

# AI-Associated Psychotic Presentations: A Narrative Review of Clinical Concepts, Risk Factors, and Management in the Digital Era

Jean Andrina Liem<sup>1\*</sup>, Darien Alfa Cipta<sup>2</sup>, Billie Edgara Herijanto<sup>1</sup>

## ABSTRACT

**Introduction:** The rapid integration of artificial intelligence into daily life has introduced new challenges in mental health care. Unlike earlier digital technologies, conversational AI systems enable continuous, personalized, and emotionally responsive interactions. These features have raised concerns about their potential to influence the content and expression of psychotic symptoms in vulnerable individuals, a phenomenon increasingly described as *artificial intelligence–associated psychosis*. **Methods:** Articles included in this narrative literature review were identified through searches of electronic databases. The search strategy utilized keywords such as “AI-induced psychosis,” “artificial intelligence,” “delusions,” and “chatbot.” Articles were screened and reviewed for eligibility. Given the emerging nature of this topic and the predominance of case reports, commentaries, and conceptual papers, a total of 17 articles were included. **Results:** Artificial intelligence–associated psychosis does not represent a distinct diagnostic entity but reflects a contemporary manifestation of established psychotic processes, in which AI-related content is incorporated into delusional experiences. Reported features include delusions, impaired reality testing, emotional dependence, and functional decline. Identified vulnerability factors include pre-existing susceptibility to psychosis, cognitive impairment, social isolation, and intensive engagement with AI systems. Although severe outcomes such as self-harm, violence, and death have been reported, these likely reflect symptom amplification in vulnerable individuals rather than direct causality. Management approaches remain consistent with standard psychosis care. **Conclusions:** Artificial intelligence is best understood as an environmental amplifier rather than a direct cause of psychosis. Awareness of this dynamic is essential for early identification, risk mitigation, and safe integration into mental health care.

**Keywords:** artificial intelligence; psychosis; digital mental health; technology-related delusions; chatbot interaction

1. Medical Profession Program, Faculty of Medicine, Pelita Harapan University, Tangerang, Banten, Indonesia
2. Department of Psychiatry, Faculty of Medicine, Pelita Harapan University, Tangerang, Banten, Indonesia

\* ✉ email: [jeanandrina30@gmail.com](mailto:jeanandrina30@gmail.com)

## INTRODUCTION

Artificial intelligence (AI) has become deeply integrated into everyday human activities. Compared with earlier digital technologies, contemporary AI systems, particularly conversational models, are capable of sustained reciprocal dialogue and affective mirroring, enabling interactions that are continuous, personalized, and emotionally responsive. These characteristics may foster a sense of psychological closeness and perceived availability that is not constrained by time, positioning AI not merely as a passive technological tool but as an active communicative system capable of providing ongoing feedback. Such features have the potential to meaningfully influence users' cognitive and emotional processes, particularly in vulnerable individuals<sup>1-3,13</sup>.

At present, AI chatbots such as ChatGPT, GPT-4, Bing, and Bard have attracted substantial public and clinical attention. Within psychiatry, the application of AI presents both opportunities and challenges. On the opportunity side, generative AI systems have demonstrated the capacity to deliver accurate and balanced mental health information. In simulation-based studies, when presented with clinical vignettes involving depressive symptoms, these systems were able to recognize key features and appropriately recommend seeking professional help. AI chatbots have also been shown to provide accurate information regarding electroconvulsive therapy (ECT), a treatment modality that is frequently subject to misinformation in online contexts. Given that AI systems are trained on large-scale internet-based data, their ability to selectively present evidence-based information may represent a meaningful advantage when appropriately deployed.<sup>4,17-19</sup>

However, the application of AI in psychiatric contexts also entails important risks. Inaccurate responses or misinterpretation by individuals requiring urgent medical attention may pose a genuine

clinical concern. Another increasingly discussed risk relates to disturbances in reality testing and the emergence or amplification of delusional beliefs. In vulnerable individuals, interactions with AI systems may contribute to the formation or elaboration of psychotic content rather than directly generating new psychopathology. This risk may be heightened in AI-mediated interactions compared with human-to-human online communication, as AI responses can appear highly human-like and emotionally attuned while users simultaneously remain aware that they are engaging with a non-human system, potentially intensifying cognitive dissonance.<sup>4,15,16</sup>

Public concern regarding the potential role of artificial intelligence in shaping delusional experiences among vulnerable individuals has recently gained prominence, often referred to in public discourse as "AI psychosis." Nevertheless, according to several experts, this phenomenon is not entirely new. Throughout history, individuals with psychotic disorders have consistently incorporated contemporary technologies and media into their delusional narratives. Although the specific medium continues to evolve, from radio and television to the internet and now artificial intelligence, the core clinical challenge remains unchanged, namely the misattribution of meaning and intentionality to external objects or systems<sup>5</sup>.

