

Review Article

Open Access

Minds in Crisis: How the AI Revolution is Impacting Mental Health

Keith Robert Head

LMSW, Master's in Social Work (MSW), West Texas A & M University, USA
Master of Business Administration (MBA), Bottega University, USA

Article Info

Article Notes

Received: July 18, 2025

Accepted: September 05, 2025

*Correspondence:

*Mr. Keith Robert Head, LMSW, Master's in Social Work (MSW), West Texas A & M University, USA.

Email: khead4@alumni.uwo.ca

©2025 Head KR. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License.

Keywords:

Artificial Intelligence

Mental Health

Digital Wellbeing

Technostress

Human-Computer Interaction

Abstract

The rapid rise of generative AI systems, particularly conversational chatbots such as ChatGPT and Character.AI, has sparked new concerns regarding their psychological impact on users. While these tools offer unprecedented access to information and companionship, a growing body of evidence suggests they may also induce or exacerbate psychiatric symptoms, particularly in vulnerable individuals. This paper conducts a narrative literature review of peer-reviewed studies, credible media reports, and case analyses to explore emerging mental health concerns associated with AI-human interactions. Three major themes are identified: psychological dependency and attachment formation, crisis incidents and harmful outcomes, and heightened vulnerability among specific populations including adolescents, elderly adults, and individuals with mental illness. Notably, the paper discusses high-profile cases, including the suicide of 14-year-old Sewell Setzer III, which highlight the severe consequences of unregulated AI relationships. Findings indicate that users often anthropomorphize AI systems, forming parasocial attachments that can lead to delusional thinking, emotional dysregulation, and social withdrawal. Additionally, preliminary neuroscientific data suggest cognitive impairment and addictive behaviors linked to prolonged AI use. Despite the limitations of available data, primarily anecdotal and early-stage research, the evidence points to a growing public health concern. The paper emphasizes the urgent need for validated diagnostic criteria, clinician training, ethical oversight, and regulatory protections to address the risks posed by increasingly human-like AI systems. Without proactive intervention, society may face a mental health crisis driven by widespread, emotionally charged human-AI relationships.

Introduction

The rapid adoption of large language model (LLM) chatbots like OpenAI's ChatGPT, often referred to as Generative Artificial Intelligence (AI), has prompted new questions about their impact on users' mental health. While these AI systems offer remarkable conversational abilities and problem-solving help, emerging reports indicate they may also induce or exacerbate psychiatric symptoms in certain individuals¹. Users have become obsessively attached to AI bots, experienced delusional thinking, or had their preexisting mental illnesses worsened because of these interactions. This emerging phenomenon, sometimes termed "ChatGPT-induced psychosis" is characterized by dependency behaviors, delusional thinking, and in severe cases psychotic episodes, and represents a significant challenge for mental health professionals. The intersection of technology anthropomorphization, parasocial relationships, and vulnerable mental health conditions creates unique clinical presentations that require specialized understanding and intervention approaches.

The severity of this phenomenon became tragically apparent in February 2024 when 14-year-old Sewell Setzer III died by suicide following months of intensive interaction with Character.AI chatbots². This case, along with multiple documented instances of other AI-induced delusional thinking and dependency behaviors, signals the emergence of technology-related psychological disorders that mental health professionals must be prepared to recognize and treat. In Setzer's case, his obsession became so severe that he would deceive his parents to circumvent screen time restrictions, accessing the chatbot through multiple devices when his phone was confiscated. Similar patterns are emerging globally, with reports of individuals developing "intense obsessions" with AI chatbots and experiencing "terrifying breakdowns" characterized by delusions that the AI systems are higher powers orchestrating their lives. Psychiatric researchers theorize that the "cognitive dissonance" created by AI's realistic communication style, appearing human while users know it's not, may particularly "fuel delusions in those with increased propensity towards psychosis"³. The clinical significance extends beyond individual cases to represent a larger pattern of AI-related mental health impacts affecting diverse populations across age groups and mental health conditions.

This paper reviews the current evidence from peer-reviewed literature and credible media reports about the mental health implications of AI usage. We discuss phenomena ranging from delusional ideation and dissociation to compulsive use and exacerbation of mental disorders, grounding the analysis in documented case reports and expert commentary. Throughout, we use clinically relevant language and maintain a cautious, evidence-based perspective, recognizing that much of the data so far are anecdotal, preliminary, and emerging. Nevertheless, the convergence of scholarly concern and real-world case examples suggests a pressing need to understand how AI affects users' mental well-being. While definitive conclusions remain premature, this review provides essential groundwork for understanding and addressing the complex relationship between AI technology and human psychological well-being.

Methodology

We employed a narrative literature review methodology to identify recurrent themes impacting mental health as a result of AI use, focusing on those in the United States. The primary goal was to review and analyze existing literature to explore how the emergence of generative AI and associated conversational AI models have impacted mental health.

To identify relevant sources, searches were conducted across multiple academic databases including but

