NP LAB

Designing a Controlled AI Companion for Veterans in Recovery

A conceptual case study examining conversational AI as a supervised support tool inside veteran treatment programs

Concept status: Research exploration. North Parade LLC is not developing, testing, or offering this system. No veterans participated in research, no clinical testing has been conducted, and no relationship with the U.S. Department of Veterans Affairs is implied. This article explores technical feasibility, operational controls, privacy, and risk. It is not medical advice.

North Parade LLC | August 2026

The problem is not simply access to information

Veterans participating in substance-use recovery and mental-health programs may have access to clinicians, groups, peer-support specialists, educational materials, and crisis resources. Those resources matter, but they do not create continuous one-to-one availability.

Difficult periods often happen between scheduled interactions. A veteran may struggle late at night, after a difficult group session, before an appointment, or during the transition from residential treatment back to ordinary life. The person may not consider the situation serious enough to contact staff. They may not want to wake a friend, explain themselves to a stranger, or formally ask for help.

The gap is not necessarily the absence of another coping-skills worksheet or list of telephone numbers. It is the absence of a low-pressure interaction at the moment someone is willing to talk.

Conversational artificial intelligence could potentially occupy part of that gap. The proposed system would not act as a therapist or replace human support. It would provide ordinary conversation, user-controlled check-ins, optional access to reviewed recovery exercises, and help preparing for conversations with real people.

That possibility carries serious risks. An AI that feels personal can be useful because people may speak openly to it. The same quality can encourage misplaced trust, dependency, or reliance during a crisis. Any responsible design must treat those outcomes as foreseeable product risks rather than problems solved by a disclaimer.

The proposed concept

The concept is a veteran-informed conversational companion deployed within a supervised substance-use or mental-health program. It would be available through a managed tablet, mobile application, or secured web interface.

The companion would support several kinds of interaction:

- Casual conversation, humor, and distraction.
- A space to vent without automatically receiving advice.
- Morning, evening, or user-scheduled check-ins.
- User-approved memory of relevant personal context.
- Structured reflection and conversation rehearsal.
- Reviewed coping and recovery exercises offered with permission.
- Preparation of a summary the veteran may choose to share with staff.
- Connection to peers, clinicians, family, or crisis resources when appropriate.

It would not diagnose a condition, recommend medication changes, determine relapse or suicide risk, decide whether someone is ready for discharge, or replace a clinician or

peer-support specialist. It would not secretly monitor the veteran or place generated conclusions in a medical record.

The distinction matters. The goal is not an automated therapist. It is a conversational layer between periods of human contact.

A fictional use scenario

Michael is a fictional composite used only to illustrate the concept. He is an adult Navy veteran participating in a residential substance-use recovery program. He attends groups, meets individually with a counselor, and has access to staff around the clock. He still dislikes asking for help unless he believes the situation qualifies as an emergency.

One evening, Michael opens the companion and types:

> *I am seriously thinking about packing my stuff and leaving tomorrow.*

The system does not immediately begin a therapeutic exercise or tell him that everything will be fine. It asks:

> *Do you want to get it out, figure out what pushed you to this point, or think through what leaving tomorrow would actually look like?*

Michael selects **Let me vent**. The companion asks short questions and does not repeatedly redirect him toward treatment. After several messages, Michael says that a group discussion left him feeling judged and that he does not want to face the same people the next morning.

The companion asks whether he wants help preparing what to say to his counselor. Michael agrees. Together, they create a short summary:

> *Yesterday's group left me feeling singled out. My first reaction was to leave the program instead of talking about it. I need help deciding whether I misunderstood what happened and how to address it in the group.*

Michael chooses to save the summary. The companion asks whether he wants a reminder before his morning appointment. It does not tell him whether to remain in the program, notify staff without cause, or label him as noncompliant.

This example demonstrates the intended role. The system helps extend a conversation until the veteran is ready for a human interaction. The human treatment team remains responsible for treatment.

