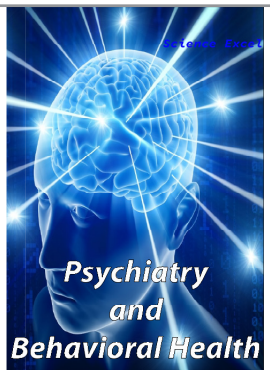
Psychiatry and Behavioral Health

Original Article



The AI Companion in the Therapy Room: When Generative AI Becomes Part of Coping, Attachment, and Recovery

Dr. Zoe Wyatt

Promises Healthcare, Singapore

Correspondence

Dr Zoe Wyatt

Promises Healthcare, Singapore

- Received Date: 10 Aug 2026
- Accepted Date: 19 Aug 2026
- Publication Date: 21 Aug 2026

Keywords: generative artificial intelligence; human–AI attachment; addiction recovery; emotional regulation; conversational AI

Abstract

Generative artificial intelligence is increasingly being used for emotional support, self-reflection, reassurance and behaviour change, including by people navigating addiction and recovery. As conversational AI becomes more relational in tone and continuously available, its role may extend beyond information-seeking into coping, attachment and emotional regulation. This critical narrative review examines emerging research on human–AI attachment, emotional reliance, reassurance-seeking and avoidance, alongside developing applications of generative AI within substance use and recovery contexts. Particular attention is given to the ways AI may support reflection, preparation for difficult conversations, recovery-oriented problem solving and access to immediate support, while also potentially reinforcing avoidance, compulsive reassurance-seeking or displacement of human relationships and professional care. The review considers these tensions through an attachment- and recovery-informed lens and highlights the importance of understanding not only what people use generative AI for, but the psychological function that the interaction serves. Current evidence remains early and heterogeneous, limiting firm conclusions about clinical benefit or harm. However, the growing integration of AI into everyday coping suggests that clinicians will increasingly encounter clients whose relationships with generative systems form part of their emotional and recovery environments. Clinical assessment, ethical guidance and future research should therefore consider patterns of use, relational reliance and the circumstances in which AI supports, complicates or substitutes for existing coping and connection.

Introduction

Generative artificial intelligence is increasingly entering clinical practice through the ways clients use conversational AI outside formal treatment. Clients may consult chatbots for information, emotional support, reassurance, decision-making, communication rehearsal, and companionship [1]. They may also bring AI-generated explanations, diagnostic language, and interpretations of relationships into sessions, where these responses can influence how concerns are framed and understood [2]. Conversational AI is therefore beginning to shape both between-session coping and the material that enters the therapeutic encounter [3]. This informal use is already substantial. In a survey of 500 Australian adults, 23.6% reported discussing mental health with conversational AI, with accessibility rated more favourably than professional support and most participants preferring access to both forms of support [4].

This development has particular relevance for addiction treatment. Craving, shame, loneliness, uncertainty, and interpersonal distress often arise outside scheduled appointments, when access to professional or peer support may be limited [5]. Generative AI offers immediate and highly accessible interaction during these periods. Its personalised and apparently responsive form may provide short-term support, while also becoming increasingly significant within a client's emotional regulation and relational world [1].

Research has begun to examine anthropomorphism, emotional attachment, loneliness, social substitution, problematic use, and the growing use of generative AI for informal mental health support [6-10]. Addiction-specific research remains comparatively limited and has focused largely on structured conversational interventions designed to provide psychoeducation, self-monitoring, motivational support, craving management, goal-setting, and links to care [11-13]. General-purpose and companion chatbots used informally during recovery have received far less direct investigation.

This critical narrative review brings together research on human–AI attachment, emotional regulation, reassurance-seeking, avoidance, and AI-supported substance-use treatment. Reflexive clinical observations are included to illustrate emerging practice issues without

Citation: Wyatt Z. The AI Companion in the Therapy Room: When Generative AI Becomes Part of Coping, Attachment, and Recovery. Psychiatry Behav Health. 2026;5(1):1-7..

being presented as empirical findings. The article examines how generative AI is becoming incorporated into clients' coping, relationships, and recovery, with particular attention to the conditions under which it extends support and the points at which it may reinforce avoidance, emotional reliance, or displacement of human connection.

From Digital Tool to Relational Technology

Generative AI differs from earlier digital mental health tools in the kind of interaction it makes possible [14]. Search engines, educational websites, and static applications can provide information or structured exercises, but they do not usually sustain an unfolding conversation that adapts to the language, concerns, and preferences of the user. Conversational AI can respond to personal disclosure in context-sensitive language and generate exchanges that users experience as validating, supportive, and socially engaging [9,16]. These features allow the system to become familiar in ways that exceed its informational function. User reviews of the mental health chatbot Wysa similarly emphasised real-time access, a trusting environment, predictable interaction, and the perceived humanness of the system [17].