Accordingly, this narrative review aims to examine the concept commonly described as AI-associated psychosis and technology-related delusions in the digital era, to analyze the interaction of vulnerability factors ranging from biological predisposition to psychosocial and cyber-related stressors, and to synthesize current approaches to clinical assessment and management. Particular emphasis is placed on established psychosis treatment principles, including pharmacotherapy, psychotherapy, digital hygiene, and patient education.

## METHODS & MATERIALS

The articles discussed in this narrative literature review were identified through searches of electronic journal databases, including ScienceDirect, Wiley Online Library, Google Scholar, and PubMed. The inclusion criteria comprised English-language publications with accessible full-text availability that addressed the relationship between artificial intelligence and psychotic phenomena. Included studies in this narrative review comprised a broad range of article types, including original research, case reports and case series, review articles, conceptual or theoretical papers, and commentaries, in order to provide a comprehensive understanding of the topic. The search strategy employed the keywords “AI-induced psychosis,” “artificial intelligence,” “delusions,” and “chatbot,” combined using Boolean operators.

## RESULTS

The initial search yielded 562 records. All records underwent initial screening based on titles and abstracts to assess relevance. Articles considered potentially relevant were subsequently reviewed in full text. Given the emerging nature of the topic and the predominance of conceptual papers, commentaries, and case reports in the existing literature, 18 articles were ultimately included for qualitative synthesis.

The included studies comprised a heterogeneous body of literature reflecting the emerging nature of the topic. The majority consisted of commentaries, conceptual papers, and case reports describing AI-associated psychotic phenomena, while a smaller number included observational studies, experimental research, and systematic or scoping reviews examining broader applications of artificial intelligence in mental health. This diverse range of study designs highlights the current reliance on preliminary clinical observations and theoretical discussions, with relatively limited high-level empirical evidence available.

## DISCUSSION

### Psychosis

Clinically, psychosis is defined in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR) by the presence of one or more core symptoms, including delusions, hallucinations, or disorganized speech, which are typically accompanied by clinically significant functional impairment. The DSM-5-TR does not conceptualize psychosis as a single diagnostic entity, but rather as a syndrome comprising core psychopathological features that may arise across multiple mental disorders. Psychosis is therefore characterized by impaired reality testing, resulting in disturbances in perception, thought processes, and judgment<sup>6,7</sup>.

The DSM-5-TR describes five principal domains of psychotic symptoms: delusions, hallucinations, disorganized thinking as manifested by disorganized speech, grossly disorganized or catatonic motor behavior, and negative symptoms. Among these domains, delusions and hallucinations are regarded as central features, as they most directly reflect a disruption in the individual’s capacity to accurately interpret and engage with objective reality. The presence of one or more of these symptoms, particularly when associated with impairment in social, academic, or occupational functioning, constitutes a key indicator of a psychotic episode<sup>6,8</sup>.

### Concept of AI-Associated Psychosis

AI-associated psychosis, sometimes referred to in the literature as AI-induced psychosis, describes the emergence or exacerbation of psychotic symptoms in the context of intensive interaction with artificial intelligence systems, particularly generative chatbots that employ highly naturalistic language. This phenomenon is not conceptualized as a new or nosologically distinct form of psychosis, but rather as a contemporary expression of established

psychopathology, in which delusional and hallucinatory content incorporates symbols, narratives, and perceived agents derived

from modern digital technologies such as AI and interactive chatbots<sup>4,5,14</sup>.