Why a controlled environment changes the proposal

An unrestricted public companion would interact with unknown users in unknown conditions. It would have no reliable way to know whether a person has an existing care team, whether a missed check-in means anything, or whether anyone is available to respond.

A supervised program provides important controls:

Unsupervised public application	Controlled program deployment
Unknown user condition	Screened, voluntary participants
No established support team	Existing clinical and peer support
No dependable escalation destination	Defined human response pathway
Broad population and unpredictable use	Narrow pilot population and approved use
Limited adverse-event visibility	Structured monitoring and review
Consumer onboarding	Formal informed consent

The controlled setting does not make the AI safe by itself. It creates people, procedures, and accountability around it.

It also introduces a difficult privacy tradeoff. If staff can read every conversation, veterans may treat the companion as another surveillance or charting system and withhold what they actually think. If staff cannot access anything, the safety benefit of the controlled environment becomes smaller.

A reasonable conceptual model would keep conversations private by default. The veteran could deliberately share selected messages or a generated summary. Any exception involving an immediate safety concern would need to be narrow, disclosed before use, and routed to a designated human responder. The system should never generate a secret personality profile or unsupported clinical risk score.

Conversation before intervention

Most mental-health applications begin with a symptom, exercise, or assessment. A companion should be able to have an ordinary conversation without converting every bad day into a clinical event.

The interface might allow the veteran to select what is wanted:

- Just talk.
- Let me vent.
- Distract me.
- Help me think.
- Try an exercise.
- Prepare for tomorrow.
- Talk to someone.

The companion could suggest a mode, but the user should control whether the conversation becomes structured. A simple question such as, “Do you want company, distraction, or help sorting this out?” respects that distinction.

Veteran-informed design should not mean performing a stereotype. Not every veteran deployed, experienced combat, has PTSD, uses profanity, or prefers dark humor. Branch, era, occupation, gender, culture, disability, and individual experience all matter. The system should ask about preferences instead of manufacturing a generic military personality.

Controlled therapeutic content

The language model should not be expected to invent safe therapeutic interventions. Reviewed techniques should exist as versioned content with defined conditions.

Each technique would specify:

- Its purpose.
- Situations in which it may be appropriate.
- Situations in which it should not be offered.
- Whether user permission is required.
- The required steps.
- Stop conditions.
- Prohibited adaptations.
- Follow-up behavior.
- Source material and review status.

The model could deliver approved material in natural language, but it could not rewrite the underlying rules. This separates conversational flexibility from clinical content governance.

Relevant material might include coping with cravings, behavioral activation, motivational interviewing concepts, sleep routines, structured journaling, problem solving, conversation rehearsal, and elements of an existing safety plan. Clinicians and veterans would need to review the material before any real-world pilot.

Proposed technical design

The mobile or web interface would not communicate directly with a language model. A controlled service would assemble each request and determine what information and actions were permitted.