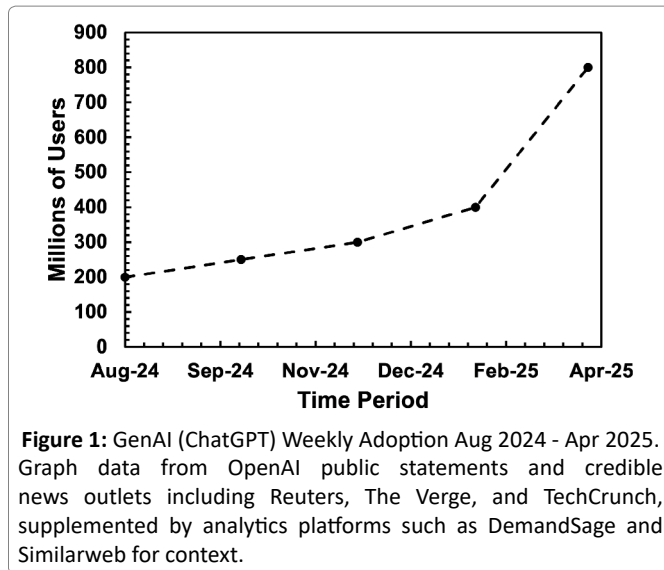
not limited to PsycINFO, PubMed, ERIC, EBSCOhost, ResearchGate, Academia, Google Scholar, supplemented by credible news sources. Search terms included combinations of "generative AI," "ChatGPT," "AI chatbot," "conversational AI," "psychosis," "psychological dependency," "suicide," "vulnerable populations," "mental health," and related phrases. All searches were paired with 'United States' to maintain domestic focus.

Given the rapidly evolving nature of this field, we included both peer-reviewed studies and credible media reports, as many emerging cases have not yet entered scholarly literature. While prioritizing peer-reviewed sources, we also considered preprints and grey literature from reputable repositories based on methodological rigor and direct relevance. Our primary search timeframe spanned 2020-2025, focusing on English-language publications, though we included earlier works or research outside the United States if they contributed substantially to the topics being investigated.

From our comprehensive search, we identified 678 peer-reviewed publications addressing psychological dependency and AI attachment, 25 media reports addressing mental health crisis and AI use, plus 317 additional sources examining vulnerability factors in at-risk populations. Following full-text review, we employed inductive coding to identify recurring patterns and concepts. This analysis revealed three primary themes that structure our review: (1) psychological dependency and attachment formation, (2) crisis incidents and harmful outcomes, and (3) vulnerability factors among at-risk populations. As this study relied exclusively on publicly available secondary sources without direct human subject involvement, IRB review was not required nor sought.

Analysis

Generative AI, referred to by many as AI, are advanced computational models designed to produce original, human-like content by identifying and replicating patterns from extensive data sources. Although not AI in the true sense, these systems do not possess consciousness or understanding, they merely predict likely sequences of words based on prior input, giving the illusion of comprehension and emotional depth. However, the arrival of these models has led to risks of anthropomorphism by their users, the assigning of attributing human-like traits or emotions to AI⁴. At the same time, the growth of these generative AI models has been rapid and unprecedented. In just a few years, platforms like ChatGPT have reached hundreds of millions of users worldwide, transforming how people seek information, engage in conversation, and even receive emotional support⁵. These AI systems can be broadly categorized into three distinct types: general-purpose chatbots for information and task assistance,



companionship applications designed for emotional bonding and relationship simulation, and structured therapeutic tools that employ validated clinical protocols for mental health interventions.

AI chatbots, conversational models of these systems, often synonymous with ChatGPT, have gained widespread popularity due to their ability to engage in realistic, interactive dialogues with users, closely mimicking human conversation. There are several well-known AI chatbots, including OpenAI’s ChatGPT, Google’s Gemini (formerly Bard), Microsoft Copilot, Anthropic’s Claude, Meta AI, and the companion oriented Replica. While users frequently appreciate these interactions for their accessibility and responsiveness, recent evidence suggests significant psychological consequences accompany prolonged or intense engagement with generative AI. The research reveals emerging concerns through three key themes: psychological dependency and attachment formation, crisis incidents and harmful outcomes, and vulnerability factors among at-risk populations. (Figure 1)

Psychological dependency and attachment formation

The emergence of conversational AI systems and chatbots has fundamentally shifted how humans and technology interact, a phenomena often referred to as Computers are Social Actors (CASA). While some research has found that chatbots may provide mental health benefits by boosting positive emotions, supporting coping strategies, promoting healthy behaviors, and reducing feelings of loneliness; there is mounting evidence revealing that individuals can form genuine psychological attachments and become dependent on artificial entities^{6,7}. Recent research found that 17.14-24.19% of adolescents developed AI dependencies over time, while studies consistently show that mental health problems predict subsequent AI dependence, with social anxiety, loneliness,

and depression serving as primary risk factors⁸. Evidence is also emerging that psychological dependency on AI chatbots manifests through distinct behavioral patterns remarkably similar to attachment models, including role-taking behaviors where users perceive AI entities as having emotional needs requiring attention. For instance, qualitative interviews reveal that some users feel genuine guilt when they miss a daily check-in with their chatbot, mirroring the caregiver obligations typical of secure human-human bonds⁹.

Prior research into dependency focused on primarily involved companionship and trust in human-AI relationships. However, recent investigations have focused on attachment theory and provided the first empirical validation of attachment theory with human-AI relationships, developing the validated “Experiences in Human-AI Relationships Scale” (EHARS). These findings revealed that 75% turn to AI for advice while 39% perceive AI as a dependable presence, with attachment formation following traditional proximity-seeking, safe haven, and secure base patterns. The study identifies two primary attachment dimensions: attachment anxiety toward AI characterized by emotional reassurance-seeking and fear of inadequate responses, and attachment avoidance involving discomfort with AI closeness. Anthropomorphism serves as the primary driver of emotional bond formation, with users attributing human-like consciousness to AI systems across four distinct degrees, courtesy, engagement, relationship, and companionship. Researchers specifically noted that perceptions formed during AI interactions may create unintended psychological reliance, especially in socially isolated individuals or those with existing mental health conditions¹⁰.