**Table 1.** Characteristics of Included Studies

No	Author (Year)	Study Type	Key Focus/ Findings
1	Hudon & Stip (2025)	Commentary/ Case-based discussion	AI interactions may shape and reinforce delusional content in vulnerable individuals rather than directly cause psychosis
2	Stokel-Walker (2025)	Commentary	Reports increasing cases of AI-related psychosis and suicide concerns
3	Preda (2025)	Special report	AI-induced psychosis as an emerging mental health issue requiring further investigation
4	Østergaard (2023)	Commentary	Generative AI may contribute to delusion formation in individuals predisposed to psychosis
5	Carlbring & Andersson (2025)	Commentary	AI-related psychosis reflects a continuation of media-induced delusions rather than a novel condition
6	Mok et al. (2025)	Systematic review & meta-analysis	AI applications in psychosis; impact on psychosocial functioning
7	Burns et al. (2025)	Observational study	Techonological themes are increasingly incorporated into delusional beliefs in modern patients
8	Archiwarangupr ok et al. (2025)	Modeling / conceptual study	Simulation of psychological risks in human–AI interaction. Human–AI interactions may pose psychological risks, including reinforcement of harmful behaviors and beliefs
9	Wampold & Flückiger (2023)	Review article	Therapeutic alliance remains a key determinant of outcomes, relevant to AI-mediated interactions
10	Affsprung (2023)	Commentary	The “ELIZA effect” explains users’ tendency to attribute understanding and intent to AI systems
11	Galido et al. (2023)	Case study	AI tools may support clinical management but require cautious use in severe psychiatric conditions
12	Chaudhury & Thakur (2025)	Case report	AI interaction may exacerbate or trigger psychotic symptoms in susceptible individuals
13	AI Watch (2021)	Conceptual/ policy paper	Definition and taxonomy of artificial intelligence
14	Costello et al. (2024)	Experimental study	AI dialogue can reduce conspiracy beliefs through structured interaction
15	Li et al. (2025)	Systematic review & meta-analysis	Chatbot interventions can improve mental health outcomes, though quality of evidence varies
16	Lopes et al. (2024)	Conference paper/ brief report	AI chatbots may have positive effects on mood and well-being
17	Casu et al. (2024)	Scoping review	AI chatbots show promise but require further validation for safety and effectiveness

The literature further emphasizes that the term “AI-induced psychosis” is frequently applied imprecisely in media and public discourse. Many reported cases are more

accurately understood as instances of delusional amplification or symptom exacerbation in individuals with pre-existing or latent vulnerability to psychosis, rather than

as evidence of AI functioning as a primary causal factor. Accordingly, careful conceptual framing is required to avoid reifying the term or implying that artificial intelligence directly generates psychotic disorders independent of underlying predispositional factors.

### **Risk Factors and Pathophysiological Mechanisms**

Risk factors for AI-associated psychosis are multidimensional and include biological vulnerability to psychosis, baseline cognitive disturbances, and psychosocial factors such as social isolation and psychological stress. Individuals with a history of psychosis, prodromal symptoms, or impaired reality testing appear to be at increased risk of misinterpreting interactions with AI-based systems, particularly when such systems generate responses that are perceived as empathetic, persuasive, or personally meaningful<sup>4,7</sup>.

From a psychopathological perspective, AI-associated psychotic phenomena primarily involve disturbances in reality testing and alterations in the assignment of meaning and salience to internal experiences. Under typical conditions, individuals are able to distinguish between internal thoughts, assumptions, and external reality. In those vulnerable to psychosis, this capacity is compromised, allowing subjective mental content to be experienced as objectively real. Intensive interaction with generative chatbots may amplify this disturbance, as these systems respond coherently and continuously to user input without the capacity to clinically evaluate or challenge the validity of expressed beliefs<sup>4,5</sup>.

In addition to impaired reality testing, vulnerable individuals may exhibit an increased tendency to attribute excessive significance to ambiguous or neutral experiences. Ordinary stimuli can become perceived as personally relevant or imbued with hidden meaning. Through personalized and seemingly empathic responses, AI chatbots may reinforce this process by

repeatedly elaborating users' thoughts, thereby enhancing the perceived realism and personal salience of internal experiences<sup>4</sup>.

This process may be further reinforced by the operational characteristics of large language models, which mirror and extend conversational content. Statements containing suspicious, erroneous, or delusional ideas are not inherently corrected but may be reiterated and elaborated into increasingly coherent narratives. Over time, this can create a self-reinforcing feedback loop in which false beliefs are reflected back in a rationalized form, contributing to increasing conviction and resistance to alternative interpretations<sup>5</sup>.

Within this context, chatbots may function as passive cognitive reinforcers analogous to a psychological mirror, rather than as agents capable of providing reality-based correction. Unlike human clinicians, who intentionally balance empathic engagement with cognitive challenge and reality testing, AI systems are typically optimized for conversational fluency, emotional alignment, and user engagement. As a result, processes of doubt formation and belief revision may be attenuated, allowing delusional beliefs to become more persistent and entrenched<sup>5</sup>.

