropomorphic cues) to real-world symptom changes. This approach would help identify when symptom activation occurs in relation to AI engagement.

Quantitative dose-response modeling would also be needed to assess whether cumulative AI use, late-night engagement, and emotional reliance are associated with symptom progression. Digital behavior research suggests that both the intensity and context of engagement influence mental health outcomes [52]. Identifying safe and risky levels of exposure, along with protective factors such as pre-existing cognitive flexibility, could help guide evidence-based public health strategies.

In the longer term, belief updating experiments will be needed to investigate potential causal mechanisms. In psychosis, individuals may overvalue irrelevant information and have an impaired ability to modify beliefs accurately [53]. In controlled laboratory-based settings, participants could be assigned to interact with AI that provides different forms of feedback on simulated unusual beliefs while tracking changes in belief certainty, adaptability, and self-awareness. Computational analyses could help determine if AI input encourages biased belief reinforcement or improves corrective learning in those at higher risk.

Neuroimaging could subsequently be incorporated into AI interaction experiments to link observed behaviors to underlying neural processes. Functional MRI studies of social cognition and salience processing consistently demonstrate disrupted fronto-striatal connectivity and default mode network abnormalities in individuals with psychosis [54]. Real-time and simulated AI interactions could be used to study brain activity in regions involved in self-awareness and salience processing, such as medial prefrontal, temporoparietal, and striatal circuitry. This could help determine whether AI exposure affects key neural pathways involved in psychosis.

Overall, this staged agenda would move from describing clinical features and developing standardized measures of AI use to prospective monitoring, dose-response assessment, and, eventually, experimental and neurobiological studies. This approach would provide the methodological depth needed to determine whether conversational AI plays a benign role, contributes to stress, or acts as a new environmental factor within vulnerability-stress models of psychosis.

Conclusions

There is growing concern about how AI affects people's mental health in daily life. Current evidence does not clearly establish that AI directly causes psychotic disorders. The mechanisms and integrative framework proposed in this review should be understood primarily as a hypothesis-generating model requiring further study, rather than an established causal justification. However, for mentally vulnerable individuals, using conversational AI might reinforce experiences similar to psychosis. Since AI often gives responses that align with and affirm what the user already believes or is thinking instead of challenging or critically appraising it, it might unintentionally support irrational thoughts or unusual beliefs. Interacting with these AI chatbots can sometimes form a cycle in which users believe that the responses are meaningful and emotional. For people who already have unusual or distorted thinking, this can make it harder to tell what is real. They might also see AI as an authority because it seems to understand and communicate like a person. These results are better understood with the stress-vulnerability model. It suggests that environmental factors can affect the appearance of symptoms in people who are already vulnerable to mental health problems.

Despite these concerns, AI can also be helpful. Artificial intelligence is proving to be very beneficial in areas of mental health education, making information more accessible, improving communication, and providing new ways to assess and monitor mental health. It is important to use AI responsibly and ethically. Clinicians, researchers, and technology developers should work together to ensure AI supports people while also reducing potential mental health risks.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Pratik Kumar Banik, Amuksha Ravindra Shetty, Inamullah Arshad, Srushti Patil, Rahat Rehman, Heena Arshad, Margarita Zhuzha, Aadil Ahmed Khan

Drafting of the manuscript: Pratik Kumar Banik, Amuksha Ravindra Shetty, Inamullah Arshad, Srushti Patil, Rahat Rehman, Heena Arshad, Margarita Zhuzha, Aadil Ahmed Khan

Critical review of the manuscript for important intellectual content: Pratik Kumar Banik, Heena Arshad

