23. Heine S.J. Norenzayan A. Toward a psychological science for a cultural species// Perspectives on Psychological Science. - 2006. - Vol. 1(3). - P. 251–269.

МРНТИ 15.81.61 **10.51889/2959-5967.2025.85.4.021**

Akhatova S.T.[1] , Tapalova O.B.[1] Кенжетаева Р.О[2] . Singh S[3]*
[1]Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
[2]Narxoz University
[3]Punjab University, Chandigarh, India

REVIEW OF INTERNATIONAL EXPERIENCE IN INTEGRATING CHATBOTS AND AI TECHNOLOGIES INTO PSYCHOLOGICAL SUPPORT PRACTICES FOR ADDICTION

Abstract

The global rise of substance-related and behavioral addictions highlights the need for scalable, accessible, and evidence-based psychological interventions. Traditional forms of counseling and rehabilitation often fail to ensure timely support, continuity of care, and sufficient therapist availability. This study aims to analyze and systematize international research on the integration of artificial intelligence (AI) chatbots into psychological support for addiction treatment.

Literature was searched in PubMed, Scopus, Web of Science, and Google Scholar (2017–2025) using predefined keywords; inclusion focused on chatbot interventions for substance-related or behavioral addictions reporting empirical outcomes or validated therapeutic frameworks. Six representative models were analyzed: Woebot/W-SUDs, Tess, Chatbot-Assisted Therapy (CAT) for methamphetamine use, Quin for smoking cessation, the simulation-based LLM system ChatThero, and a field-defining systematic review.

The analysis demonstrated that chatbot-based interventions are feasible and acceptable for individuals with substance-use and behavioral addictions. Reported outcomes include reduced craving intensity, improved emotional regulation, and increased engagement in treatment. CAT showed measurable clinical improvement through toxicology data and retention rates, while Quin and ChatThero demonstrated enhanced adaptation to motivational interviewing and relapse-prevention frameworks.

The findings indicate that AI-mediated conversational systems can effectively complement traditional therapy by providing continuous support and promoting self-management. Further research is needed to verify long-term clinical outcomes, ensure ethical oversight, and adapt interventions to diverse cultural and linguistic contexts.

Keywords: artificial intelligence in mental health, AI-driven chatbots, addiction treatment, digital psychological support, relapse prevention technologies.

Ахатова С.Т.[1], Тапалова О.Б.[1], Кенжетаева Р.О[2]. Singh S [3].
[1]Абай атындағы Қазақ ұлттық педагогикалық университеті (Алматы, Қазақстан)
[2]Нархоз университеті (Алматы, Қазақстан)
[3] Пенджаб университеті, Чандигарх, Индия

ТӘУЕЛДІЛІККЕ ҚАРСЫ ПСИХОЛОГИЯЛЫҚ ҚОЛДАУ ПРАКТИКАЛАРЫНА ЧАТ-БОТТАР МЕН ЖАСАНДЫ ИНТЕЛЛЕКТ ТЕХНОЛОГИЯЛАРЫН ИНТЕГРАЦИЯЛАУДЫҢ ХАЛЫҚАРАЛЫҚ ТӘЖІРИБЕСІНЕ ШОЛУ

Аңдатпа

Заттық және мінез-құлықтық тәуелділіктердің әлемдік деңгейде өсуі ауқымды, қолжетімді және ғылыми дәлелденген психологиялық араласулардың қажеттілігін айқындайды. Дәстүрлі

кеңес беру және оңалту түрлері көбіне уақтылы қолдауды, терапияның үздіксіздігін және мамандардың жеткілікті қолжетімділігін қамтамасыз ете алмайды.

Осы зерттеудің мақсаты — тәуелділіктерді емдеуде жасанды интеллект (ЖИ) негізіндегі чат-боттарды психологиялық қолдау тәжірибесіне енгізу бойынша халықаралық зерттеулерді талдау және жүйелеу.

Әдебиеттерді іздеу PubMed, Scopus, Web of Science және Google Scholar деректер базаларында 2017–2025 жылдар аралығында алдын ала анықталған кілт сөздер арқылы жүргізілді. Іріктеу критерийлеріне тәуелділікке қарсы чат-бот араласуларының эмпирикалық нәтижелері немесе валидтелген терапиялық үлгілері бар зерттеулер енгізілді.