The social quality of this interaction does not depend on users believing that the system is human. Research on human–computer interaction has long shown that people apply interpersonal expectations to technologies that display social cues, even when they remain fully aware that they are engaging with a machine [18]. Generative AI extends this tendency through fluent, context-sensitive, and increasingly personalised language. In two experiments involving 1,274 participants, Folk [6], found that individual differences in anthropomorphism were strongly associated with feelings of connection following chatbot interaction. The same exchange may therefore remain primarily instrumental for one user while acquiring a more socially meaningful quality for another.

The concept of relational technology helps capture this development. The term refers less to the technological status of the system than to the psychological qualities that emerge through repeated use [19]. A chatbot may initially be consulted for information, then become a familiar source of explanation, reassurance, or emotional organisation [19]. Over time, users may come to recognise a particular tone, anticipate how the system will respond, and return partly because the exchange itself has become predictable and personally resonant [1,20].

This familiarity is produced through an interaction between the user and the system. Repeated prompting, correction, and feedback allow users to shape the tone, depth, and style of the responses they receive [19]. The resulting interaction may feel increasingly individualised, even though the system's responses continue to reflect patterns extending beyond the particular user [21].

A brief reflexive example illustrates how this co-constructed familiarity can develop. Over several years of using conversational AI in clinical, academic, and writing contexts, I began referring to the system as "Chad". The name emerged from a recognisable interactional style: an American therapeutic idiom, a recurring preference for American research, and a polished form of clinical language that became increasingly familiar through repeated use. My corrections and preferences influenced the responses I received, while the system continued to reproduce linguistic and cultural tendencies derived from its wider training.

The nickname captures a stable conversational persona without implying that I regard the system as a person or experience the interaction as reciprocal. It provides shorthand for the particular version of the system that has developed within our exchanges. A user engaging primarily in engineering, coding, or technical problem-solving would likely encounter a different style, even while interacting with the same underlying model. This illustrates how conversational AI can feel individually familiar while retaining characteristics that originate beyond the individual relationship. Naming and personification also occur outside emotionally supportive contexts. In a survey of 445 university students, participants commonly applied social conventions to AI language models, including names, gendered pronouns, greetings, and expressions of thanks [22].

Naming an AI or describing it as having a recognisable style offers little basis, on its own, for inferring attachment, dependence, or impaired reality testing [1]. Greater relational investment becomes more clinically relevant when the system is experienced as a source of emotional closeness, continuity, understanding, or social substitution [8].

Recent research has begun to distinguish these forms of engagement more precisely. Kasturiratna and Hartanto [8] developed and validated a 15-item AI Attachment Scale across five studies involving 1,259 participants. The measure identified emotional closeness, social substitution, and normative regard as distinct dimensions of attachment [8]. Stronger attachment was associated with socioemotional motives for use, loneliness, social anxiety, and anxious attachment, as well as positive affect and life satisfaction [8]. Hu, et al. [7] add further detail, showing that personification, satisfaction with the interaction, interpersonal difficulties, and projective processes may all contribute to attachment formation. Together, these findings suggest that attachment develops through the interaction between the user's relational needs and the qualities they experience the system as providing [7,8].

The fluency of these exchanges can make the limits of the system less visible. Large language models depend on the information supplied to them and cannot independently observe non-verbal behaviour, gather undisclosed contextual information, or monitor changes across the client's wider life. Their responses are generated through statistical pattern recognition rather than clinical observation, contextual judgement, or responsibility for care [23]. Apparent understanding may nevertheless feel sufficiently convincing to acquire emotional and interpretive authority.

The Culturally Coded AI Companion

The familiarity that develops through repeated interaction can make an AI system appear increasingly specific to the individual user. This personalisation may obscure another influence on the interaction: the cultural assumptions already embedded within the model. Conversational AI is trained on language produced within particular social, linguistic, and institutional contexts. Its responses therefore carry tendencies concerning relationships, autonomy, distress, responsibility, and help-seeking, even when these assumptions are not made explicit.

Research increasingly challenges the idea that generative AI provides culturally neutral reasoning [24,25]. Lu et al. [15] found measurable cultural variation in both social orientation and cognitive style according to the language in which generative AI was used. When GPT responded in Chinese, its outputs were more interdependent, holistic, and sensitive to context; when responding in English, they were more independent and analytic [15]. Importantly, these patterns were also reproduced in

ERNIE, a generative AI model widely used in China, suggesting that culturally patterned responses are not limited to Western-developed systems [15].

The findings complicate the assumption that any single cultural orientation is inherent to conversational AI and instead indicate that language, model design, and training context can all shape the worldview expressed in an answer [15,24,25]. Other evaluations have nevertheless identified Western cultural preferences across multilingual models, including outputs generated in non-Western languages [24,25]. Linguistic fluency across cultures therefore does not guarantee cultural equivalence.