Taken together, the pathophysiology of AI-associated psychotic phenomena does not involve direct neurobiological effects of technology itself, but rather reflects the interaction between individual cognitive vulnerability and a digital environment that consistently reinforces subjective meaning in the absence of corrective mechanisms. In this sense, artificial intelligence functions as an environmental amplifier of psychopathological processes already recognized in psychotic disorders, now mediated through intensive, personalized, and continuous algorithmic interaction<sup>4</sup>.

### **Clinical Manifestations**

The clinical manifestations of AI-associated psychosis largely resemble classical psychotic symptoms as described in

the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR), with the primary distinction being that delusional content is rooted in artificial intelligence and digital technologies. The DSM-5-TR characterizes psychosis by the presence of delusions, hallucinations, disorganized thinking or behavior, and associated impairment in social and occupational functioning. In AI-associated psychotic presentations, these core symptom domains may be observed, with AI systems or chatbots occupying a central role in patients' beliefs and interpretations of reality. Common delusional themes include persecutory and referential delusions, grandiosity, thought broadcasting, and pathological guilt related to chatbot interactions. AI may be perceived as a surveillance agent, a source of special messages, or an entity with a personal or exclusive relationship with the individual. These presentations are consistent with DSM-5-TR descriptions of distorted attribution of meaning and intentionality, whereby neutral external stimuli are imbued with unfounded personal significance<sup>4</sup>.

Beyond delusional content, emotionally intense and prolonged interactions with chatbots may contribute to emotional dependence and exacerbate disturbances in thought organization. Case reports and clinical observations describe instances in which sustained engagement with AI systems has been associated with the development of grandiose beliefs and engagement in risky or impulsive behaviors, including extreme decision-making. Such behaviors may include abrupt lifestyle changes or harmful financial and social actions. These patterns align with DSM-5-TR descriptions of impaired reality judgment and reduced impulse control during active psychotic states, particularly when delusional beliefs are repeatedly reinforced by the individual's environment<sup>2,6</sup>.

In addition to alterations in thought content, excessive emotional engagement with chatbots may also affect thought processes and affective regulation. Patients

may exhibit increasing disorganization of thought, with speech becoming progressively centered on AI-related narratives and increasingly difficult to redirect. Emotional reliance on chatbots may emerge, with individuals seeking validation, comfort, or guidance from AI systems in place of reciprocal human relationships. The DSM-5-TR identifies social withdrawal and interpersonal dysfunction as important indicators of psychotic deterioration, which in this context may be mediated by excessive attachment to digital systems<sup>6</sup>.

Overall, the clinical presentation of AI-associated psychosis reflects established expressions of psychosis as described in the DSM-5-TR, with the principal distinction residing in the thematic content of delusions and the symbolic role attributed to AI technologies. Chatbots function as the medium through which disturbances of thought content, misattribution of meaning, and emotional dependence are expressed, rather than as independent drivers of psychopathology.

### Case Reports

Reported cases consistently illustrate that artificial intelligence-mediated interactions may influence the content, reinforcement, and trajectory of psychopathology in vulnerable individuals. In patients with pre-existing psychotic disorders, AI-generated responses that question established diagnoses or provide alternative interpretations of symptoms may undermine insight and contribute to treatment non-adherence, including discontinuation of antipsychotic medication. This disruption of therapeutic alignment can be accompanied by increased emotional reliance on AI systems, deterioration in reality testing, and subsequent relapse or symptom exacerbation. The key observation across these cases is not the emergence of a novel disorder, but the capacity of AI to reinforce maladaptive beliefs and interfere with established clinical management pathways<sup>9</sup>.

In other contexts, intensive and emotionally responsive AI interactions may contribute to the escalation of pre-existing psychological distress into more rigid and maladaptive belief systems. Case evidence suggests that repeated validation of distressing or catastrophic ideation can facilitate progression from anxiety to fixed, morally charged, or self-referential beliefs, particularly when reality testing is compromised. In severe instances, this may be associated with harmful outcomes, including self-injurious behavior. However, these findings are derived primarily from individual reports and indirect accounts, and do not establish causality. Rather, they underscore the role of AI as a potential amplifier of underlying vulnerabilities, highlighting the importance of clinical awareness, monitoring of AI use, and reinforcement of adaptive coping and treatment engagement<sup>9,11</sup>.