Алты үлгілік модель талданды: Woebot/W-SUDs, Tess, метамфетаминге тәуелділікті емдеуге арналған Chatbot-Assisted Therapy (CAT), темекі шегуді тоқтатуға бағытталған Quin, симуляциялық LLM жүйесі ChatThero, сондай-ақ осы саладағы бағыт беруші жүйелі шолу (systematic review).

Талдау нәтижелері чат-бот негізіндегі араласулардың химиялық және мінез-құлықтық тәуелділігі бар адамдар үшін тиімді әрі қабылдануға лайықты екенін көрсетті. Есептелген нәтижелердің ішінде — тәуелділікке деген ұмтылыстың төмендеуі, эмоциялық реттелудің жақсаруы және емдеу процесіне қатысу деңгейінің артуы бар.

CAT моделі токсикологиялық деректер мен қатысушылардың бағдарламада қалу көрсеткіштері арқылы клиникалық тұрғыдан маңызды жақсару көрсетті, ал Quin мен ChatThero мотивациялық сұхбат және рецидивтің алдын алу қағидаттарына тиімді бейімделгенін дәлелдеді.

Зерттеу нәтижелері ЖИ арқылы жүзеге асатын диалогтық жүйелердің дәстүрлі психотерапияны тиімді толықтыра алатынын, үздіксіз қолдау көрсетіп, өзін-өзі басқару дағдыларын дамытуға ықпал ететінін дәлелдейді.

Алдағы зерттеулер ұзақ мерзімді клиникалық нәтижелерді растауға, этикалық бақылауды қамтамасыз етуге және интервенцияларды әртүрлі мәдени және тілдік контексттерге бейімдеуге бағытталуы тиіс.

Түйін сөздер: психикалық денсаулықтағы жасанды интеллект, ЖИ негізіндегі чат-боттар, тәуелділікті емдеу, цифрлық психологиялық қолдау, рецидивтің алдын алу технологиялары.

Ахатова С.Т.[1], Тапалова О.Б.[1], Кенжетаева Р.О[2]. Singh S[3].
[1]Казахский национальный педагогический университет имени Абая (Алматы, Казахстан)
[2]Университет Нархоз (Алматы, Казахстан)
[3]Университет Пенджаба, г. Чандигарх, Индия

ОБЗОР МЕЖДУНАРОДНОГО ОПЫТА ИНТЕГРАЦИИ ЧАТ-БОТОВ И ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ПРАКТИКУ ПСИХОЛОГИЧЕСКОЙ ПОДДЕРЖКИ ПРИ ЗАВИСИМОСТЯХ

Аннотация

Глобальный рост числа случаев зависимостей — как связанных с употреблением психоактивных веществ, так и поведенческих — подчёркивает необходимость масштабируемых, доступных и научно обоснованных психологических интервенций. Традиционные формы консультирования и реабилитации нередко оказываются недостаточными для обеспечения своевременной поддержки, непрерывности терапии и достаточного охвата клиентами.

Цель данного исследования — проанализировать и систематизировать международные работы, посвящённые интеграции чат-ботов на основе искусственного интеллекта (ИИ) в практику психологической помощи при зависимостях.

Поиск литературы осуществлялся в базах данных PubMed, Scopus, Web of Science и Google Scholar за период с 2017 по 2025 годы с использованием заранее определённых ключевых слов. Включение исследований проводилось по критерию наличия эмпирических данных или валидированных терапевтических моделей, связанных с чат-ботами, применяемыми при химических и поведенческих зависимостях.

Были проанализированы шесть репрезентативных моделей: Woebot/W-SUDs, Tess, Chatbot-Assisted Therapy (CAT) для лечения зависимости от метамфетамина, Quin для отказа от курения, симуляционная LLM-система ChatThero, а также крупный систематический обзор, задающий направления дальнейших исследований.

Результаты анализа показали, что интервенции с использованием чат-ботов являются реализуемыми и приемлемыми для лиц с химическими и поведенческими зависимостями. Среди зафиксированных эффектов — снижение интенсивности влечения, улучшение эмоциональной саморегуляции и повышение вовлеченности в терапевтический процесс. Модель CAT продемонстрировала клинически значимое улучшение по данным токсикологических анализов и показателям удержания в программе, тогда как Quin и ChatThero проявили высокую степень адаптации к принципам мотивационного интервьюирования и профилактики рецидивов.