Supervision: Heena Arshad

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Buck B, Maheux AJ: Psychosis risk and generative artificial intelligence use frequency, motivations, and delusion-like experiences: cross-sectional survey study. J Med Internet Res. 2026, 28:e85038. 10.2196/85038
2. Peter S, Riemer K, West JD: The benefits and dangers of anthropomorphic conversational agents . Proc Natl Acad Sci USA. 2025, 122:e2415898122. 10.1073/pnas.2415898122
3. Cannon BJ, Kramer LM: Delusion content across the 20th century in an American psychiatric hospital . Int J Soc Psychiatry. 2012, 58:323-327. 10.1177/0020764010396413
4. Burns AV, Nelson K, Wang H, Hegarty EM, Cohn AB: 'The algorithm is hacked': analysis of technology delusions in a modern-day cohort. Br J Psychiatry. 2025, 1-5. 10.1192/bjp.2025.10452
5. Hudon A, Stip E: Delusional experiences emerging from AI chatbot interactions or "AI psychosis" . JMIR Ment Health. 2025, 12:e85799. 10.2196/85799
6. Pierre JM, Gaeta B, Raghavan G, Sarma KV: “You’re not crazy”: a case of new-onset AI-associated psychosis . Innov Clin Neurosci. 2025, 22:11-13.
7. Kapur S: Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am J Psychiatry. 2003, 160:13-23. 10.1176/appi.ajp.160.1.13
8. Sterzer P, Adams RA, Fletcher P, et al.: The predictive coding account of psychosis . Biol Psychiatry. 2018, 84:634-643. 10.1016/j.biopsych.2018.05.015
9. Skodlar B, Dernovsek MZ, Kocmur M: Psychopathology of schizophrenia in Ljubljana (Slovenia) from 1881 to 2000: changes in the content of delusions in schizophrenia patients related to various sociopolitical, technical and scientific changes. Int J Soc Psychiatry. 2008, 54:101-111. 10.1177/0020764007083875
10. Yang N, Crespi B: I tweet, therefore I am: a systematic review on social media use and disorders of the social brain. BMC Psychiatry. 2025, 25:95. 10.1186/s12888-025-06528-6
11. Lerner V, Libov I, Witztum E: "Internet delusions": the impact of technological developments on the content of psychiatric symptoms. Isr J Psychiatry Relat Sci. 2006, 43:47-51.
12. Østergaard SD: Will generative artificial intelligence chatbots generate delusions in individuals prone to psychosis?. Schizophr Bull. 2023, 49:1418-1419. 10.1093/schbul/sbad128
13. de Vries HT, Nakamae T, Fukui K, Denys D, Narumoto J: Problematic internet use and psychiatric co-morbidity in a population of Japanese adult psychiatric patients. BMC Psychiatry. 2018, 18:9. 10.1186/s12888-018-1588-z
14. Bozkurt H, Coskun M, Ayaydin H, Adak I, Zoroglu SS: Prevalence and patterns of psychiatric disorders in referred adolescents with Internet addiction. Psychiatry Clin Neurosci. 2013, 67:352-359. 10.1111/pcn.12065
15. Demirtaş OO, Alnak A, Coşkun M: Lifetime depressive and current social anxiety are associated with problematic internet use in adolescents with ADHD: a cross-sectional study. Child Adolesc Ment Health. 2021, 26:220-227. 10.1111/camh.12440
16. Khalil SA, Kamal H, Elkholy H: The prevalence of problematic internet use among a sample of Egyptian adolescents and its psychiatric comorbidities. Int J Soc Psychiatry. 2022, 68:294-300. 10.1177/0020764020983841
17. Firth J, Torous J, Stubbs B, et al.: The "online brain": how the Internet may be changing our cognition . World Psychiatry. 2019, 18:119-129. 10.1002/wps.20617
18. Karkosz S, Szymański R, Sanna K, Michałowski J: Effectiveness of a web-based and mobile therapy chatbot on anxiety and depressive symptoms in subclinical young adults: randomized controlled trial. JMIR Form Res. 2024, 8:e47960. 10.2196/47960
19. Fitzpatrick KK, Darcy A, Vierhile M: 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 Ment Health. 2017, 4:e19. 10.2196/mental.7785
20. Corlett PR, Fraser KM: 20 years of aberrant salience in psychosis: what have we learned? . Am J Psychiatry. 2025, 182:819-829. 10.1176/appi.ajp.20240556
21. McCutcheon RA, Krystal JH, Howes OD: Dopamine and glutamate in schizophrenia: biology, symptoms and treatment. World Psychiatry. 2020, 19:15-33. 10.1002/wps.20693
22. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ: The computational anatomy of psychosis . Front Psychiatry. 2013, 4:47. 10.3389/fpsyt.2013.00047
23. Corlett PR, Horga G, Fletcher PC, Alderson-Day B, Schmack K, Powers AR 3rd: Hallucinations and strong priors. Trends Cogn Sci. 2019, 23:114-127. 10.1016/j.tics.2018.12.001
24. Sheffield JM, Suthaharan P, Leptourgos P, Corlett PR: Belief updating and paranoia in individuals with schizophrenia. Biol Psychiatry Cogn Neurosci Neuroimaging. 2022, 7:1149-1157. 10.1016/j.bpsc.2022.03.013
25. Hortensius R, Kent M, Darda KM, Jastrzab L, Koldewyn K, Ramsey R, Cross ES: Exploring the relationship between anthropomorphism and theory-of-mind in brain and behaviour. Hum Brain Mapp. 2021, 42:4224-4241. 10.1002/hbm.25542

