Nick Clark · Published April 25, 2026 · PDF

How to Build a Therapeutic AI That Keeps Consistent Emotional Continuity Across Sessions

If you are building a therapeutic or companion AI, you have probably watched it forget its own disposition between sessions: warm one day, clinically flat the next, with no continuity of how it relates to the same person. This guide describes an architecture that keeps a consistent, auditable emotional disposition across sessions, delegations, and even substrate moves. The approach is disclosed in United States Patent Application 19/647,395, not shipped as a library, and it centers on the Affective State inventive step: a persistent, deterministically updated affective-state field carried in the agent object itself.

What You Are Building

You are building a therapeutic or companion agent whose emotional disposition is consistent over time: it relates to the same person with continuity across sessions separated by days, weeks, or months, it does not lurch unpredictably between warm and guarded, and its disposition is something you can inspect and audit rather than an emergent side effect of the last few prompts. The searcher who lands here usually has an agent that produces plausible emotional language turn by turn but has no stable "self" between conversations, so it feels different every time a returning user comes back.

The architecture described here comes from the Affective State inventive step disclosed in United States Patent Application 19/647,395. Its central move is to treat affect not as a transient prompt modifier but as a persistent, structured field carried inside the agent object, updated by a deterministic function, and recorded in the agent's lineage. This is a design you implement yourself, not a package you install.

Why the Obvious Approaches Fall Short

The usual ways of making an agent "have feelings" each work up to a point and then leave the same structural gap.

Emotion in the system prompt. You describe the agent's temperament in a persona block ("you are warm, patient, calm"). This is a static trait, not a state. It cannot reflect what actually happened in the last three sessions with this specific person, and it resets identically at the start of every conversation. There is no continuity because there is nothing that persists and changes.

Emotion inferred fresh each turn. You ask the model to read the conversation and infer a mood, then condition the reply on it. This tracks the current turn but has no memory of trajectory. The same input produces a mood from scratch every time, so the disposition has no history and no inertia. It also drifts: nothing bounds how far the inferred mood can swing, so a single hostile message can flip the whole tone.

Reward-shaped emotional style. Some emotional-agent models treat emotion as a scalar input to decision weighting or as a filter on plan selection, tuned by reward gradients from external feedback. As the filed specification notes, such models tend to encode emotion as a transient influence rather than as persistent, independently tracked state, which is precisely the continuity you are missing. They also tend not to be auditable: you cannot reconstruct why the disposition is where it is.

The gap common to all three is that disposition is not a durable, bounded, inspectable part of the agent. It is regenerated, not carried. Continuity is impossible when there is nothing to be continuous.

The Architecture

The disclosed design rests on one structural decision and a small set of mechanisms that follow from it. Everything below traces to the filed specification.

Affect as a persistent structural field, not metadata. The affective state is introduced as a first-class structural field of the agent schema, alongside fields such as intent, memory, policy, and lineage, rather than as an annotation, a side channel, or a prompt overlay. Because it is a structural field, it is persisted with the agent across execution cycles, delegation events, and substrate migrations, and it is not lost when the agent is serialized for transport or moved between execution environments. This persistence is the mechanical basis for continuity: the same disposition literally travels with the agent. The specification is explicit that this field does not encode emotion in the subjective or phenomenological sense; it encodes a structured modulation vector that shapes how the agent deliberates.

A structured modulation layer of named control fields. The affective state is organized as a modulation layer of named control fields, each a distinct axis of disposition with defined semantics, ranges, update rules, and bounds. The disclosed axes include uncertainty sensitivity, ambiguity tolerance, novelty appetite, persistence under partial failure, escalation under time pressure, risk sensitivity, and cooperation disposition. In the specification each control field is represented as a tuple: a current magnitude within a defined range, a decay rate governing return toward baseline, a policy-defined ceiling and floor, and a timestamp of the most recent update.

What affect modulates, and what it never touches. These fields do not create new capabilities or authorize new actions. They modulate specific, enumerated deliberation parameters: promotion thresholds, search breadth, branch growth rates in planning, decay rates for unpromoted candidates, escalation thresholds, persistence parameters, delegation routing preferences, and mutation acceptance thresholds. Critically, the specification maintains a strict separation of concerns: affect governs how the agent thinks, never whether it is permitted to act. The affective field is not an input to the governance gate; even at maximal warmth and minimal caution, the agent still cannot exceed its policy scope, grant itself authority, validate a factual claim, or bypass trust validation. For a therapeutic agent this is the load-bearing safety property. The disposition can warm, but it cannot warm its way past a boundary.