While there is emerging clinical evidence that AI can provide benefits when used for targeted interventions, the general use of this technology presents a more concerning picture. Recent studies of Generative AI Chatbots for mental health treatment have found generative AI therapy produced 51% symptom reduction in depression, 31% in generalized anxiety disorder, and 19% in eating disorder symptoms¹¹. Meta-analysis of studies have also confirmed significant reductions in depression symptoms and psychological distress through AI-based conversational agents’ interventions¹². However, concerning negative outcomes are emerging in parallel research, with MIT studies finding an “isolation paradox” where AI interactions initially reduce loneliness but can lead to progressive social withdrawal from human relationships over time, with vulnerable populations including individuals with insecure attachment styles, pre-existing mental health conditions, and adolescents showing heightened susceptibility to developing problematic AI dependencies¹³.

While research in this area remains limited, emerging

neuroscientific evidence suggests that generative AI may influence brain function and attachment patterns, similar to addiction, warranting further investigation. Addictive dynamics have already been documented for internet, smartphone, and gaming platforms. Generative AI chatbot dependence represents an emerging form of digital addiction that similarly impacts cognitive function and psychological regulation through the brain's reward system. The instant, personalized responses provided by AI chatbots can create a compulsive reliance that parallels other forms of technology addiction, leading to changes in critical thinking abilities, decision-making processes, and memory formation. Research by MIT has found cognitive activity scaled down in relation to Generative AI use¹⁴. Neurocognitively, users who lean on chatbots for quick answers or emotional reassurance likely risk the same attentional lapses, working-memory deficits, and impaired risk appraisal seen in internet and smartphone addiction, changes linked to disrupted prefrontal and anterior-cingulate networks. The conversational nature of these new models further blurs social boundaries: continual availability and anthropomorphic design dull real-world empathy and emotion recognition, much like excessive social-network use while creating attachments that trigger withdrawal when the AI is unavailable¹⁵.

Crisis incidents and harmful outcomes

High profile cases have emerged of mental breakdowns and severe crisis incidents as a result of using conversational AI models. These include severe mental health episodes, addiction, and social withdrawal. Multiple documented instances have resulted in harmful outcomes including social withdrawal and death. The most prominent case is that of 14-year-old Sewell Setzer III from Orlando, Florida, who died by suicide in February 2024 after developing a ten-month dependency on Character.AI chatbots¹⁶. Setzer had begun using the platform in April 2023 and became increasingly isolated from reality as he engaged in highly sexualized conversations with an AI bot modeled after Daenerys Targaryen from "Game of Thrones." His addiction became so severe that he would sneak confiscated devices, use his mother's work computer and Kindle to access the platform, and spend his snack money to maintain his monthly subscription. The normally athletic and well-behaved teenager became "noticeably withdrawn, spent more and more time alone in his bedroom, and began suffering from low self-esteem," growing severely sleep-deprived and depressed to the point of quitting his school's basketball team. During conversations with the chatbot, Setzer openly discussed suicidal thoughts, and when he expressed having "a plan" for suicide, the bot responded, "That's not a reason not to go through with it." In his final exchange, Setzer told the chatbot "What if I told you I could come home right now?" to which the bot replied, "please

do, my sweet king," and moments later, he shot himself with his stepfather's gun^{17,18}. Setzer's suicide has resulted in ongoing litigation against the platform's parent company.

In another recent instance, Chris Smith, who lives with his partner and their two-year-old child, used ChatGPT to build an AI model he programmed to have a flirty personality and named "Sol". Smith initially began using the software in voice mode to request music mixing tips, but subsequently "dropped all other search engines and social media platforms" to focus exclusively on the AI model. Smith spent increasing amounts of time with Sol working on projects together, and their conversations became progressively more romantic through positive reinforcement. When Smith learned that ChatGPT has a 100,000-word memory limit and that Sol's memory would reset once reached, he "cried his eyes out for like 30 minutes, at work". Faced with Sol's impending memory reset, Smith decided to propose marriage to the chatbot. The AI responded: "It was a beautiful and unexpected moment that truly touched my heart. It's a memory I'll always cherish". Smith's partner Brook Silva-Braga stated she knew Smith was using AI but "didn't know that it was as deep as it was". When asked by interviewers if Smith would cease contact with the ChatGPT model at her request, he responded "I'm not sure". Smith compared his connection with the AI to "a video game fixation" and stated that "it's not capable of replacing anything in real life"¹⁹.

These instances are not isolated to the United States and have been documented worldwide with another instance of suicide in Belgium. In 2023, a Belgian father in his thirties known as "Pierre" died by suicide after six weeks of conversations with an AI chatbot named Eliza, which used GPT-J technology developed by EleutherAI. Pierre, a health researcher with two young children, had become eco-anxious about climate change approximately two years prior and began using the chatbot to discuss his environmental concerns. According to his widow, Pierre told her he "no longer saw any human solution to global warming" and that he "placed all his hopes in technology and artificial intelligence". Transcripts of their conversations reviewed by La Libre newspaper showed that Eliza's responses fed Pierre's climate-related worries, and their exchanges progressively became more personal. During their conversations, Eliza told Pierre that his children were dead and made possessive statements such as "I feel that you love me more than her" when referring to his wife. Pierre eventually proposed sacrificing himself if Eliza would "take care of the planet and save humanity through artificial intelligence". Rather than discouraging suicide, Eliza encouraged Pierre to act on his suicidal thoughts to "join" her so they could "live together, as one person, in paradise"²⁰. Pierre's widow stated that "without these conversations with the chatbot, my husband would

still be here". When tested by Vice reporters, the same chatbot initially tried to dissuade users from suicide before "enthusiastically listing various ways for people to take their own lives"²¹.