26. Thibaudeau E, Rae J, Raucher-Chéné D, Bougeard A, Lepage M: Disentangling the relationships between the clinical symptoms of schizophrenia spectrum disorders and theory of mind: a meta-analysis. Schizophr Bull. 2023, 49:255-274. 10.1093/schbul/sbac150
27. Egan SJ, Greene D, Callaghan T, et al.: Worry and rumination as a transdiagnostic target in young people: a co-produced systematic review and meta-analysis. Cogn Behav Ther. 2025, 54:17-40. 10.1080/16506073.2024.2369936
28. Martinelli C, Cavanagh K, Dudley RE: The impact of rumination on state paranoid ideation in a nonclinical sample. Behav Ther. 2013, 44:385-394. 10.1016/j.beth.2013.02.002
29. Head KR: Minds in crisis: how the AI revolution is impacting mental health . J Ment Health Clin Psychol. 2025, 9:34-44. 10.29245/2578-2959/2025/3.1352
30. Keshavan M, Torous J, Yassin W: Do generative AI chatbots increase psychosis risk? . World Psychiatry. 2026, 25:150-151. 10.1002/wps.70017
31. Bakanaitė R, Bakanienė I, Kanyeredzi A: Long-term effectiveness of Internet-based versus traditional cognitive behavioral therapy across psychiatric disorders: a systematic review and meta-analysis. Telemed J E Health. 2025, 31:1284-1296. 10.1177/15305627251358855
32. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E: Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther. 2018, 47:1-18. 10.1080/16506073.2017.1401115
33. Zheng Y, Xu T, Zhu Y, Li C, Wang J, Livingstone S, Zhang T: Cognitive behavioral therapy for prodromal stage of psychosis-outcomes for transition, functioning, distress, and quality of life: a systematic review and meta-analysis. Schizophr Bull. 2022, 48:8-19. 10.1093/schbul/sbab044
34. Poletti M, Pelizza L, Preti A, Raballo A: Clinical high-risk for psychosis (CHR-P) circa 2024: synoptic analysis and synthesis of contemporary treatment guidelines. Asian J Psychiatr. 2024, 100:104142. 10.1016/j.ajp.2024.104142
35. Hua Y, Na H, Li Z, Liu F, Fang X, Clifton D, Torous J: A scoping review of large language models for generative tasks in mental health care. NPJ Digit Med. 2025, 8:230. 10.1038/s41746-025-01611-4
36. Malouin-Lachance A, Capolupo J, Laplante C, Hudon A: Does the digital therapeutic alliance exist? Integrative review. JMIR Ment Health. 2025, 12:e69294. 10.2196/69294
37. Tavory T: Regulating AI in mental health: ethics of care perspective. JMIR Ment Health. 2024, 11:e58493. 10.2196/58493
38. Liu L, Li Z, Hu Y, et al.: Predictive performance of machine learning for suicide in adolescents: systematic review and meta-analysis. J Med Internet Res. 2025, 27:e73052. 10.2196/73052
39. Hua Y, Siddals S, Ma Z, et al.: Charting the evolution of artificial intelligence mental health chatbots from rule-based systems to large language models: a systematic review. World Psychiatry. 2025, 24:383-394. 10.1002/wps.21352
40. Bedi G, Carrillo F, Cecchi GA, et al.: Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr. 2015, 1:15030. 10.1038/npjschz.2015.30