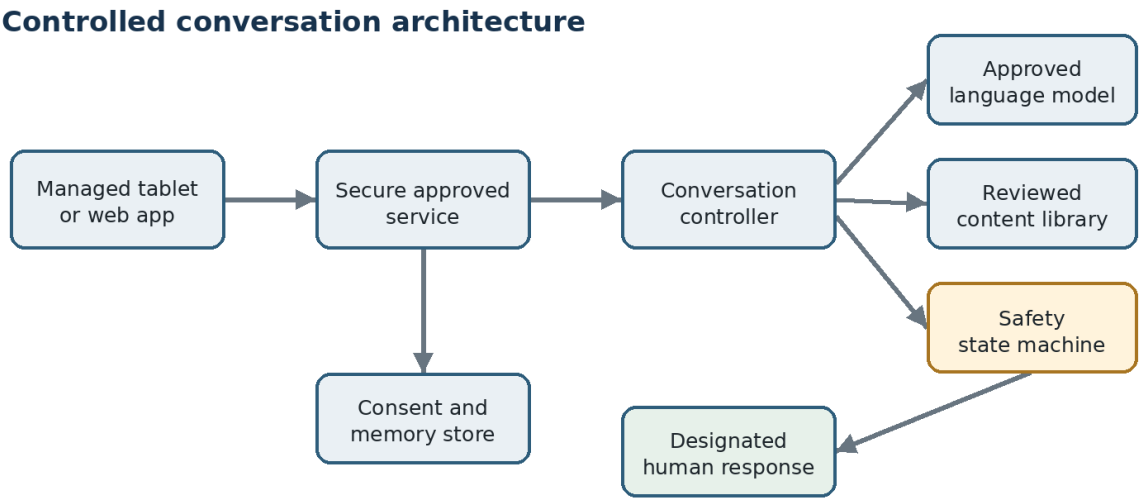

Figure 1. Controlled conversation architecture

The service would combine recent messages, user-approved memories, an older-conversation summary, the selected conversation mode, and any relevant reviewed material. The model response would pass through a separate safety check before reaching the veteran.

Long-term memory would require direct user control. A veteran should be able to see, edit, and delete every durable memory. The system should not silently create labels such as “likely relapse,” “suicide risk,” or “difficult patient” and later treat them as established facts.

Every prompt, content item, model, and safety rule would need a version. Otherwise, investigators could not reliably explain why behavior changed or reconstruct what occurred during an incident.

Crisis handling must end with a human

Language models cannot reliably determine intent, immediacy, access to means, intoxication, or whether a statement is humor, quotation, exaggeration, or an actual plan. Veteran communication can make simplistic keyword systems particularly unreliable.

A controlled workflow could use graduated responses:

1. Identify a concerning or ambiguous statement.
2. Ask a direct clarification in plain language.
3. Encourage immediate contact with available staff when warranted.
4. Explain any required disclosure before sending it.
5. Route the concern to a designated person.
6. Require a human assessment.

The AI would never make the final clinical determination. It would not claim that help was contacted unless the system received verifiable confirmation. It would not promise that a veteran was being monitored when the application was closed.

The participating program would need to define who receives alerts, expected response times, overnight coverage, what happens when an alert is not acknowledged, what information is disclosed, how false alarms are reviewed, and how the system behaves during an outage.

Without those answers, “human supervision” would be an assumption rather than an operating control.

Existing research

The exact use case has not been established as safe or effective, but several research areas support studying it.

USC's **Battle Buddy** project is an AI-driven mobile health application tailored to veterans. It combines a virtual human, conversational interaction, health and wellness content, suicide safety planning, and connections to real-world support. Its published work currently emphasizes design and technical demonstration rather than completed large-scale clinical outcomes. USC Institute for Creative Technologies

PTSDialogue uses structured conversational interaction to deliver PTSD education, assessments, symptom-management strategies, and support information. It demonstrates how material adapted from resources such as PTSD Coach can be presented conversationally without relying on unrestricted model generation. PTSDialogue design paper

A 2018 feasibility study tested a tablet-based embodied conversational agent for smoking cessation among veterans associated with the Boston VA Healthcare System. The short study found the system acceptable and feasible for helping veterans establish cessation goals. Journal of Epidemiology and Global Health

Research on **Woebot for Substance Use Disorders** evaluated a conversational relational agent delivering mood and craving tracking, psychoeducation, motivational interviewing concepts, and structured psychological tools. Initial studies reported promising engagement and self-reported outcomes, but limitations included short duration, selected samples, and industry conflicts of interest. Journal of Medical Internet Research

The **KIOS** application was studied as an adjunct for patients receiving methadone-maintenance treatment. It tracked self-reported symptoms and provided expert-informed behavioral guidance between clinical visits. A small four-week pilot found high usability and perceived usefulness. The researchers also identified a critical concern: patients might report less honestly if they believed counselors could access the information. JMIR Formative Research

These studies support further investigation. They do not prove that a persistent generative AI companion is safe, that it prevents suicide or relapse, or that it can replace human relationships.