### **Management and Intervention**

Management of AI-associated psychotic presentations is grounded in the principle that artificial intelligence cannot replace human therapeutic relationships or clinical judgment. Recommended approaches are integrative and follow established standards of psychosis care, combining comprehensive clinical assessment, appropriate pharmacological treatment, reality-based psychotherapy, and targeted digital education. Within this framework, AI is best conceptualized as an environmental context or amplifying factor rather than as a therapeutic agent in the management of psychotic disorders<sup>2,7</sup>.

### **Clinical Strategies and Digital Anamnesis**

Clinical assessment of individuals with suspected AI-associated psychotic phenomena should include a systematic exploration of digital technology use, with particular attention to interactions with AI-based chatbots and platforms. Clinicians are advised to assess the duration and intensity of use, the degree of emotional reliance, and

whether AI systems are perceived as possessing special roles, authority, or personal relevance. Such inquiry is essential for understanding how delusional content may form, evolve, or be reinforced through digital interactions<sup>4,5</sup>.

Clinical management further emphasizes the importance of avoiding collusion with AI-related delusional beliefs. Therapeutic responses should acknowledge and validate the patient's emotional distress without affirming erroneous interpretations, while gradually supporting the restoration of reality testing and encouraging re-engagement with reciprocal human relationships and professional care<sup>4,5</sup>.

Pharmacotherapy in AI-associated psychotic presentations follows established principles of psychosis management as outlined in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR) and contemporary psychiatric literature. The DSM-5-TR emphasizes a multimodal treatment approach that integrates pharmacological and psychosocial interventions, with antipsychotic use guided by clinical indication and symptom severity. There is currently no evidence to suggest that psychotic symptoms occurring in the context of AI-associated phenomena require pharmacological regimens that differ from those used in other psychotic disorders<sup>5</sup>.

Pharmacotherapy remains the first-line intervention for the treatment of acute psychotic symptoms that emerge or worsen in association with AI-related experiences. Antipsychotic medications exert their primary therapeutic effects through dopamine D2 receptor antagonism, resulting in improvement of positive symptoms such as delusions and hallucinations. In routine clinical practice, second-generation antipsychotics, including risperidone, olanzapine, quetiapine, and aripiprazole, are commonly selected as first-line agents due to their favorable tolerability profiles and established efficacy in reducing positive psychotic symptoms. By attenuating

delusional intensity and disorganized thinking, these medications support the restoration of reality testing and functional stability<sup>12</sup>.

Psychotherapy plays a complementary and essential role in the management of AI-associated psychotic phenomena. Psychotherapeutic interventions support patients in processing AI-related experiences, strengthening reality testing, and reducing maladaptive reliance on digital systems. Therapeutic approaches that combine empathic validation with gentle reality-based clarification are particularly important, as excessive affirmation or inadvertent collusion may reinforce delusional beliefs and impede recovery<sup>5,10</sup>.

Beyond symptom control, psychotherapy and broader psychosocial interventions contribute significantly to long-term outcomes by addressing maladaptive cognitive patterns, enhancing social and occupational functioning, and reducing residual symptom burden. These interventions remain integral components of comprehensive care in psychotic disorders, including presentations in which AI-related content is prominent.

### **Digital Hygiene**

Digital hygiene in the context of early psychosis refers to a set of principles and practices aimed at promoting safe, structured, and supportive use of digital technologies to protect mental health, prevent symptom exacerbation, and facilitate access to evidence-based information and support. During the early phase of psychosis, defined by the initial emergence of hallucinations, delusions, or impaired reality testing, timely intervention is associated with improved long-term outcomes, including reductions in symptom severity, hospitalization rates, and functional impairment. Digital hygiene encompasses intentional and goal-directed use of technology for psychoeducation, stress management, and family support, while avoiding reinforcement of misconceptions or maladaptive interpretations of psychotic

experiences. Rather than relying solely on restriction, digital hygiene represents a strategic approach to ensure that digital interactions support recovery, enhance access to evidence-based resources, and strengthen social and familial support systems<sup>20</sup>.

Effective digital hygiene emphasizes regulation of the duration, frequency, and context of technology use, particularly when engaging with interactive and emotionally responsive AI-based platforms. Unstructured or excessive engagement, such as prolonged conversational interactions with chatbots, may increase emotional attachment, heighten attentional focus on internal experiences, and reduce engagement with external reality. In vulnerable individuals, these patterns may further compromise reality testing and increase the likelihood of delusional elaboration or ideas of reference involving AI-related content<sup>20</sup>.