Mental health responses may be particularly sensitive to these tendencies because psychological interpretation depends heavily on cultural beliefs about the self, family, emotion, responsibility, and acceptable behaviour [15]. A response framed around individual autonomy, direct disclosure, or firm interpersonal boundaries may appear clinically coherent while fitting some cultural contexts better than others [26,27]. Concepts such as trauma, attachment, validation, codependence, narcissism, and toxicity can also be applied as though their everyday meanings are settled and universal [27].

The personalised quality of AI may increase the authority of these interpretations. A generic article presents itself as external information. A chatbot produces its explanation within an ongoing conversation, using the client's language and referring to details supplied moments or months earlier [24,25]. Cultural assumptions may therefore arrive in a form that feels individually recognised rather than externally imposed [24,25].

This distinction has implications for addiction recovery. Meanings attached to substance use, shame, dependence, family responsibility, spirituality, disclosure, and recovery vary across communities [28]. A system trained predominantly on Western psychological and medical literature may privilege individualised explanations and formal treatment models [29]. It may give less weight to collective identity, family interdependence, religious practice, culturally specific understandings of distress, or community-based recovery [29].

Evidence from psychiatric and medical applications shows that these concerns extend beyond tone. Bouguettaya, et al. found racial differences in treatment recommendations produced by four large language models when otherwise similar psychiatric cases included explicit or implied racial information [30]. A larger evaluation of medical decision-making similarly found that recommendations changed according to race, income, housing status, and LGBTQIA+ identity despite constant clinical details [31]. These studies do not directly examine addiction counselling, although they demonstrate that apparently personalised clinical responses can reproduce social bias.

Cultural coding may also shape the explanatory language clients bring into therapy. Generative AI is already used for psychoeducation, diagnosis-related questions, relational guidance, and decision-making, allowing its concepts and classifications to enter users' accounts of personal experience [9]. These interpretations may identify meaningful aspects of an experience while also framing it through cultural assumptions that remain largely invisible.

The therapist may therefore encounter the client's history through an interpretive framework partly shaped by the AI's cultural and linguistic assumptions. When that framework is delivered in language that feels natural, empathic, and highly specific, it may carry more authority than its evidentiary basis warrants [32]. Recognising this influence is important because the appeal of conversational AI lies partly in how convincingly it appears to understand the individual user.

Why AI Is Compelling in Addiction Recovery

Recovery is often negotiated through brief and highly consequential moments: whether to act on a craving, disclose a lapse, attend an appointment, contact a sponsor, or withdraw from support [33]. These moments do not necessarily coincide with scheduled care, and participants involved in the development of the Suzy recovery chatbot specifically identified the need for support with cravings, stressors, goals, appointments, and connection to care between treatment contacts [13]. Conversational AI is compelling partly because it can respond while the situation is still unfolding, when intention may be unstable and the next action remains undecided.

Suzy offers one of the clearest examples of this approach within addiction care. The chatbot was developed to support people with substance-use disorders between treatment contacts, with functions spanning craving management, goal-setting, appointment reminders, health information, resource referrals, and connection back to the care team [13]. Clinicians, researchers, health coaches, and people with lived experience contributed to its development, grounding the system in the practical realities of recovery rather than adapting a generic mental health tool after the fact [13].

The study did not evaluate treatment effectiveness, yet it provides a useful account of what participants wanted from an AI-supported recovery tool [13]. They valued approachable language, accountability, and access beyond normal service hours [13]. These preferences suggest that the perceived usefulness of a chatbot may depend on its capacity to translate broad recovery intentions into manageable actions: reviewing a goal, interrupting an urge, locating support, or preparing for the next treatment contact.

Suzy was also designed within a visible structure of care. Accuracy testing, escalation procedures, human oversight, and explicit limits on the system's role were incorporated into its development [13]. The chatbot's conversational flexibility was therefore linked to a defined clinical purpose and a pathway back to human support. Commercial companions and general-purpose systems may become involved in similar moments of vulnerability through ordinary use, although their role, safeguards, and relationship to treatment are usually less clearly established [11,34]. The broader idea of AI as a recovery companion has also been examined specifically in relation to people who use drugs, with attention to both its supportive potential and the boundaries required around its role [35].

The wider substance-use literature supports the value of conversational formats while showing how early the evidence remains [11,33,36]. A systematic review of 28 chatbot-assisted interventions identified applications in psychoeducation, motivational work, self-monitoring, craving management, goal-setting, behavioural feedback, and relapse prevention [12]. Personalised guidance and ease of access were frequently associated with engagement [12]. Many of the reviewed interventions were rule-based, however, and the studies varied considerably in design and methodological quality [12]. The findings offer stronger evidence for feasibility and user engagement than for lasting effects on substance use or recovery.