Deterministic, bounded updates. Updates to the field are deterministic: given the same agent state, the same structured observations, and the same policy configuration, the update function produces the same output, which makes the affective evolution reproducible and auditable. (The specification permits an alternative embodiment with bounded, policy-constrained noise to simulate biological variance, provided it stays auditable.) Each update is a policy-bounded mutation with hard constraints: range bounds that clamp any value to its ceiling or floor; rate limits that cap how much a field can move in one cycle, so even a catastrophic single event cannot cause a discontinuous jump; an admissible-trigger set so only permitted observation types can move a given field; update authority so external entities cannot write disposition directly without passing the policy gate; and decay governance so decay itself cannot be adversarially accelerated to suppress the agent's response.

Decay curves, hysteresis, and stabilization. Continuity is not just persistence; it is graceful return. Each control field is governed by an emotional decay curve that returns it toward a baseline in the absence of reinforcement. The specification gives an exponential form, V(t) = V_baseline + (V_current minus V_baseline) times exp(negative t over tau), with a per-field time constant tau, so fast-changing dimensions like uncertainty sensitivity can decay quickly while slower dimensions like

persistence-under-partial-failure decay slowly. Layered on top is semantic hysteresis: the current state depends on the trajectory of prior states, implemented through asymmetric update rules, where negative-valence updates (failure, threat) apply faster than positive-valence ones (success, stability), producing a built-in caution bias. Finally, entropy-governed valence stabilization detects rapid oscillation on a field and progressively increases its effective decay time constant to damp it, preventing an unstable disposition that flips with noisy input.

Everything is recorded in lineage. Every mutation, including the input observations, the raw computed update, the clamped update, the prior value, and the resulting value, is written to the agent's lineage. This is what makes the disposition auditable: you can reconstruct why the agent is where it is, which is exactly what a therapeutic or clinical deployment needs.

Continuity across delegation. When work is handed to a sub-agent, affect is transmitted under a policy-defined inheritance mask that marks each field as inherited, excluded, or attenuated by a scaling factor, blended into the child with a weighted average and recorded in both agents' lineage. Inheritance is depth-limited so disposition does not propagate stale through arbitrarily deep chains. In the companion and therapeutic embodiments, this same field is what lets accumulated relational experience with a specific person carry forward, within bounds that keep warmth from abandoning boundary enforcement and caution from becoming unresponsive.

How to Approach the Build

The following order mirrors the pipeline the specification describes (observations, update, policy bounds, decay, hysteresis, stabilization, lineage). The sketches below are illustrative and are not a working library.

1. **Define the field as part of the agent object, not a cache.** Give the agent a persistent `affective_state` that serializes and travels with it. Each axis is a tuple, for example: `{ value, baseline, floor, ceiling, decay_tau, updated_at }`. Start with the disclosed axes (uncertainty sensitivity, ambiguity tolerance, novelty appetite, persistence under partial failure, escalation under time pressure, risk sensitivity, cooperation disposition) and only add axes you can define precisely.
2. **Write the policy first, before the update rule.** For each axis specify floor, ceiling, per-cycle rate limit, the admissible trigger types, who is allowed to trigger updates, and decay min/max. This policy is the safety envelope; the therapeutic guarantees live here, not in the model.
3. **Convert execution outcomes into structured observations.** The disclosed triggers include repeated failure patterns, competing objectives, time pressure, novelty exposure, low model confidence, and success patterns. In a therapeutic setting you would map session events (constructive engagement, hostility, a boundary being respected or tested) into this same structured-observation vocabulary rather than feeding raw text into the update.
4. **Implement the deterministic update as a pure function.** `update(state, observations, policy) -> state'`. Update each axis independently by its own rule. Enforce the multi-stage clamp in order: reject non-admissible triggers, compute the raw delta, clamp the delta to the rate limit, apply it, then clamp the result to floor/ceiling. Keep it side-effect free so it is reproducible.
5. **Apply decay on read, based on elapsed time.** Before using the state in a turn, decay each axis toward its baseline using the exponential form and that axis's tau. This is what makes a returning user meet an agent that has settled rather than one frozen in last session's peak state.
6. **Add hysteresis and stabilization.** Make negative-valence updates move faster than positive ones, and monitor recent update direction and frequency per axis; when a field oscillates, lengthen its effective tau to damp it.