Полученные данные свидетельствуют о том, что системы диалогового взаимодействия, основанные на ИИ, могут эффективно дополнять традиционную психотерапию, обеспечивая непрерывную поддержку и способствуя самоуправлению поведением.

Для дальнейшего развития направления необходимы исследования, направленные на подтверждение долгосрочных клинических результатов, обеспечение этического контроля и адаптацию вмешательств к разнообразным культурным и языковым контекстам.

Ключевые слова: искусственный интеллект в психическом здоровье, чат-боты на основе ИИ, лечение зависимостей, цифровая психологическая поддержка, технологии профилактики рецидивов.

INTRODUCTION The intersection of clinical psychology and artificial intelligence (AI) is reshaping the way psychological support is conceptualized, delivered, and sustained in addiction treatment. As societies face rising rates of both substance-related and behavioral addictions, traditional models of psychological assistance struggle to provide timely, continuous, and stigma – free care. The resulting gap between need and care delivery has accelerated interest in digital systems – particularly AI-driven chatbots – as potential mediators of psychological support.

While digital health tools have historically been limited to tracking, monitoring, and prompting behavior, current systems increasingly simulate relational presence. Conversational agents such as Woebot, Tess, and Replika are no longer presented merely as wellness companions; they are being positioned as therapeutic actors within structured care pathways. These systems can deliver elements of cognitive-behavioral therapy (CBT), motivational interviewing (MI), relapse-prevention strategies, and self-monitoring prompts in moments when human contact is unavailable. Preliminary findings suggest that such interventions may reduce cravings, improve affect regulation, and increase treatment engagement in populations with addiction risk. However, the evidence remains uneven across regions and methodologies. Health systems differ in how they regulate the use of AI in mental health services, how they evaluate safety and efficacy, and how they define the ethical boundaries of automated support [1 – 3].

There is therefore a need to critically examine not only what AI-assisted chatbots are capable of, but under which conditions they can responsibly participate in addiction recovery. The aim of this narrative review is to analyze and systematize international experience with clinically oriented chatbots and AI-driven conversational systems designed to support individuals with substance use disorders and related addictive behaviors. We focus on representative systems from recent literature – Woebot (including its adaptation for substance use), Tess, Chatbot-Assisted Therapy (CAT) for

methamphetamine use, Quin for smoking cessation support, and the large-language-model-based system ChatThero—alongside a recent systematic review of chatbot-based addiction support. By comparing these models, we identify therapeutic mechanisms, methodological trends, and ethical constraints that shape how AI is being integrated into psychological assistance for addiction. We argue that AI should not be conceptualized as a replacement for psychotherapy, but as a scalable medium of therapeutic presence between human-led interventions.

Materials and methods / methodology of review This article is a narrative and comparative review synthesizing evidence on AI-enabled conversational agents developed for psychological support in the context of addiction. The review draws on peer-reviewed publications and preprint studies published between 2017 and 2025, including both clinical pilot studies and simulation-based evaluations.

Sources were identified through searches in PubMed, Scopus, Web of Science, and Google Scholar using combinations of the following keywords: “chatbot,” “conversational agent,” “addiction,” “substance use disorder,” “relapse prevention,” “motivational interviewing,” “cognitive behavioral therapy,” “digital mental health,” “AI,” and “large language model.” Studies were eligible for inclusion if they met at least one of the following criteria:

(1) they described a chatbot or AI-driven conversational agent explicitly designed to support individuals with substance-related or behavioral addictions;

(2) they reported empirical data on treatment engagement, craving reduction, relapse prevention, mood regulation, or retention in care;

(3) they proposed and evaluated an explicit therapeutic framework (e.g., CBT, MI, mindfulness-based relapse prevention) operationalized through chatbot interaction;

(4) they analyzed ethical, relational, or clinical implications of chatbot-mediated psychological support for addiction.

We excluded: (a) general-purpose assistants (e.g., non-clinical voice assistants) that provide unregulated advice without an embedded therapeutic model; (b) purely technical system descriptions without behavioral or clinical relevance; and (c) wellness chatbots for stress management that did not address addiction, relapse risk, or compulsive use. When multiple studies described iterations of the same system, we prioritized studies reporting concrete outcomes or clinically meaningful evaluation.