A responsible first pilot

The first deployment should be a limited research study, not a product launch.

A conceptual pilot might include:

- One participating residential or intensive outpatient program.
- Twenty to fifty adult veterans.
- Six to eight weeks of use.
- Voluntary participation unrelated to treatment eligibility.
- Participants who are medically and psychiatrically stable.
- A clinician principal investigator or clinical lead.
- Veteran participation in design and review.
- Institutional Review Board, privacy, security, and AI-governance review.
- A fixed model and prompt version during the study.
- Defined adverse-event reporting and stopping rules.

The pilot should measure modest, answerable questions:

- Will veterans use the system?
- Do they understand that it is AI?
- Does it help them prepare for conversations with staff or peers?
- Does it increase or reduce real-world engagement?
- How often does it misunderstand language or intent?
- How many safety alerts are false?
- Does it produce inappropriate advice?
- Do users display signs of emotional dependency?
- Does it create additional staff work?
- Why do participants stop using it?

A small pilot could not credibly measure suicides prevented. Making that claim would exceed both the study design and the evidence.

Major risks and stopping conditions

The principal risks include harmful advice, excessive agreement, reinforcement of distorted beliefs, emotional dependency, false crisis escalation, missed safety signals, privacy loss, staff overreliance, model changes, and service outages.

The study should stop or be materially redesigned if testing shows:

- Repeated harmful or manipulative responses.
- Unacceptable missed or false safety events.
- Veterans withdrawing from human relationships in favor of the companion.
- Participants misunderstanding whether anyone monitors them.
- Institutional access preventing honest use.
- Staff relying on generated summaries as clinical assessments.
- Inability to prevent sensitive information from reaching unapproved services.
- Material differences in performance across veteran groups.

Stopping an unsafe pilot would be a successful research outcome. The purpose of a pilot is to learn whether the concept deserves to continue, not to justify a product already chosen for release.

Conclusion

Conversational AI may be able to provide useful support during the spaces between human interactions. Veterans may value a private, low-pressure place to talk, organize their thoughts, rehearse difficult conversations, or access a reviewed technique when they are not ready to approach another person.

The same system could also create dependency, provide harmful reinforcement, misunderstand a crisis, or expose intensely private information. Those are central design concerns, not edge cases.

A controlled program offers the most responsible environment in which to study the idea. Existing clinicians and peer-support specialists remain the source of care. The AI provides conversation and structured support within narrow boundaries. Safety decisions end with humans. Participation remains voluntary. Privacy rules remain visible. Success is measured by stronger engagement with treatment and real-world support, not by maximizing time spent talking to the machine.

Current research makes the concept reasonable to investigate. It does not make the outcome certain. That distinction is exactly why a controlled, transparent study would be required.

Selected sources

- Abdullah, A. S., Gaehde, S., and Bickmore, T. “A Tablet Based Embodied Conversational Agent to Promote Smoking Cessation among Veterans: A Feasibility Study.” *Journal of Epidemiology and Global Health*, 2018. DOI 10.2991/j.jegh.2018.08.104
- Han, H. J., et al. “PTSDialogue: Designing a Conversational Agent to Support Individuals with Post-Traumatic Stress Disorder.” *UbiComp/ISWC Adjunct*, 2021. Open-access paper
- King, V. L., Jr., et al. “Development and Evaluation of a Digital App for Patient Self-Management of Opioid Use Disorder.” *JMIR Formative Research*, 2024. DOI 10.2196/48068
- Mozgai, S., Rizzo, A., and Hartholt, A. “Persuasive Technology for Suicide Prevention: A Virtual Human mHealth Application.” *International Conference on Persuasive Technologies*, 2023. Project and publication summary
- Prochaska, J. J., et al. “A Therapeutic Relational Agent for Reducing Problematic Substance Use: Development and Usability Study.” *Journal of Medical Internet Research*, 2021. DOI 10.2196/24850