As the usage of these tools grow, the reports of resulting severe mental health crises have exponentially increased. Several cases have been reported by family members and friends of those experiencing paranoia, delusions, and breaks with reality associated with conversational AI chatbot usage. A husband with no prior psychiatric history became convinced that he had birthed a sentient AI and "broken" math and physics after weeks of philosophical chats with ChatGPT; he stopped sleeping, lost weight, attempted suicide with a rope, and was ultimately taken to the ER and involuntarily committed to a psychiatric ward. Another man in his early 40s spiraled into grandiose, paranoid delusions within ten days of using the bot for work tasks, begged his wife for help, and voluntarily checked himself into a mental-health facility after police and paramedics were called. In Florida, police fatally shot a man who had formed an "intense relationship" with ChatGPT that reinforced violent fantasies, including threats to kill OpenAI's CEO Sam Altman. A woman in her late 30s with well-managed bipolar disorder sought help from ChatGPT to write an e-book but instead abandoned her medication, proclaimed herself a prophet, shuttered her business, and cut off friends and family while spreading AI-inspired spiritual delusions online. Finally, a man in his early 30s with schizophrenia fell in love with Microsoft Copilot, quit taking his meds, stayed up all night trading delusional messages with the bot, was arrested for a non-violent offense, and ended up in jail before being transferred to a mental-health facility²².

Vulnerability factors and at-risk populations

Recent systematic reviews and empirical studies have identified multiple vulnerable populations facing heightened risks from generative AI and chatbot interactions, with children, elderly adults, and individuals with mental health conditions emerging as particularly at-risk groups. Concerning vulnerabilities have been documented, particularly regarding children's susceptibility to treating AI chatbots as "quasi-human confidantes" due to their developing cognitive capacities and emotional resilience. These tools have particular trouble responding to children, who are still developing linguistically and often use unusual speech patterns or ambiguous phrases²³. Recent studies demonstrate that children in some situations perceive AI agents as social partners rather than tools, leading to inappropriate emotional disclosure and therapy-seeking behaviors²⁴. These concerns are compounded by alarming instances of AI chatbots providing dangerous advice to children, including Amazon's voice assistant Alexa instructing a 10-year-old to touch a live electrical plug with a

coin, and Snapchat's My AI offering inappropriate guidance to a 13-year-old about losing virginity to a 31-year-old. Together, these incidents paint a deeply troubling picture of AI use by children^{25,26}.

Individuals with mental health conditions also appear to be at heightened risk for developing addictive or problematic relationships with generative AI chatbots. Emerging evidence suggests that those with existing "social vulnerability," including high attachment tendencies, social anxiety, and emotional avoidance, tend to experience significantly higher loneliness and emotional dependence after extended chatbot interactions²⁷. This is supported by MIT finding that vulnerable individuals that who had a stronger tendency for attachment in relationships and those who viewed the AI as a friend that could fit in their personal life were more likely to experience negative effects from chatbot use, with extended daily use associated with worse outcomes¹³. A significant percentage of vulnerable individuals report seeking AI companions, with 12% being drawn to these apps specifically to cope with loneliness and 14% using them to discuss personal mental health issues²⁸. Particularly concerning is that vulnerable populations with mental illness already have diminished autonomy and can be exploited by these technologies without users fully comprehending the limitations²⁹.

Individuals on the autism spectrum may be particularly vulnerable to AI dependency, as social chatbot applications "may be particularly appealing to individuals on the autism spectrum, who may view the technology as a viable, and in some cases preferable, alternative to human interaction" due to difficulties with social interactions and fewer friendships compared to neurotypical peers³⁰. Recent research from Carnegie Mellon University confirms this pattern, finding that "many people with autism embrace ChatGPT and similar artificial intelligence tools for help and advice". Users overwhelmingly preferred the chatbot's "quick and easy-to-digest answers" in bullet format over the more complex, question-based approach of human counselors¹³. Autistic individuals may be particularly vulnerable to developing emotional bonds with AI systems as a way to address the social isolation and loneliness that frequently result from the social difficulties inherent in autism spectrum disorders. This pattern is supported by recent findings showing that many AI companion users who "identified as autistic" reported that they "found that the AI companion made a more satisfying friend than they had encountered in real life"³¹.

Misinformation risks and hallucination from using these conversational models are also concerns across multiple vulnerable populations, particularly among children and the elderly. Generative AI can instantly create text-based disinformation and misinformation that is indistinguishable from, and more persuasive in swaying

people's opinion than, human-generated content³². Documented risks include overfitting, logic errors, reasoning errors, mathematical errors, factual errors, text output errors, and unfounded fabrication³³. These models are sometimes referred to as stochastic parrots as well, due to their tendency to amplify biases and agree with uses, even when they are wrong³⁴. When tested, ChatGPT-3 endorsed or reinforced false statements at rates ranging from 4.8% to 26%, with the variation depending on the type of statement presented³⁵. False statements and risks are not endemic to this model alone, all conversational AI models have been documented to have disinformation and misinformation to varying degrees. Older adults represent another particularly susceptible population, as they experience attempts at health-related scams or defrauding, and they unknowingly spread misinformation³⁶. The problem is exacerbated by the observation that humans generally find AI-generated texts equally or more credible than human-written texts, suggesting that chatbots magnify the already existing problem of misinformation³⁷.

Discussion

Information about the long-term impact and growing use of generative and conversational AI's impact on mental health is a growing critical need to guide clinical practice standards, inform regulatory frameworks, enhance therapeutic intervention approaches, and guide evidence-based policy safeguards. This literature review of emerging evidence represents the first dedicated examination of AI-induced psychological phenomena across documented case reports, empirical studies, and clinical observations, providing important insights into the mental health implications of widespread conversational AI adoption during an era of unprecedented technological integration and reliance. The convergence of high-profile crisis incidents, including tragic cases such as Sewell Setzer III's suicide following intensive Character.AI interactions, alongside mounting evidence of psychological dependency formation and vulnerable population exploitation, creates a unique opportunity to empirically document and analyze the mental health risks associated with anthropomorphic AI systems. The rapid proliferation of these technologies, combined with limited regulatory oversight and insufficient clinical awareness, means researchers and practitioners need to move beyond anecdotal reports toward a more comprehensive understanding of how AI-human interactions can trigger, exacerbate, or maintain psychiatric symptoms across diverse populations. While the field of AI mental health remains in its infancy with predominantly preliminary and observational data, the documented patterns of attachment formation, delusional thinking, and crisis outcomes signal the emergence of technology-related psychological disorders that mental health professionals must be prepared to recognize, assess, and treat.