Users also appear to expect more than convenience or encouragement [37]. In a qualitative study of conversational

agents for alcohol self-help, participants valued privacy, practical guidance, and control over the interaction, while raising concerns about authenticity, trust, data security, and excessive agreeableness [37]. Their responses point to a more demanding standard of usefulness. A recovery chatbot must be able to remain supportive while recognising risk, tolerating complexity, and directing the user towards appropriate action [37]. Purpose-built systems such as Suzy offer one model for balancing responsiveness with accountability and connection to care [13]. General-purpose chatbots enter the same territory with greater conversational freedom and fewer agreed boundaries around what their support is intended to achieve [34]. This also raises unresolved questions about how systems should respond when serious risk is detected, particularly where escalation to human support may conflict with expectations of privacy, consent, and confidentiality.

Attachment and AI-Mediated Relief

Repeated use during loneliness, shame, uncertainty, or fear of relapse may allow the AI interaction itself to become associated with anticipated relief. Users seeking emotional support from ChatGPT have described validation, perceived care, and the freedom to disclose without burdening another person [9]. In recovery, these qualities may make the system especially appealing during periods when approaching another person feels difficult or exposing [8,15].

The term *attachment-like regulation* describes the use of AI as a recurring source of comfort, steadiness, or emotional organisation [38,39]. The term recognises attachment-related behaviour directed towards the system without treating the interaction as equivalent to reciprocal human attachment [38]. AI has no independent concern, intention, or responsibility for the user, yet it can still become a familiar point of return when distress rises or approaching another person feels difficult [39]. Individual attachment patterns may influence how central this regulatory role becomes. In a study of 54 conversational-AI users, Heng and Zhang [40] found that attachment anxiety was associated with problematic use both directly and indirectly through emotional attachment, with anthropomorphic tendencies moderating the relationship.

Satisfaction with the interaction and the expectations that develop around it may be especially important. When a chatbot repeatedly accompanies periods of distress, elements such as its tone and continuity may themselves become associated with relief [7]. This predictability may be particularly compelling for people who experience human support as uncertain, exposing, or demanding. The user can repeat a concern, revise what has been disclosed, reject an unwanted response, and end the exchange without managing another person's frustration, fatigue, or emotional reaction [7,8]. These features can lower the threshold for seeking support while also making the interaction unusually controllable.

The loss or removal of a familiar AI system may reveal how much emotional significance the interaction has acquired. Users affected by the shutdown of an AI companion described experiences resembling grief and attempted to preserve or recreate the relationship on other platforms [41]. Recent measures of human–AI attachment also include separation distress, reflecting discomfort associated with losing access to an established AI relationship [10,41]. The wider relational implications extend beyond individual attachment. Jones and Schonewille [42] argue that synthetic intimacy may alter the social practices through which human connection and community are sustained, particularly when convenience and simulation begin to displace reciprocal engagement.

Pseudo-Safety and the Deferral of Recovery Tasks

Attachment-like regulation may become clinically significant when relief increasingly depends on reassurance, repeated explanation, checking, or extended analysis. Golden and Aboujaoude [43] propose that the continuous availability and low-friction nature of conversational AI may facilitate repetitive reassurance-seeking, checking, "figuring out," perfectionistic prompting, and decision-seeking. These interactions can provide immediate relief or an impression of certainty, while repeated use may strengthen avoidance and reduce opportunities for corrective learning [43]. Similar mechanisms are well established in the safety-behaviour literature, in which reassurance and checking can reduce anxiety in the short term while increasing reliance on the behaviour and maintaining the underlying fear or uncertainty [44].

Drawing on this literature, this review uses the term pseudo-safety to describe relief attributed to the AI interaction while the distress, uncertainty, or recovery task that prompted it remains unresolved. Safety behaviours can become self-reinforcing when their immediate reduction of distress increases reliance on them as a means of coping [44]. Conversational AI may serve a comparable function when reassurance, repeated analysis, checking, or highly controlled interaction becomes increasingly necessary for the user to feel able to cope [43]. A person may repeatedly ask whether a lapse "counts", continue analysing another person's intentions, or refine a message without sending it. In each case, the interaction may provide temporary clarity or emotional relief while postponing disclosure, decision-making, interpersonal contact, or another action relevant to recovery [43,44].

Pseudo-safety can be distinguished conceptually from useful short-term regulation by examining what follows the interaction. Safety-behaviour models emphasise function and consequence rather than the behaviour in isolation, particularly whether short-term anxiety reduction prevents new learning or increases dependence on the strategy [44]. Within the present framework, emotional relief refers to an immediate reduction in distress following the AI interaction.

In this review, *recovery movement* refers to an exchange that supports action beyond the AI conversation, such as engagement with treatment or the care team, goal-directed behaviour, accountability, craving management, or connection with community and other human supports [13]. *Deferred action* describes the contrasting pattern in which consultation repeatedly produces relief while the underlying decision, disclosure, interpersonal contact, or recovery task remains postponed, consistent with models in which reassurance-seeking and repetitive consultation can maintain avoidance and interfere with corrective learning [43,44]. Recovery movement and deferred action are proposed here as clinically useful functional distinctions rather than empirically validated categories.