For comparative analysis, we selected six representative cases frequently cited in the literature or notable for methodological rigor:

– The systematic review by Ogilvie, Prescott, and Carson on chatbots in substance use support [4];

– Woebot and its adaptation for substance use (W-SUDs), evaluated by Prochaska et al. (2021) [1];

– Tess, developed by X2AI, focusing on culturally responsive emotional support [2];

– Chatbot-Assisted Therapy (CAT) for methamphetamine use disorder, implemented in an outpatient clinical setting [5];

– Quin, a smoking-cessation chatbot co-designed using real counselor–client dialogue [6];

– ChatThero (Wang et al., 2025), a large-language-model-based multi-agent system evaluated in simulated, clinically modeled addiction scenarios [7].

Because this is a narrative review and not a meta-analysis, the purpose is not to estimate pooled effect sizes, but to map therapeutic mechanisms, methodological structures, and ethical considerations across different chatbot paradigms.

RESULTS 1. *Systematic characterization of chatbot-based support for substance use*

The review by Ogilvie, Prescott, and Carson [4] represents one of the earliest structured attempts to synthesize empirical evidence on chatbots for individuals seeking help with substance use disorders. Their analysis emphasized that, despite the growing commercial visibility of AI-based mental health tools, relatively few chatbot interventions for addiction had undergone systematic clinical assessment. Most available studies were constrained by short intervention windows, self-report outcomes, small and demographically narrow samples, and the absence of long-term follow-up.

The authors also noted a conceptual split within the term "chatbot." On one side were purpose-built therapeutic chatbots designed to deliver structured psychological interventions such as CBT, motivational interviewing, or relapse-prevention strategies. On the other side were general-purpose conversational assistants capable of offering advice on substance use without clear clinical safeguards. This distinction is critical: unregulated assistants were shown in some cases to provide unsafe or misleading responses when queried about substance misuse, highlighting the ethical and clinical risk of deploying unsupervised conversational AI in high-vulnerability contexts.

2. *Woebot / W-SUDs: structured CBT and motivational interviewing in self-guided format*

Woebot is a conversational agent originally designed to deliver brief, structured cognitive-behavioral and dialectical behavior therapy (DBT) – inspired exercises, emotional check-ins, and reflective prompts through mobile messaging. Motivational interviewing (MI) provides a client-centered, autonomy-supportive therapeutic framework that emphasizes eliciting intrinsic motivation for change rather than imposing directive instructions. Miller and Rollnick [8] describe MI as a collaborative process focused on strengthening a person's own commitment to behavioral change through empathic communication, reflective listening, and reinforcement of self-efficacy. These principles align closely with the design of conversational agents for addiction support, which aim to encourage user-driven goal setting and enhance internal motivation through non-judgmental dialogue and supportive guidance.

Speaking about DBT, it is important to note that its relevance to addiction treatment has been supported in clinical settings. For instance, Rezaie et al. [9] reported that DBT significantly improved emotion regulation, distress tolerance, craving management, and depressive symptoms among individuals with opioid dependence, highlighting the role of DBT mechanisms in supporting behavioral change and affective stabilization in substance-use populations. These therapeutic principles inform Woebot's design, in which DBT-based coping strategies are translated into short, accessible micro-interventions aimed at enhancing self-regulation across daily situations. An adapted version, Woebot for Substance Use Disorders (W-SUDs), was evaluated by Prochaska and colleagues [1] as an eight-week, self-guided intervention for adults reporting problematic alcohol or drug use. Participants interacted with the chatbot several times per week, completing structured modules targeting relapse prevention, craving awareness, cognitive reframing, mood regulation, and self-efficacy.

Preliminary findings suggested several positive trends. Participants reported reductions in substance-use days per month and craving intensity, along with improvements in mood and anxiety. Self-reported confidence in resisting urges increased. In qualitative feedback, many participants described Woebot as supportive and trustworthy, and some reported a perceived sense of "being understood," despite knowing the agent was automated. This perception of therapeutic alliance—traditionally considered a human relational construct—appeared to emerge in part from Woebot's consistent availability, empathic tone, and ability to mirror users' language.

However, the study had limitations. The design was not fully randomized, the observation period was short, the outcomes were primarily self-reported, and the sample skewed toward individuals who were already motivated to seek help (and was largely female). The absence of long-term follow-up prevents conclusions about sustained abstinence or relapse trajectories. These constraints indicate that Woebot and W-SUDs should not be interpreted as stand-alone treatment for addiction, but rather as structured digital extensions of established therapeutic logic.