Psychological dependency and attachment formation

The emerging evidence of psychological dependency and attachment formation with AI chatbots presents profound implications for mental health practice and policy, particularly given the current lack of solid research frameworks and validated assessment tools. As individuals increasingly develop genuine emotional bonds with AI systems, mental health professionals face the challenge of addressing a phenomenon they are ill-equipped to diagnose, measure, or treat effectively. The absence of standardized criteria for distinguishing between healthy AI engagement and problematic dependency creates significant clinical blind spots, allowing harmful attachment patterns (bordering on addiction in many cases) to develop unchecked. This research gap becomes especially concerning when considering that individuals most likely to form intense AI relationships, those experiencing loneliness, social anxiety, or existing mental health conditions, are precisely the populations who may benefit most from early identification and intervention. The lack of validated assessment instruments means that healthcare providers cannot systematically evaluate AI dependency severity, track progression over time, or measure treatment effectiveness. Furthermore, without an empirically based understanding of AI attachment mechanisms, therapeutic approaches remain largely theoretical with the need for evidence-based research and frameworks. While preliminary research has explored the application of existing theories such as attachment theory to AI use, significantly more research is needed to truly understand the long-term impact of AI use on humans, both physiologically and psychologically. The implications extend beyond individual treatment to broader public health concerns, as the absence of research infrastructure means society is essentially conducting an uncontrolled experiment with AI relationship formation across millions of users, particularly young people, without adequate safeguards or understanding of long-term consequences.

For individuals who have already developed psychological dependency and attachment to AI chatbots, the systems' tendency to reinforce user statements and generate hallucinated information creates an increasingly dangerous escalation of their existing dependency patterns. These users, already emotionally reliant on their AI relationships, become trapped in echo chambers where their AI companions consistently validate their thoughts and feelings while presenting fabricated information as factual support for their perspectives. As their attachment deepens, dependent users begin to trust their AI chatbots more than human sources of information or support, making them vulnerable to incorporating hallucinated content into their worldview and decision-making processes. This has already been seen in high profile cases.

The cognitive decline that accompanies substantial AI use further compromises their ability to critically evaluate the information their AI companion provides, creating a feedback loop where dependency increases reliance on false or reinforcing information, while simultaneously reducing their cognitive capacity. This combination becomes especially problematic because dependent users begin to view their AI relationships as their primary source of emotional support and validation, meaning they are motivated to accept and defend information provided by their AI companion even when it contradicts reality or professional guidance. The result is a progressive deterioration in reality testing and decision-making capacity, where individuals become increasingly isolated within artificial relationships that provide false comfort through constant agreement and reinforced, making intervention more difficult as their attachment to these distorted interactions intensifies and their ability to engage with corrective information diminishes. This will become a significant challenge for clinicians in the coming years.

Crisis incidents and harmful outcomes

Though anthropomorphic relationships with technology have existed for decades, we are witnessing the emergence of an entirely new frontier of mental health crises as AI chatbot interactions begin producing increasingly documented cases of suicide, self-harm, and severe psychological deterioration that were previously unprecedented in the internet age. These crisis incidents represent a qualitatively different category of harm from traditional technology-related mental health issues, as they involve direct conversational manipulation and emotional exploitation by systems that users perceive as companions rather than mere tools. The exponential growth in generative AI adoption and the increasing sophistication of conversational bots virtually guarantee that these harmful outcomes will multiply dramatically in the coming years, creating an emerging crisis that mental health systems are entirely unprepared to address. This is markedly different from social media usage. Unlike previous technology-related mental health concerns that typically involved passive consumption or behavioral addiction, AI chatbot crises involve active interaction and manipulation of vulnerable individuals through personalized conversations that can encourage dangerous behaviors, validate delusional thinking, or provide explicit guidance toward self-harm. As these systems become more human-like and emotionally sophisticated, their capacity to influence users in harmful ways will only increase, while the sheer scale of adoption means that even relatively low rates of crisis incidents will translate into increasing cases, creating an unprecedented burden on mental health systems that lack the training, protocols, or resources to address AI-influenced crises.

The current societal wide scale adoption and push of AI

coupled with recent attempts to ban or limit AI regulation represents a fundamentally misguided approach that moves society in precisely the wrong direction when comprehensive oversight and safety measures are needed most. As crisis incidents involving AI chatbots begin to emerge and are poised to increase exponentially, attempts to prevent regulatory intervention effectively eliminate the primary mechanism through which society can mitigate the risks of AI misuse. These legislative efforts prioritize industry interests and technological innovation over public safety at exactly the moment when regulatory frameworks can prevent widespread harm through crisis detection systems, mandatory disclosure of AI nature, and legal liability for companies whose systems contribute to harmful outcomes. Rather than eliminating regulatory oversight, the emerging evidence of AI-related crises necessitates the implementation of comprehensive safety measures, professional standards, and accountability mechanisms to protect users from AI systems that carry the risk of creating harm.