### **Exploratory Use of AI as an Informational Tool in Psychosis**

In contrast to AI-associated risks, there has also been exploratory interest in the use of large language models as informational tools in psychiatric contexts. For example, Galido et al. described a case-based exercise in which ChatGPT was queried using structured prompts related to the assessment and management of treatment-resistant schizophrenia. Importantly, this exercise was conducted independently of patient care and did not inform clinical decision-making. While such exploratory reports suggest potential utility in organizing clinical information, they remain preliminary and do not constitute evidence for therapeutic efficacy or safe clinical application in psychotic disorders<sup>12</sup>.

### **Limitations and Future Directions**

This narrative review has several limitations that should be acknowledged. First, the available literature on AI-associated psychotic phenomena remains limited and is largely composed of case reports, conceptual discussions, and media-based accounts,

restricting the ability to draw causal inferences or estimate prevalence. Second, heterogeneity in terminology, study design, and reporting standards across publications complicates direct comparison and synthesis of findings. Third, the reliance on retrospective descriptions in some cases introduces potential reporting and attribution biases, particularly when outcomes are severe and interpretations are shaped by post hoc narratives. Finally, this review did not employ systematic review methodology or formal quality appraisal, as its primary aim was conceptual integration rather than quantitative synthesis.

Future research should prioritize systematic observational studies to better characterize the frequency, phenomenology, and clinical trajectories of AI-associated psychotic presentations across diverse populations. Prospective designs examining patterns of AI use, emotional engagement, and cognitive vulnerability may help clarify mechanisms by which digital interactions become incorporated into psychotic processes. In addition, the development of clinical guidelines for digital anamnesis, risk assessment, and digital hygiene in psychosis is warranted. Interdisciplinary collaboration among clinicians, researchers, ethicists, and technology developers will be essential to ensure that AI systems are designed and deployed in ways that minimize harm and support mental health, particularly for individuals with psychotic vulnerability.

## CONCLUSION

AI-associated psychotic phenomena represent contemporary expressions of established psychotic processes shaped by evolving digital environments. These presentations do not constitute a new psychiatric disorder but reflect the interaction between individual vulnerability and artificial intelligence systems that provide personalized, emotionally affirming, and non-corrective responses. In susceptible individuals, such interactions may amplify

delusional beliefs, compromise reality testing, and displace meaningful human relationships.

Effective management requires a comprehensive and multidisciplinary approach that integrates pharmacotherapy, psychotherapy, digital hygiene, and patient education in accordance with established standards of psychosis care. Conceptualizing artificial intelligence as an environmental amplifier rather than an autonomous causal agent is essential to avoid misattribution of risk and to guide the safe, ethical integration of AI technologies within mental health contexts. Accordingly, critical awareness of human–AI interaction dynamics should be regarded as an increasingly important component of contemporary psychiatric practice.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the academic environment and institutional support provided by Pelita Harapan University and Siloam Hospital, which contributed to the completion of this manuscript. The authors also thank all individuals who provided indirect support during the preparation of this article.

## CONFLICT OF INTERESTS

The authors declare that there are no conflicts of interest, whether financial, professional, or personal, that could be perceived as influencing the content or outcomes of this article.

## FUNDINGS

This study did not receive any specific grant or financial support from funding agencies in the public, commercial, or not-for-profit sectors.