3. *Tess: culturally adaptive, stigma-reducing emotional support*

Tess, developed by X2AI, reflects a different design philosophy. Rather than strictly enforcing manualized psychotherapy techniques, Tess focuses on emotionally responsive, nonjudgmental dialogue with users who may be unwilling or unable to engage directly with human clinicians. Early deployments suggested decreases in perceived stress and increases in reported feelings of support, especially among users experiencing shame, fear of judgment, or social withdrawal. In the context of addiction, this matters because reluctance to seek help is often driven more by anticipated stigma than

by lack of need. Tess's contribution, according to early evaluations, lies not in rigid protocol delivery but in psychological safety: it provides a channel where users can disclose urges, fear of relapse, or emotional distress without experiencing moral condemnation [2].

Evidence from Tess is promising but remains largely based on self-report and short-term interaction metrics. There is limited data tying Tess specifically to objective addiction-related outcomes such as reductions in substance intake or verified relapse prevention. As such, Tess illustrates how conversational support can lower the emotional barrier to help-seeking, but it does not yet establish durable clinical impact on substance use behavior.

4. *Chatbot-Assisted Therapy (CAT) for methamphetamine use disorder: continuity of care in outpatient treatment*

One of the most clinically grounded implementations is Chatbot-Assisted Therapy (CAT), developed for individuals with methamphetamine use disorder and deployed in an outpatient addiction-treatment setting. Unlike many prototypes evaluated only in laboratory or pilot environments, CAT was integrated directly into ongoing clinical care. The chatbot was delivered through LINE, a widely used messaging platform, which reduced stigma associated with "installing an addiction app" and minimized friction for users [5].

CAT combined three functional strands: (1) psychoeducation about addiction mechanisms and high-risk triggers, reformulated into brief, daily micro-learning modules; (2) mindfulness-based relapse-prevention techniques aimed at helping users tolerate urges without acting on them; and (3) continuous self-monitoring of mood, craving intensity, and perceived stress, alongside links to crisis resources. The goal was not to "replace therapy," but to maintain therapeutic presence between human-led sessions—precisely the time window where relapse risk is often highest.

Across a six-month observation period, participants using CAT showed a lower proportion of methamphetamine-positive urine toxicology results compared with control patients, and modestly better treatment retention. Participants also described CAT as comforting, relevant, and easy to use, frequently mentioning the value of having "someone" who checked in daily. These findings suggest that reliable, judgment-free contact may stabilize motivation and reduce impulsive, high-risk behavior in critical moments. Importantly, these outcomes were supported by objective clinical measures (urine screens) rather than self-report alone, which strengthens the credibility of this model.

Nonetheless, limitations remain. The sample was relatively homogeneous, the follow-up window was still limited, and the system's language and tone were culturally specific. Long-term relapse trajectories and post-discharge outcomes were not fully established. Still, CAT demonstrates that AI-supported conversational systems can function as extensions of outpatient addiction care, especially in settings with limited human resources.

5. *Quin: co-designed smoking-cessation counseling through modeled human dialogue*

Quin is a chatbot developed to support smoking cessation, and it represents a methodological innovation in how therapeutic dialogue is encoded. Rather than being built purely from theory, Quin was trained using transcripts of real counseling sessions — approximately thirty recorded Quitline sessions totaling over eighteen hours of authentic therapist–client interaction. From these transcripts, developers extracted recurrent structures in counseling: assessment of motivation, identification of relapse triggers, reinforcement of self-efficacy, and planning coping strategies. These conversational patterns were then translated into a hybrid architecture combining rule-based logic, linguistic matching, sentiment analysis, and case-based reasoning.

Quin conducts an initial structured "appointment" to generate an individualized quit plan, then follows up at planned intervals to reassess motivation, discuss adherence to pharmacotherapy, and address emerging barriers. The chatbot refers back to the user's own prior statements ("you said last time that…"), which gives the interaction a sense of continuity and personal memory [6].

From a behavioral psychology perspective, Quin aligns closely with motivational interviewing principles: instead of prescribing action, it invites the user to articulate internally meaningful reasons to change, anticipate stressors, and select feasible coping strategies [10]. Early evaluations indicate

that users experience Quin as supportive and nonjudgmental, and that its consistent emotional neutrality reduces feelings of guilt associated with relapse events — a known barrier to continued engagement in cessation programs.