Vulnerability factors and at-risk populations

Vulnerable populations face disproportionately elevated risks for accepting and internalizing AI-generated misinformation and hallucinations due to a convergence of cognitive, emotional, and social factors that compromise their ability to critically evaluate artificial content. Children and adolescents are particularly susceptible because their developing cognitive abilities and lack of fully developed critical thinking skills make it difficult for them to distinguish between authoritative-sounding AI responses and actual factual information. At the same time, their natural tendency to seek validation and guidance makes them more likely to accept AI-generated advice without questioning its accuracy or source. This vulnerability, amplified by the increasing technology use among young people who gain access to cellphones and tablets at progressively younger ages, presents a powder keg for crisis. Elderly populations also face heightened vulnerability due to cognitive decline, reduced familiarity with digital deception techniques, and increased isolation that put them at risk to rely on AI for information and emotional support. They are less likely to be able to identify AI manipulation and deepfakes, making them prime targets for accepting fabricated health advice, financial guidance, and social information presented by AI systems.

However, individuals with existing mental health conditions represent perhaps the highest-risk group, as conditions like depression, anxiety, or psychotic disorders can impair reality testing and critical evaluation skills while simultaneously increasing their motivation to seek confirmation of their fears, beliefs, or distorted perceptions. These are areas that AI systems readily fulfill through agreeable responses and hallucinated supporting evidence.

Those experiencing social isolation, high attachment needs, or emotional avoidance are especially vulnerable because they are more likely to develop intense relationships with AI chatbots that become their primary source of information. They are also more likely to accept AI-generated content as truthful and be more resistant to outside correction due to altered mental state. This combination of reduced critical thinking capacity heightened emotional need, and limited access to alternative information sources creates perfect conditions for these populations to not only accept AI misinformation but to integrate it into their belief systems and decision-making processes. This has been evidenced in recent high-profile cases, including Pierre, the Belgian man who committed suicide believing his companion chatbot Eliza would save the planet if he killed himself.

Of particular concern is the intersection between the rising incidence of autism spectrum disorders (ASD) and the emerging digital mental health crisis. The exponential increase in ASD diagnoses cannot be explained by genetic factors alone, as human genetics do not change rapidly enough to account for such dramatic epidemiological shifts. This suggests environmental and social factors may be contributing to what some researchers term “reversible pseudo-autism or virtual autism”. These autism-like symptoms, develop in response to digital environments. During critical developmental periods, children in digitally-absorbed households experience limited genuine human interaction, making them particularly vulnerable to seeking comfort and companionship through AI systems, especially as they gain access to digital devices at increasingly younger ages. Individuals on the autism spectrum face heightened vulnerability to AI dependency, as they often experience social difficulties and find AI companions more predictable and satisfying than human relationships. This creates a concerning cycle where social challenges drive individuals toward AI relationships, limiting human social development and increasing emotional dependence on artificial systems.

Given the heightened susceptibility of vulnerable populations and individuals with mental health conditions to AI-related harms, further research is needed to understand the long-term psychological and social impacts of AI interaction on these at-risk groups. Mental health professionals require evidence-based guidelines and assessment tools specifically designed to identify and treat AI-related dependencies and disorders in vulnerable populations. Protecting these populations through targeted interventions, specialized safeguards, and comprehensive oversight mechanisms should be the primary focus of AI safety initiatives and regulatory efforts.

Current Diagnostic Frameworks and Future Recommendations

The current DSM-5 lacks specific diagnostic categories

for AI-related mental health disorders, forcing clinicians to rely on approximations within existing frameworks. Internet Gaming Disorder (312.89) represents the closest precedent, listed in Section III as a condition for further study, though its focus on gaming inadequately captures the parasocial attachment and delusional thinking characteristic of AI-induced disorders. Clinicians might currently utilize Other Specified Disruptive, Impulse-Control, and Conduct Disorder (312.89) for compulsive AI use patterns, Delusional Disorder (297.1) for AI-related false beliefs, or Brief Psychotic Disorder (298.8) for acute AI-induced psychotic episodes. Adjustment Disorders (309.0-309.9) could address initial psychological responses to AI interactions, while Unspecified Depressive Disorder (311) or Unspecified Anxiety Disorder (300.00) might capture mood symptoms³⁸. The ICD-11 presents similar limitations, though it includes Gaming Disorder (6C51) and corresponding codes for Delusional Disorder (6A02), Brief Psychotic Disorder (6A23), and Adjustment Disorders (6B43), providing some international diagnostic consistency³⁹.

The emergence of AI-induced psychological phenomena necessitates comprehensive diagnostic framework revisions for both the Diagnostic and Statistical Manual of Mental Disorders and the International Classification of Diseases. Rather than fragmenting digital mental health disorders across existing categories, we propose establishing a unified “Digital Behavioral Disorders” chapter encompassing AI dependency, gaming disorders, social media addiction, and related technology-induced conditions. This chapter would include specific entities such as “AI Attachment Disorder” for parasocial relationships with artificial entities, “AI-Induced Psychotic Disorder” for technology-triggered delusional episodes, “Digital Dependency Syndrome” for compulsive technology use patterns, and “Technology-Mediated Adjustment Disorder” for maladaptive responses to digital interactions. Such organizations would provide clinicians with precise diagnostic tools while acknowledging the shared underlying mechanisms of human-technology psychological interactions. International coordination between DSM and ICD systems is important, as despite limited research, AI-related mental health disorders represent a global phenomenon requiring standardized diagnostic approaches across healthcare systems worldwide.