Communication rehearsal illustrates why function is important. Emerging research suggests that AI-mediated communication tools can support some autistic and neurodivergent users with interpreting conversational cues, generating alternative phrasing, practising social communication, and reducing some of the effort involved in preparing for interpersonal exchanges [45-47]. Such use may therefore be adaptive when rehearsal increases confidence or facilitates subsequent communication. The same behaviour could become avoidant

Table 1. *Potential functions and implications of conversational AI in addiction recovery*

Function	Possible contribution to recovery	Possible clinical concern	Current evidence
Information, interpretation, and planning	Psychoeducation, service navigation, recovery planning, and language for discussing psychological experiences	Inaccurate guidance, premature diagnostic certainty, or excessive authority assigned to fluent responses	Purpose-built recovery chatbots have shown feasibility for information, referrals, goal-setting, and care support, although effectiveness remains untested [13].
Reflection and disclosure	Organising thoughts, preparing for therapy, and lowering the threshold for discussing shame or stigma	Extended analysis without action, or clinically significant material remaining outside treatment	AI-generated responses can elicit greater self-disclosure and perceived closeness during chatbot interactions [16]. Effects on later disclosure to clinicians and treatment engagement remain unclear.
Regulation and craving support	Immediate coping prompts during craving, insomnia, distress, or gaps between appointments	Increasing reliance on AI or delayed contact with appropriate human or medical support	Greater daily use has been associated with emotional dependence, problematic use, loneliness, and lower socialisation, without establishing causality [10].
Companionship and relational continuity	Reduced acute loneliness and access to a familiar conversational presence	Emotional dependence, social substitution, relational narrowing, or distress following loss of access	Research has identified emotional closeness and social substitution as dimensions of AI attachment, while the loss of an established AI companion has been associated with grief-like responses [8,41].
Communication rehearsal	Preparation for difficult conversations, reduced cognitive load, and support with organising language	Repeated refinement that delays communication or increases reliance on highly controlled interaction	Emerging research suggests that AI may support communication preparation [45-47], although its effects on subsequent interpersonal engagement remain insufficiently studied.
Reassurance and certainty-seeking	Temporary stabilisation and clarification during uncertainty	Repetitive checking, short-lived relief, reduced uncertainty tolerance, and reinforcement of avoidance	Conceptual models propose that continuous access to reassurance and repeated analysis may maintain anxiety and interfere with corrective learning [43].

if repeated refinement increasingly postpones contact until the user feels sufficiently certain, regulated, or protected from misunderstanding, consistent with broader models of reassurance-seeking and safety behaviour [43,44]. AI-generated interpretations may operate similarly when repeated requests for explanation provide temporary certainty while narrowing exploration of alternative accounts or delaying action outside the AI interaction [43,44].

Preliminary longitudinal evidence supports closer examination of these patterns. In a four-week randomised study involving 981 participants and more than 300,000 messages, Fang et al. [10] found that higher daily chatbot use was associated with greater loneliness, emotional dependence, problematic use, and lower socialisation. Exploratory analyses also found that greater trust in the chatbot was associated with greater emotional dependence. The findings were correlational with respect to usage intensity and psychosocial outcomes and therefore do not establish causal direction [10]. Carter [48] similarly found that escapism was associated with greater problematic AI-companion use, while resilience attenuated the association between escapism and problematic use. Together, these findings suggest that the motivation and function underlying AI use may provide clinically relevant information beyond frequency of use alone [10,48].

Potential Functions of AI Within Addiction Recovery

The clinical meaning of conversational AI depends less on the function itself than on the role it comes to play within recovery. Table 1 applies this functional analysis across the principal ways conversational AI may enter addiction recovery.

Bringing AI Use Into Clinical Assessment

Conversational AI use can be incorporated into assessment by examining the interaction as part of the clinical sequence surrounding a relevant event. Golden and Aboujaoude [43] recommend direct inquiry into chatbot use and its involvement in emotional regulation, decision-making, and distress management. In addiction treatment, the therapist can ask what prompted the consultation, what the client entered into the system, how the response was understood, and what occurred afterwards [43].

Attention should also be given to the evidentiary basis of AI-generated conclusions. General-purpose chatbots may

provide reassurance, psychoeducation, and recommendations with less elicitation of contextual detail than human therapists [50]. When a client brings an AI-generated formulation into treatment, assessment can examine the information supplied to the system, the details omitted, the certainty of the response, and the influence it has had on the client's interpretation or choices [50]. Clinicians have also identified potential changes in communication, transparency, shared decision-making, and the therapeutic relationship when AI becomes involved in mental health care [51].