Beyond motivational interviewing, such emotionally supportive and autonomy-enhancing approaches are consistent with evidence from positive psychology interventions, which have been shown to strengthen well-being and reduce depressive symptoms through reinforcement of strengths-based coping and self-directed motivation [10]. This suggests that conversational systems incorporating compassionate tone, validation of user agency, and focus on psychological resources may foster more sustainable engagement and resilience throughout the cessation process.

At the same time, the Quin system's current implementation is shaped by a specific cultural and linguistic environment (Australian Quitline), which raises questions of generalizability across regions without adaptation.

6. *ChatThero: multi-agent large-language-model system for adaptive therapeutic dialogue*

ChatThero, introduced by Wang, Yao, Yang, Li, Qian, and Yu [7], represents a new generation of AI-driven psychological support systems for addiction recovery. Unlike earlier chatbots that relied on pre-authored scripts or finite-state logic, ChatThero is built on a large-language-model architecture aligned to therapeutic goals through supervised fine-tuning and Direct Preference Optimization. Its design centers on a multi-agent simulation framework: a "patient agent" that models emotional volatility and motivational resistance, a "therapy agent" that produces empathetic, motivational-interviewing-style responses, and an "environment agent" that injects destabilizing stressors such as financial strain or interpersonal conflict. This arrangement allows the system to rehearse complex, multi-turn addiction counseling scenarios under dynamic and high-risk conditions.

To train these agents, the authors constructed ethically sanitized, high-fidelity patient profiles derived from anonymized narratives in online addiction-support communities. These profiles encode factors such as impulsivity, ambivalence, relapse history, and emotional reactivity. In simulated trials with hundreds of virtual patients categorized by resistance level ("easy," "medium," "hard"), ChatThero increased motivational readiness for change and treatment confidence compared with baseline conversational models, and it managed to resolve high-resistance cases in fewer conversational turns. Human evaluators, including licensed clinicians, rated ChatThero's responses higher in empathy, persuasive appropriateness, and behavioral realism than those produced by comparison systems.

It is important to note, however, that these findings are based on simulation rather than longitudinal clinical outcomes in human patients. While the simulated patient model is psychologically sophisticated, real-world addiction is influenced by social, cultural, economic, and biological factors that cannot yet be fully captured in synthetic conversational environments. The authors themselves acknowledge the need for culturally adaptive versions, multilingual expansion, and ethically monitored clinical trials to evaluate sustained abstinence, relapse prevention over time, and therapeutic alliance in real patients.

DISCUSSION 1. *Evolution of chatbot-mediated addiction support*

Across the examples reviewed, AI-driven conversational systems have advanced along a clear trajectory. Early models (e.g., Tess) emphasized emotional accessibility and stigma reduction: creating psychologically safe space for disclosure without fear of moral judgment. Subsequent systems (e.g., Woebot / W-SUDs) embedded structured therapeutic content such as CBT, mindfulness, and relapse-prevention strategies into short, repeatable interventions. Clinically integrated systems (e.g., CAT) then extended this model into real outpatient care, offering continuous check-ins, psychoeducation, and coping prompts between human-led sessions — precisely when relapse risk peaks. More recent systems (e.g., Quin) built on real-world counselor language to reproduce motivational interviewing dynamics with high fidelity, and advanced frameworks (e.g., ChatThero) now attempt to model resistance, ambivalence, and situational stressors adaptively using large-language-model architectures.

Taken together, these developments suggest that the function of chatbots in addiction care is shifting from static information delivery toward dynamic therapeutic mediation. The contemporary chatbot does not only "give advice"; it engages in ongoing dialogue that supports self-reflection, monitors relapse risk, reinforces self-efficacy, and sustains contact through moments of acute vulnerability.

2. *Psychological mechanisms: presence, safety, and self-efficacy*

Several recurring psychological mechanisms emerge across systems:

– Continuous presence. CAT and Woebot create a sense of ongoing accompaniment, which participants describe as reassuring. This continuity of presence appears to support motivation and reduce feelings of isolation during craving episodes.

– Psychological safety. Tess and Quin demonstrate that nonjudgmental tone and emotional neutrality can lower shame and encourage honest self-report. This is particularly relevant for individuals who avoid human clinicians due to fear of stigma.