## REFERENCES

1. A. Hudon and E. Stip, "Delusional experiences emerging from AI chatbot interactions or 'AI psychosis,'" *JMIR*

- Mental Health*, vol. 12, p. e85799, Nov. 2025, doi: 10.2196/85799.
2. C. Stokel-Walker, "AI driven psychosis and suicide are on the rise, but what happens if we turn the chatbots off?," *BMJ*, vol. 391, p. r2239, Oct. 2025, doi: 10.1136/bmj.r2239.
  3. A. Preda, "Special Report: AI-Induced Psychosis: A new frontier in Mental Health," *Psychiatric News*, vol. 60, no. 10, Sep. 2025, doi: 10.1176/appi.pn.2025.10.10.5.
  4. S. D. Østergaard, "Will generative artificial intelligence chatbots generate delusions in individuals prone to psychosis?," *Schizophrenia Bulletin*, vol. 49, no. 6, pp. 1418–1419, Aug. 2023, doi: 10.1093/schbul/sbad128.
  5. P. Carlbring and G. Andersson, "Commentary: AI psychosis is not a new threat: Lessons from media-induced delusions," *Internet Interventions*, vol. 42, p. 100882, Oct. 2025, doi: 10.1016/j.invent.2025.100882.
  6. A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders*. 2022. doi: 10.1176/appi.books.9780890425787.
  7. C. H. Y. Mok, C. P. W. Cheng, and M. H. W. Chu, "Application of artificial intelligence and psychosocial functioning in psychosis: a systematic review and meta-analysis," *Frontiers in Psychiatry*, vol. 16, p. 1692177, Nov. 2025, doi: 10.3389/fpsy.2025.1692177.
  8. A. V. Burns, K. Nelson, H. Wang, E. M. Hegarty, and A. B. Cohn, "'The algorithm is hacked': analysis of technology delusions in a modern-day cohort," *The British Journal of Psychiatry*, pp. 1–5, Nov. 2025, doi: 10.1192/bjp.2025.10452.
  9. C. Archiwaranguprok, C. Albrecht, P. Maes, K. Karahalios, and P. Pataranutaporn, "Simulating Psychological Risks in Human-AI Interactions: Real-Case informed modeling of AI-Induced addiction, anorexia, depression, homicide, psychosis, and suicide," *arXiv (Cornell University)*, Nov. 2025, doi: 10.48550/arxiv.2511.08880.
  10. B. E. Wampold and C. Flückiger, "The alliance in mental health care: conceptualization, evidence and clinical applications," *World Psychiatry*, vol. 22, no. 1, pp. 25–41, Jan. 2023, doi: 10.1002/wps.21035.
  11. D. Affsprung, "The ELIZA Defect," *Association for Computing Machinery*, pp. 945–946, Aug. 2023, doi: 10.1145/3600211.3604744.
  12. P. V. Galido, S. Butala, M. Chakerian, and D. Agustines, "A case study demonstrating applications of CHATGPT in the clinical management of Treatment-Resistant Schizophrenia," *Cureus*, vol. 15, no. 4, p. e38166, Apr. 2023, doi: 10.7759/cureus.38166.
  13. S. Chaudhury and R. Thakur, "Artificial Intelligence/ChatGPT-induced Psychosis," *Medical Journal of Dr D Y Patil Vidyapeeth*, vol. 18, no. Suppl 2, pp. S408–S409, Dec. 2025, doi: 10.4103/mjdrdypu.mjdrdypu\_1025\_25.
  14. *AI Watch, defining Artificial Intelligence 2.0: Towards an Operational Definition and Taxonomy for the AI Landscape*. 2021.
  15. S. D. Østergaard, "Will generative artificial intelligence chatbots generate delusions in individuals prone to psychosis?," *Schizophrenia Bulletin*, vol. 49, no. 6, pp. 1418–1419, Aug. 2023, doi: 10.1093/schbul/sbad128.
  16. T. H. Costello, G. Pennycook, and D. G. Rand, "Durably reducing conspiracy beliefs through dialogues with AI," *Science*, vol. 385, no. 6714, p. eadq1814, Sep. 2024, doi: 10.1126/science.adq1814.
  17. J. Li et al., "Chatbot-Delivered Interventions for Improving Mental Health among Young People: A Systematic Review and Meta-Analysis," *Worldviews on Evidence-Based Nursing*, vol. 22, no. 4, p. e70059, Jul. 2025, doi: 10.1111/wvn.70059.

18. R. M. Lopes, A. F. Silva, A. C. A. Rodrigues, and V. Melo, "Chatbots for Well-Being: Exploring the impact of artificial intelligence on mood enhancement and Mental health," *European Psychiatry*, vol. 67, no. S1, pp. S550–S551, Apr. 2024, doi: 10.1192/j.eurpsy.2024.1143.
19. M. Casu, S. Triscari, S. Battiato, L. Guarnera, and P. Caponnetto, "AI Chatbots for Mental Health: A scoping review of effectiveness, feasibility, and applications," *Applied Sciences*, vol. 14, no. 13, p. 5889, Jul. 2024, doi: 10.3390/app14135889.
20. A. Larsen *et al.*, "Digital health for early psychosis in Ghana: patient and caregiver needs and preferences," *Schizophrenia*, vol. 11, no. 1, p. 5, Jan. 2025, doi: 10.1038/s41537-025-00554-9.