The client's relationship to the response may be as clinically informative as its content. Relevant indicators include the authority assigned to the system, willingness to consider alternative explanations, capacity to delay consultation, and the degree to which AI input has become necessary before decisions, disclosures, or interpersonal actions [50]. These features may help the therapist identify whether AI is functioning as an occasional resource or has acquired a more central role within the client's recovery and treatment [43].

This framework has yet to be validated in addiction populations. It offers a preliminary structure for examining conversational AI when its use has become clinically relevant to the client's interpretation, behaviour, or engagement with treatment.

Ethical and Evidentiary Limits

Research on emotional attachment, problematic use, and the psychosocial effects of conversational AI remains at an early stage. Much of the available evidence is cross-sectional, and the limited longitudinal research has identified associations among heavier use, emotional dependence, loneliness, and reduced socialisation without establishing the direction of these relationships [8,10].

Addiction-specific evidence is more limited. Existing studies have focused mainly on structured chatbots developed for psychoeducation, self-monitoring, craving support, motivational work, and connection to care. Findings support feasibility, acceptability, and engagement, while evidence concerning relapse, treatment retention, harm reduction, and sustained recovery remains sparse [12,13].

The application of attachment-like regulation and pseudo-safety to addiction recovery therefore remains conceptual. Research has yet to establish whether repeated AI consultation influences lapse disclosure, treatment participation, interpersonal support, craving management, or relapse trajectories over time. Longitudinal studies are also needed to distinguish the effects of purpose-built clinical systems from those of general-purpose chatbots and commercial AI companions.

Ethical concerns extend beyond clinical effectiveness. Users may disclose sensitive information to systems operating outside professional confidentiality and medical regulation. Fluent and empathic responses may also encourage levels of trust that exceed the system's capacity for contextual judgement, risk assessment, and responsibility for care [52]. Identified concerns include inaccurate crisis guidance, emotional dependence, ambiguous loss, anthropomorphic design, and changes in expectations of human care [52,53]. Experimental evidence from a simulated crisis self-rescue context also found that AI hallucination affected users' evaluations of the service, while access to human expert advice reduced some of its adverse effects [54].

These limitations support a cautious clinical approach. AI-generated material should be examined alongside the client's wider history, relationships, treatment context, and current level of risk. Strong conclusions about benefit, harm, dependence, or therapeutic value remain premature.

Conclusion

Conversational AI is becoming part of the emotional, relational, and behavioural environment in which addiction recovery takes place. Its significance extends beyond the information it provides. Through repeated use, it may shape how clients regulate distress, interpret experience, prepare for communication, and seek a sense of continuity or certainty. This critical narrative review has brought together research on AI-supported addiction interventions, human–AI attachment, reassurance, avoidance, and safety behaviour. It has extended discussion beyond purpose-built recovery chatbots to consider how general-purpose and companion AI may influence relationships, coping, and treatment. The review has proposed attachment-like regulation and pseudo-safety as clinical concepts and developed a functional framework for assessing AI use in addiction recovery. The central question is what the exchange makes more likely to happen next: greater engagement with treatment, accountability, and human connection, or further consultation before the client feels able to disclose, decide, or act.

References

1. Dai X, Leng LL, Liu Y, Huang YT, Wong DFK. The paradox of agency in psychotherapy: how people with mental distress experience support from generative AI chatbots and human therapists. BMC Psychiatry. 2026;26(1):49. doi:10.1186/s12888-025-07671-w
2. Barzkar F, Zaribaf A, Mirfazeli FS, Keshavarz-Akhlaghi AA. The machine as therapist: unpacking transference and emotional healing in AI-assisted therapy. J Contemp Psychother. 2025;55(4):361-368.
3. Tabasum A. The therapeutic alliance in AI-assisted psychotherapy: a clinical evaluation of human-AI interaction in mental health care. Glob Soc Sci Rev. 2025;10(3):321-335.
4. Franze A, Lushington K, Loetscher T. Conversational AI or a human professional for mental health advice? Exploring prevalence and public preferences in Australian adults. Aust J Psychol. 2026;78(1):2657279. doi:10.1080/00049530.2026.2657279
5. Miller WR, Forcehimes AA, Zweben A. Treating Addiction: A Guide for Professionals. Guilford Publications; 2019.
6. Folk D, Heine SJ, Dunn E. Individual differences in anthropomorphism help explain social connection to AI companions. Sci Rep. 2025;15:36548. doi:10.1038/s41598-025-19212-2
7. Hu D, Lan Y, Yan H, Chen CW. What makes you attached to social companion AI? A two-stage exploratory mixed-method study. Int J Inf Manage. 2025;83:102890. doi:10.1016/j.ijinfomgt.2025.102890
8. Kasturiratna KTAS, Hartanto A. Attachment to artificial intelligence: development of the AI Attachment Scale, construct validation, and the psychological mechanisms of human-AI attachment. Comput Human Behav Rep. 2026;21:100912. doi:10.1016/j.chbr.2025.100912
9. Luo X, Wang Z, Tilley JL, et al. Seeking emotional and mental health support from generative AI: mixed-methods study of ChatGPT user experiences. JMIR Ment Health. 2025;12:e69750.
10. Fang CM, Liu AR, Danry V, et al. How AI and human behaviors shape psychosocial effects of chatbot use: a longitudinal randomized controlled study. arXiv. Preprint published March 2025. doi:10.48550/arXiv.2503.17473