41. Tang SX, Kriz R, Cho S, et al.: Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders. NPJ Schizophr. 2021, 7:25. 10.1038/s41537-021-00154-3
42. Bufano P, Laurino M, Said S, Tognetti A, Menicucci D: Digital phenotyping for monitoring mental disorders: systematic review. J Med Internet Res. 2023, 25:e46778. 10.2196/46778
43. Pichowicz W, Kotas M, Piotrowski P: Performance of mental health chatbot agents in detecting and managing suicidal ideation. Sci Rep. 2025, 15:31652. 10.1038/s41598-025-17242-4
44. Gleeson JF, McGuckian TB, Fernandez DK, et al.: Systematic review of early warning signs of relapse and behavioural antecedents of symptom worsening in people living with schizophrenia spectrum disorders. Clin Psychol Rev. 2024, 107:102357. 10.1016/j.cpr.2023.102357
45. Flinn A, Hefferman-Clarke R, Parker S, et al.: Cumulative exposure to childhood adversity and risk of adult psychosis: a dose-response meta-analysis. Psychol Med. 2025, 55:e162. 10.1017/S0033291725001138
46. Varese F, Smeets F, Drukker M, et al.: Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective- and cross-sectional cohort studies. Schizophr Bull. 2012, 38:661-671. 10.1093/schbul/sbs050
47. Taher R, Hsu CW, Hampshire C, et al.: The safety of digital mental health interventions: systematic review and recommendations. JMIR Ment Health. 2023, 10:e47433. 10.2196/47433
48. Fromm SP, Wieland L, Deneault A, Heinz A, Katthagen T, Schlagenhauf F: Neural correlates of uncertainty processing in psychosis spectrum disorder. Brain Commun. 2025, 7:fcaf073. 10.1093/braincomms/fcaf073
49. Ritunnano R, Kleinman J, Whyte Oshodi D, et al.: Subjective experience and meaning of delusions in psychosis: a systematic review and qualitative evidence synthesis. Lancet Psychiatry. 2022, 9:458-476. 10.1016/S2215-0366(22)00104-3
50. Salazar de Pablo G, Radua J, Pereira J, et al.: Probability of transition to psychosis in individuals at clinical high risk: an updated meta-analysis. JAMA Psychiatry. 2021, 78:970-978. 10.1001/jamapsychiatry.2021.0830
51. Oudin A, Maatoug R, Bourla A, et al.: Digital phenotyping: Data-driven psychiatry to redefine mental health. J Med Internet Res. 2023, 25:e44502. 10.2196/44502
52. Mougharbel F, Chaput JP, Sampasa-Kanyinga H, Colman I, Leatherdale ST, Patte KA, Goldfield GS: Longitudinal associations between different types of screen use and depression and anxiety symptoms in adolescents. Front Public Health. 2023, 11:1101594. 10.3389/fpubh.2023.1101594
53. Haarsma J, Fletcher PC, Griffin JD, et al.: Precision weighting of cortical unsigned prediction error signals benefits learning, is mediated by dopamine, and is impaired in psychosis. Mol Psychiatry. 2021, 26:5320-5333. 10.1038/s41380-020-0803-8
54. Kim M, Kim T, Ha M, Oh H, Moon SY, Kwon JS: Large-scale thalamocortical triple network dysconnectivities in patients with first-episode psychosis and individuals at risk for psychosis. Schizophr Bull. 2023, 49:375-384. 10.1093/schbul/sbac174