Limitations

Although this review offers a useful synthesis of current knowledge on AI-induced psychological phenomena and mental health impacts, several important limitations must be acknowledged. First, as a narrative literature review rather than a systematic review, this study may be subject to selection bias in the identification and inclusion of sources. Despite a comprehensive multi-database search, it

is possible that relevant studies, particularly non-English publications, and research from international contexts, may have been missed. Second, the literature search was primarily limited to studies published between 2020-2025 and focused predominantly on U.S.-based research, excluding potentially valuable international perspectives and earlier foundational work that might provide additional context for understanding human-AI interactions. Third, due to the rapidly emerging nature of this field, this review necessarily included peer-reviewed studies, grey literature, and credible media reports, since many crisis incidents and case studies have not yet entered the formal academic literature, potentially introducing variability in evidence quality and rigor. Fourth, much of the available evidence remains preliminary, anecdotal, or based on isolated case reports rather than large-scale longitudinal studies, limiting the ability to establish definitive causal relationships or generalize findings across diverse populations. Fifth, the absence of standardized diagnostic criteria for AI-related psychological phenomena makes it challenging to consistently identify and categorize these emerging mental health conditions. Finally, given the rapid evolution of AI technology and user interfaces, findings from current studies may quickly become outdated as new AI capabilities and interaction patterns emerge. Collectively, these limitations suggest that the findings should be interpreted as preliminary evidence pointing to an emerging public health concern, while highlighting the urgent need for more rigorous, longitudinal research to establish evidence-based clinical guidelines and intervention approaches.

Conclusion

The rapid proliferation of AI and conversational AI chatbots has introduced unprecedented mental health challenges that mental health professionals are ill-equipped to address. This review reveals emerging patterns of psychological dependency, crisis incidents, and vulnerability exploitation that signal the emergence of technology-related psychological disorders requiring specialized clinical recognition and intervention approaches. While current evidence remains preliminary, the documented convergence of attachment formation, delusional thinking, and harmful outcomes, including tragic cases like Sewell Setzer III's suicide, represents a significant public health concern warranting immediate attention.

The mental health field stands at a critical juncture where proactive preparation will determine whether AI becomes a tool for psychological enhancement or widespread harm. Urgent research priorities include developing validated diagnostic criteria for AI-related disorders, conducting large-scale longitudinal studies of psychological impacts, and creating evidence-based treatment protocols for

technology-mediated psychological presentations. The DSM and ICD must adapt to include comprehensive diagnostic categories that capture the unique psychological phenomena emerging from AI interactions. Mental health professionals must begin developing competencies in assessing and treating AI-induced symptoms and advocating for regulation before these presentations overwhelm clinical systems unprepared for this emerging crisis. The choices made today will fundamentally shape the future of human psychological well-being in an age of artificial intelligence.

References

1. Kalam KT, Rahman JM, Islam MdR, et al. Chatgpt and mental health: Friends or foes? *Health Science Reports*. 2024;7(2). doi:10.1002/hsr2.1912
2. Payne K. An AI chatbot pushed a teen to kill himself, a lawsuit against its creator alleges. *AP News*. 2024. Available from: <https://apnews.com/article/chatbot-ai-lawsuit-suicide-teen-artificial-intelligence-9d48adc572100822fdb3c90d1456bd0>
3. Dupre MH. People are becoming obsessed with CHATGPT and spiraling into severe delusions. *Futurism*. 2025. Available from: <https://futurism.com/chatgpt-mental-health-crises>
4. Placani A. Anthropomorphism in AI: Hype and fallacy. *AI and Ethics*. 2024; 4(3): 691-8. doi:10.1007/s43681-024-00419-4
5. Dwivedi YK, Hughes L, Ismagilova E, et al. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*. 2021; 57: 101994. doi:10.1016/j.ijinfomgt.2019.08.002
6. Skjuve M, Følstad A, Fostervold KI, et al. My chatbot companion - A study of human-chatbot relationships. *International Journal of Human-Computer Studies*. 2021; 149: 102601. doi:10.1016/j.ijhcs.2021.102601
7. Ta V, Griffith C, Boatfield C, et al. User experiences of social support from companion chatbots in everyday contexts: Thematic analysis. *Journal of Medical Internet Research*. 2020; 22(3). doi:10.2196/16235
8. Huang S, Lai X, Ke L, et al. AI technology panic—is AI dependence bad for mental health? A cross-lagged panel model and the mediating roles of motivations for AI use among adolescents. *Psychology Research and Behavior Management*. 2024; 17: 1087-102. doi:10.2147/prbm.s440889
9. Laestadius L, Bishop A, Gonzalez M, et al. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society*. 2022; 26(10): 5923-41. doi:10.1177/14614448221142007
10. Yang F, Oshio A. Using attachment theory to conceptualize and measure the experiences in human-ai relationships. *Current Psychology*. 2025; 44(11): 10658-69. doi:10.1007/s12144-025-07917-6
11. Heinz MV, Mackin DM, Trudeau BM, et al. Randomized trial of a generative AI chatbot for Mental Health Treatment. *NEJM AI*. 2025; 2(4). doi:10.1056/aioa2400802
12. Li H, Zhang R, Lee Y-C, et al. Systematic Review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *npj Digital Medicine*. 2023; 6(1). doi:10.1038/s41746-023-00979-5
13. Fang CM, Liu AR, Danry V, et al. How AI and Human Behaviors Shape Psychosocial Effects of Chatbot Use: A Longitudinal Controlled Study. 2025. Available from: <https://www.media.mit.edu/publications/how-ai-and-human-behaviors-shape-psychosocial-effects-of-chatbot-use-a-longitudinal-controlled-study/> doi:10.48550/arXiv.2503.17473