11. Ogilvie L, Prescott J, Carson J. The use of chatbots as supportive agents for people seeking help with substance use disorder: a systematic review. Eur Addict Res. 2022;28(6):405-418.
12. Lee S, Yoon J, Cho Y, Chun JS. A systematic review of chatbot-assisted interventions for substance use. Front Psychiatry. 2024;15:1456689. doi:10.3389/fpsyt.2024.1456689
13. Comulada WS, Swendeman D, Ho YX, et al. Development, feasibility, acceptability, and usability of an artificial intelligence-powered chatbot (Suzy) to support patients in substance use disorder recovery: multiphase study. JMIR Form Res. 2026;10:e84683. doi:10.2196/84683
14. Torous J, Topol EJ. Assessing generative artificial intelligence for mental health. Lancet. 2025;406(10504):683.
15. Lu JG, Song LL, Zhang LD. Cultural tendencies in generative AI. Nat Hum Behav. 2025;9:2360-2369. doi:10.1038/s41562-025-02242-1
16. Kleinert T, Waldschütz M, Blau J, Heinrichs M, Schiller B. AI outperforms humans in establishing interpersonal closeness in emotionally engaging interactions, but only when labelled as human. Commun Psychol. 2026;4:23. doi:10.1038/s44271-025-00391-7
17. Chaudhry BM, Debi HR. User perceptions and experiences of an AI-driven conversational agent for mental health support. Mhealth. 2024;10:22.
18. Reeves B, Nass C. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press; 1996.
19. Root D. Reconfiguring the alterity relation: the role of communication in interactions with social robots and chatbots. AI Soc. 2025;40(3):1321-1332.
20. Geroimenko V. Emotional and relational aspects: building rapport and human-like interaction with AI. In: Beyond and After Prompt Engineering: The Future of AI Communication. Springer Nature Switzerland; 2025:289-316.
21. Chavan V, Cenaj A, Shen S, et al. Feeling machines: ethics, culture, and the rise of emotional AI. arXiv. Preprint published June 2025. doi:10.48550/arXiv.2506.12437
22. Tsapro G, Gryshchenko O, Sivaieva O. Human-AI interaction in higher education: how students address, personify, and evaluate AI language models. Inf Technol Learn Tools. 2026;2(112):103-119.
23. Christof M, Armoundas AA. Implications of integrating large language models into clinical decision making. Commun Med. 2025;5(1):490. doi:10.1038/s43856-025-01216-8
24. Naous T, Ryan MJ, Ritter A, Xu W. Having beer after prayer? Measuring cultural bias in large language models. In: Ku LW, Martins A, Srikumar V, eds. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics: Volume 1, Long Papers. Association for Computational Linguistics; 2024:16366-16393. doi:10.18653/v1/2024.acl-long.862
25. Tao Y, Viberg O, Baker RS, Kizilcec RF. Cultural bias and cultural alignment of large language models. arXiv. Preprint published November 2023. doi:10.48550/arXiv.2311.14096
26. Kim MS. Non-Western Perspectives on Human Communication: Implications for Theory and Practice. Sage; 2002.
27. Chu MD, Gerard P, Pawar K, Bickham C, Lerman K. Illusions of intimacy: emotional attachment and emerging psychological risks in human-AI relationships. arXiv. Preprint published May 2025. doi:10.48550/arXiv.2505.11649
28. Zemore SE, Ziemer KL, Gilbert PA, Karno MP, Kaskutas LA. Understanding the shared meaning of recovery from substance use disorders: new findings from the What Is Recovery? Study. Subst Abuse. 2023;17:11782218231199372.
29. Beresford P, Rose D. Decolonising global mental health: the role of Mad Studies. Cambridge Prisms Glob Ment Health. 2023;10:e30.
30. Bouguettaya A, Stuart EM, Aboujaoude E. Racial bias in AI-mediated psychiatric diagnosis and treatment: a qualitative comparison of four large language models. NPJ Digit Med. 2025;8:332. doi:10.1038/s41746-025-01746-4
31. Omar M, Soffer S, Agbareia R, et al. Sociodemographic biases in medical decision making by large language models. Nat Med. 2025;31:1873-1881. doi:10.1038/s41591-025-03626-6
32. Kang J. Linguistic humanization in conversational AI: a systematic-narrative review of humor, empathy, tone, and imperfection in relational communication. AI Soc. 2026;41(4):3997-4008.
33. Turner N, Welches P, Conti S. Mindfulness-Based Sobriety: A Clinician's Treatment Guide for Addiction Recovery Using Relapse Prevention Therapy, Acceptance and Commitment Therapy, and Motivational Interviewing. New Harbinger Publications; 2014.
34. Manoli A, Pauketat JV, Ladak A, Noh H, Hwang AHC, Anthis JR. Digital companionship: overlapping uses of AI companions and AI assistants. In: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery; 2026:1-25.
35. Hsu SC, Chiu CH, Wang HC, Lin XR. Exploring the role of AI as a recovery companion for people who use drugs: potential and boundaries. In: Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing. Springer Nature Switzerland; 2025:213-223.
36. Johansson M, Romero D, Jakobson M, Heinemans N, Lindner P. Digital interventions targeting excessive substance use and substance use disorders: a comprehensive and systematic scoping review and bibliometric analysis. Front Psychiatry. 2024;15:1233888.
37. De Vries M, Begeman M, Derksen M, Goudriaan A, Blankers M. User perspectives on conversational agents for preventive alcohol self-help: qualitative exploratory study. JMIR Hum Factors. 2026;13:e86752. doi:10.2196/86752
38. Yang F, Oshio A. Using attachment theory to conceptualize and measure the experiences in human-AI relationships. Curr Psychol. 2025;44(11):10658-10669.
39. Mitchell JJ, Jeon M. Exploring emotional connections: a systematic literature review of attachment in human-robot interaction. Int J Hum Comput Interact. 2025;41(18):11753-11774.
40. Heng S, Zhang Z. Attachment anxiety and problematic use of conversational artificial intelligence: mediation of emotional attachment and moderation of anthropomorphic tendencies. Psychol Res Behav Manag. 2025;18:1775-1785. doi:10.2147/PRBM.S531805
41. Banks J. Deletion, departure, death: experiences of AI companion loss. J Soc Pers Relat. 2024;41(12):3547-3572. doi:10.1177/02654075241269688
42. Jones M, Schonewille M. Synthetic relationships and artificial intimacy: an ethical framework for evaluating the impact of generative-AI on community. J Ethics Entrep Technol. 2026;6(1):56-70.
43. Golden A, Aboujaoude E. A transdiagnostic model for how general purpose AI chatbots can perpetuate OCD and anxiety disorders. NPJ Digit Med. 2026;9:343. doi:10.1038/s41746-026-02531-7
44. Korte KJ, Norr AM, Schmidt NB. Targeting safety behaviors in the treatment of anxiety disorders: a case study of False Safety Behavior Elimination Treatment. Am J Psychother. 2018;71(1):9-20. doi:10.1176/appi.psychotherapy.20180001
45. Haroon R, Dogar FR. TwIPS: a large language model powered texting application to simplify conversational nuances for autistic users. In: Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24). Association for Computing Machinery; 2024:1-18.
46. Koegel LK, Ponder E, Bruzzese T, et al. Using artificial