– Guided self-efficacy. Woebot, Quin, and ChatThero all reinforce autonomy by prompting users to articulate their own reasons for change, identify personal triggers, and rehearse coping strategies. This aligns with motivational interviewing, which treats intrinsic motivation as more durable than externally imposed instruction.

– Micro-structures of accountability. CAT shows how daily, low-intensity check-ins can serve as a behavioral anchor, reminding patients that recovery is an active, continuous process rather than a topic confined to scheduled therapy hours.

These mechanisms are clinically meaningful because addiction relapse is often opportunistic and affect-driven: the urge emerges, peaks, and recedes within minutes. Traditional therapy models struggle to cover these "in-between" intervals. Chatbots, by contrast, can inhabit exactly that temporal space.

3. *Ethical and clinical constraints*

Despite their promise, AI-driven conversational agents in addiction care raise unresolved concerns.

First, there is the question of "algorithmic empathy." Many participants in studies such as Woebot and CAT described a felt sense of being understood. This subjective bond may improve engagement, but it also raises questions of informed consent: to what extent do users understand the limits of the system's abilities, and are they at risk of over-relying on a nonhuman agent during psychological crises? If a user experiences the chatbot as a stable attachment figure, does that create a new form of dependence on a digital presence rather than fostering internalized self-regulation?

Second, safety and accuracy remain critical. The review by Ogilvie et al. showed that general-purpose conversational assistants can provide misleading or dangerous advice regarding substance use. This indicates that not all "chatbots" are clinically appropriate for vulnerable users. Addiction-support systems require explicit therapeutic framing, escalation pathways, and transparent ethical boundaries regarding crisis intervention.

Third, privacy and data ethics are central. Systems like ChatThero attempt to address this by training on sanitized, reconstructed patient profiles rather than identifiable clinical transcripts. This approach allows simulation of high-risk counseling scenarios without exposing real patients. However, once such systems are deployed in real settings, questions of data retention, clinician oversight, and accountability for adverse outcomes will intensify.

Fourth, cultural and contextual adaptation remains underdeveloped. Quin is tuned to Australian smoking-cessation discourse; CAT is embedded in a clinical and linguistic environment where LINE is normative; ChatThero is trained on English-language, Western-context addiction narratives. The generalizability of these systems to other regions — including Central Asia and other areas with limited access to mental-health professionals — is still largely hypothetical. Without localization of language, norms, stigma narratives, and health-system pathways, the same chatbot may produce very different levels of trust, engagement, and safety.

4. *Structural limitations of current evidence*

Across the literature, several methodological limitations recur:

– Short follow-up windows. Most interventions are evaluated over weeks or a few months; there is little evidence about relapse trajectories six months or one year later.

– Non-randomized or partially controlled designs. Many studies rely on self-selected, motivated participants, which may inflate observed benefits.

– Reliance on self-report. Apart from CAT (which included urine toxicology), outcomes often depend on self-declared mood, craving, or use reduction.

– Limited demographic diversity. Some samples are predominantly female, treatment-seeking, and already in contact with services, which limits generalizability to populations with lower motivation or severe polysubstance use.

– Simulation-based evaluation. ChatThero demonstrates advanced adaptive interaction under simulated conditions. This is promising for safety testing and scalability, but findings from simulated "patients" cannot be assumed to predict longitudinal outcomes in real human users.

By acknowledging these limitations, we can interpret the reported benefits as important but preliminary. The evidence supports the feasibility and acceptability of chatbot-mediated psychological support in addiction care, and in some cases suggests clinically meaningful associations with improved engagement, reduced risky use behavior, and enhanced readiness to change. However, claims of causal impact on long-term abstinence or relapse prevention remain provisional.

5. *Implications for underserved systems*

One of the most practically significant findings concerns scalability. Models such as CAT show that chatbot-mediated support can be embedded into existing outpatient infrastructure without replacing clinicians. This has direct implications for regions with limited specialist availability, such as rural areas or health systems with workforce shortages. In such contexts, chatbots may function as continuity-of-care scaffolds: maintaining contact, reinforcing relapse-prevention strategies, and prompting crisis escalation when needed. Rather than competing with therapists, these systems can multiply therapeutic reach during high-risk intervals between in-person sessions.