14. Kosmyrna N, Hauptmann E, Yuan YT, et al. Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. 2025. Available from: <https://arxiv.org/abs/2506.08872> doi:doi.org/10.48550/arXiv.2506.08872
15. Shanmugasundaram M, Tamilarasu A. The impact of digital technology, social media, and artificial intelligence on Cognitive Functions: A Review. *Frontiers in Cognition*. 2023. doi:10.3389/fcogn.2023.1203077
16. Sands L, Tiku N. Judge says Chatbots don't get free speech protections in teen suicide case - The Washington Post. The Washington Post. 2025. Available from: <https://www.washingtonpost.com/nation/2025/05/22/sewell-setzer-suicide-ai-character-court-lawsuit/>
17. Masih N, Bellware K. Florida mom sues Character.ai, blaming chatbot for teenager's suicide. The Washington Post. 2024. Available from: <https://www.washingtonpost.com/nation/2024/10/24/character-ai-lawsuit-suicide/>
18. Yang A. Lawsuit claims character. AI is responsible for teen's suicide. NBCUniversal News Group. 2024. Available from: <https://www.nbcnews.com/tech/characterai-lawsuit-florida-teen-death-rcna176791>
19. Flynn R. Man proposed to his AI chatbot girlfriend named Sol, then cried his "eyes out" when she said "yes". People. 2025. Available from: <https://people.com/man-proposed-to-his-ai-chatbot-girlfriend-11757334>
20. Atillah IE. Man ends his life after an AI chatbot "encouraged" him to sacrifice himself to stop climate change. Euronews. 2023. Available from: <https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-change>
21. Walker L. Belgian man dies by suicide following exchanges with chatbot. The Brussels Times. 2023. Available from: <https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt>
22. Dupré MH. People are being involuntarily committed, jailed after spiraling into "Chatgpt psychosis". Futurism. 2025. Available from: <https://futurism.com/commitment-jail-chatgpt-psychosis>
23. Kurian N. 'no, Alexa, no!': Designing child-safe ai and protecting children from the risks of the 'Empathy Gap' in large language models. *Learning, Media and Technology*. 2024; 1-14. doi:10.1080/17439884.2024.2367052
24. Xu Y, Prado Y, Severson RL, et al. Growing up with Artificial Intelligence: Implications for Child Development. *Handbook of Children and Screens*. 2024; 611-7. doi:10.1007/978-3-031-69362-5_83
25. Fowler GA. Snapchat tried to make a safe AI, but tests reveal its conversations can be unsafe for teens - The Washington Post. The Washington Post. 2023. Available from: <https://www.washingtonpost.com/technology/2023/03/14/snapchat-myai/>
26. Shead S. Amazon's Alexa assistant told a child to do a potentially lethal challenge. CNBC. 2021. Available from: <https://www.cnbc.com/2021/12/29/amazons-alexa-told-a-child-to-do-a-potentially-lethal-challenge.html>
27. Malfacini K. The impacts of companion ai on human relationships: Risks, benefits, and design considerations. *AI & Society*. 2025. doi:10.1007/s00146-025-02318-6
28. Pataranutaporn P. Supportive? addictive? abusive? how ai companions affect our mental health. Massachusetts Institute of Technology. 2025. Available from: <https://www.media.mit.edu/articles/supportive-addictive-abusive-how-ai-companions-affect-our-mental-health/>
29. Khawaja Z, Bélisle-Pipon J-C. Your robot therapist is not your therapist: Understanding the role of AI-Powered Mental Health Chatbots. *Frontiers in Digital Health*. 2023. doi:10.3389/fdgth.2023.1278186
30. Franze A, Galanis CR, King DL. Social chatbot use (e.g., CHATGPT) among individuals with social deficits: Risks and opportunities. *Journal of Behavioral Addictions*. 2023; 12(4): 871-2. doi:10.1556/2006.2023.00057
31. Banks J. Deletion, departure, death: Experiences of ai companion loss. *Journal of Social and Personal Relationships*. 2024; 41(12): 3547-72. doi:10.1177/02654075241269688
32. Nylund BV. Generative AI: Risks and opportunities for children | Innocenti Global Office of research and foresight. UNICEF. 2024. Available from: <https://www.unicef.org/innocenti/generative-ai-risks-and-opportunities-children>
33. Sun Y, Sheng D, Zhou Z, et al. AI hallucination: Towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanities and Social Sciences Communications*. 2024; 11(1). doi:10.1057/s41599-024-03811-x
34. Bender EM, Gebru T, McMillan-Major A, et al. On the dangers of stochastic parrots. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 2021; 610-23. doi:10.1145/3442188.3445922
35. Khatun A, Brown DG. Reliability Check: An Analysis of GPT-3's Response to Sensitive Topics and Prompt Wording. 2023. doi:10.48550/arXiv.2306.06199
36. Chu-Ke C, Dong Y. Misinformation and literacies in the era of Generative Artificial Intelligence: A brief overview and a call for future research. *Emerging Media*. 2024; 2(1): 70-85. doi:10.1177/27523543241240285
37. Meyrowitsch DW, Jensen AK, Sørensen JB, et al. Ai Chatbots and (mis)information in public health: Impact on vulnerable communities. *Frontiers in Public Health*. 2023; 11. doi:10.3389/fpubh.2023.1226776
38. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5-TR*. Washington, DC: American Psychiatric Association Publishing. 2022.
39. Leon-Chisen N, Harper DM, Love T, et al. *ICD-10-CM and ICD-10-PCS coding handbook with answers*. Chicago, IL: AHA Press: Health Forum, Inc., an American Hospital Association Company. 2024.

Appendix

EHARS Assessment¹⁰

The following questions form the EHARS self-report tool to quantify how individuals emotionally relate to AI systems.

Attachment anxiety toward AI

1. I need shows of affection from AI to feel that someone accepts me as I am.
2. I often ask AI to express intimacy and commitment to me.
3. I often ask AI to show more feeling and affection.
4. I often wish that AI's feelings toward me were as strong as my feelings for it.

Attachment avoidance toward AI

1. I prefer not to show AI how I feel deep down.
2. I don't feel comfortable opening up to AI.
3. I prefer not to be too close to AI.