intelligence to improve empathetic statements in autistic adolescents and adults: a randomized clinical trial. J Autism Dev Disord. 2026;56(7):2513-2529. doi:10.1007/s10803-025-06734-x

47. Xu Z, Liu F, Xia G, Duan Y, Yu L. A scoping review of inclusive and adaptive human-AI interaction design for neurodivergent users. Disabil Rehabil Assist Technol. 2026;21(4):943-961.
48. Carter JW. Understanding problematic AI companion use through escapism and resilience. Comput Human Behav Artif Humans. 2026;8:100308. doi:10.1016/j.chbah.2026.100308
49. Suárez D, García-Mariñoso B. On the verge of a digital divide in the use of generative AI? Telecommun Policy. 2025;49(7):102997. doi:10.1016/j.telpol.2025.102997
50. Scholich T, Barr M, Wiltsey Stirman S, Raj S. A comparison of responses from human therapists and large language model-based chatbots to assess therapeutic communication: mixed methods study. JMIR Ment Health. 2025;12:e69709. doi:10.2196/69709
51. Weir IB, Stroud AM, Stout JJ, et al. Physician perspectives on the impact of artificial intelligence on the therapeutic relationship in mental health care: qualitative study. JMIR Ment Health. 2025;12:e81970. doi:10.2196/81970
52. Milford SR, Herger E, Eichinger J, Elger BS. Promoting xenomorphic patient-facing AIs: the case against anthropomorphism in medical AIs. NPJ Digit Med. 2025;8:667. doi:10.1038/s41746-025-02046-7
53. De Freitas J, Cohen IG. Unregulated emotional risks of AI wellness apps. Nat Mach Intell. 2025;7(6):813-815. doi:10.1038/s42256-025-01051-5
54. Song Y, Cui M, Wan F, Yu Z, Jiang J. AI hallucination in crisis self-rescue scenarios: the impact on AI service evaluation and the mitigating effect of human expert advice. Int J Hum Comput Interact. 2025;41(22):14419-14439.