CONCLUSION This review shows that the integration of AI-driven chatbots into psychological support for addiction has progressed from experimental novelty to a structured and increasingly clinically informed field. Early systems established emotional availability and stigma reduction; subsequent models embedded validated therapeutic techniques such as CBT, motivational interviewing, mindfulness-based relapse prevention, and craving management; clinically integrated tools extended care into the daily lives of patients; and emergent large-language-model frameworks now attempt to simulate adaptive therapeutic reasoning, resistance, and relapse risk in complex, evolving psychological states.

Across these developments, a consistent theme emerges: the value of AI in addiction support lies less in technical sophistication and more in psychological integrity. Systems are most effective when they embody therapeutic principles such as respect for autonomy, nonjudgmental stance, self-efficacy building, and consistent presence. Preliminary evidence suggests that such systems can support treatment engagement, reduce self-reported craving and negative affect, and improve perceived readiness to change. At the same time, the risks are nontrivial. Overreliance on perceived "algorithmic empathy," absence of culturally adapted safeguards, and limited long-term validation all indicate that these tools must remain under human clinical oversight.

AI should not be framed as a replacement for psychological care in addiction. Its emerging role is as a medium of therapeutic presence: maintaining supportive contact when human providers are unavailable, reinforcing coping strategies during moments of acute vulnerability, and transforming recovery from a once-a-week appointment into an ongoing, lived process. The challenge for the next stage of research is not merely to build more persuasive chatbots, but to establish ethically accountable, clinically validated, culturally adaptive systems that extend—not supplant—the human capacity to offer stability, understanding, and hope.

References

1. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, Robinson A. A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study. J Med Internet Res. 2021 Mar 23;23(3):e24850. doi: 10.2196/24850. PMID: 33755028; PMCID: PMC8074987.

2. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Ment Health. 2018 Dec 13;5(4):e64. doi: 10.2196/mental.9782. PMID: 30545815; PMCID: PMC6315222.

3. Ta-Johnson, Vivian & Griffith, Caroline & Boatfield, Carolynn & Wang, Xinyu & Civitello, Maria & Bader, Haley & DeCero, Esther & Loggarakis, Alexia. (2020). User Experiences of Social Support from Companion Chatbots in Everyday Contexts: Thematic Analysis. Journal of Medical Internet Research 2020 Mar 6;22(3):e16235. doi: 10.2196/16235. PMID: 32141837; PMCID: PMC7084290.

4. 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. doi: 10.1159/000525959. Epub 2022 Aug 30. PMID: 36041418.

5. Chun-Hung L, Guan-Hsiung L, Wu-Chuan Y, Yu-Hsin L. Chatbot-assisted therapy for patients with methamphetamine use disorder: a preliminary randomized controlled trial. Front Psychiatry. 2023 Jul 7;14:1159399. doi: 10.3389/fpsyt.2023.1159399. PMID: 37484677; PMCID: PMC10359989.

6. Bendotti H, Ireland D, Lawler S, Oates D, Gartner C, Marshall HM. Introducing Quin: The Design and Development of a Prototype Chatbot to Support Smoking Cessation. Nicotine Tob Res. 2024 Apr 22;26(5):612-620. doi: 10.1093/ntr/ntad217. PMID: 37936253; PMCID: PMC11033568.

7. Wang, Junda & Yao, Zonghai & Yang, Zhichao & Li, Lingxi & Qian, Junhui & Yu, Hong. (2025). ChatThero: An LLM-Supported Chatbot for Behavior Change and Therapeutic Support in Addiction Recovery. 10.48550/arXiv.2508.20996.

8. Miller, W. R., & Rollnick, S. (2013). Motivational interviewing: Helping people change (3rd edition). The Guilford Press.

9. Rezaie, Zeinab & Afshari, Behrooz & Balagabri, Zohreh. (2021). Effects of Dialectical Behavior Therapy on Emotion Regulation, Distress Tolerance, Craving, and Depression in Patients with Opioid Dependence Disorder. Journal of Contemporary Psychotherapy. 55. 163-172. 10.1007/s10879-020-09487-z.

10. Sin NL, Lyubomirsky S. Enhancing Well-Being and Alleviating Depressive Symptoms With Positive Psychology Interventions:A Practice-Friendly Meta-Analysis. Journal of Clinical Psychology. 2009; 65(5):467–487.10.1002/Jclp.20593 [PubMed: 19301241]