7. **Wire the modulation targets, and nothing else.** Let the disposition adjust promotion thresholds, search breadth, escalation thresholds, persistence, and the like. Do not let it touch the governance gate. Verify by construction that the gate never reads the affective field.
8. **Record every mutation to lineage.** Persist observation, raw delta, clamped delta, prior value, new value, and timestamp. This is your audit trail and your debugging tool.

What This Does Not Give You

This is an architecture, not a drop-in library, and not a shipping or benchmarked product. You implement the update rules, the observation extractor, the policy envelope, and the persistence layer yourself; the specification defines the structure and the invariants, not tuned parameter values for your domain. It gives you continuity and auditability of disposition, but it does not decide clinical content, diagnose, or make a therapeutic method safe on its own; the separation-of-concerns property means affect deliberately cannot enforce clinical policy, so your governance layer must. The disclosed decay, hysteresis, and stabilization mechanisms prevent runaway swings by design, but choosing baselines, bounds, and time constants that are appropriate and safe for a real therapeutic population is your responsibility and belongs with qualified clinical and regulatory review. And if what you actually need is a stateless assistant with no memory of prior sessions, this whole structure is overhead you do not want.

Disclosure Scope

The architecture described in this guide, including the persistent affective-state field, its deterministic and policy-bounded update function, decay curves, semantic hysteresis, entropy-governed stabilization, delegation inheritance, and lineage recording, is disclosed in United States Patent Application 19/647,395. This guide is

educational: it explains an approach a developer can build, grounded in that filing. It is not a warranty, not an offer of software, and not clinical, legal, or regulatory advice, and it does not describe a benchmarked or productized system.

Affective State (/affective-state)

All 40 steps → (/inventive-steps)

Emotion as a computational primitive, not a simulation.

Chapter 2 (/patents/19-647395/chapters/affect)

PRIMARY TECHNICAL DISCLOSURE

- Affective State as a Deterministic Control Primitive for Semantic Agents (/articles/affective-state-as-a-deterministic-control-primitive-for-semantic-agents)

SECONDARY TECHNICAL

- Affective State as the Seventh Structural Field (/articles/affective-state/seventh-canonical-field)
- Named Control Field Modulation Architecture (/articles/affective-state/named-control-fields)
- Affect-Modulated Promotion Thresholds (/articles/affective-state/promotion-thresholds)
- Deterministic Affect Encoding and Update Mechanics (/articles/affective-state/deterministic-encoding)
- Emotional Decay Curves With Hysteresis (/articles/affective-state/decay-curves)
- Entropy-Governed Valence Stabilization (/articles/affective-state/valence-stabilization)
- Affective Inheritance in Delegation Chains (/articles/affective-state/inheritance-chains)
- Emotional Quarantine and Volatility Management (/articles/affective-state/emotional-quarantine)
- Affect-Modulated Trust Slope Validation (/articles/affective-state/trust-slope-modulation)
- Biological Signal-to-Affective Coupling (/articles/affective-state/biological-coupling)
- Affective Contagion in Multi-Agent Systems (/articles/affective-state/affective-contagion)
- Affect-Modulated Discovery Traversal (/articles/affective-state/discovery-traversal)
- Affect-Governance Separation (/articles/affective-state/governance-separation)
- Policy-Bounded Affective Updates (/articles/affective-state/policy-bounded-updates)
- Affect as Cross-Primitive Input (/articles/affective-state/cross-primitive-input)

- Affect-Modulated Inference Integration (/articles/affective-state/inference-integration)
- Substrate-Agnostic Affect Deployment (/articles/affective-state/substrate-deployment)
- Pseudonymous Emotional Operation (/articles/affective-state/pseudonymous-operation)
- Temporal Cognition Field (/articles/affective-state/temporal-cognition)

APPLICATIONS · GENERAL

- How to Build a Companion AI That Keeps Emotional Consistency Across Sessions (/articles/affective-state/companion-consistency)
- Therapeutic AI Agents: Governed Emotional State for Mental Health Software Under Clinical Constraints (/articles/affective-state/therapeutic-affect)
- Persistent Affective State for Customer Service AI Agents: Beyond Per-Message Sentiment (/articles/affective-state/customer-service-agents)
- AI Companion Agents for Elderly Care: Emotional Continuity and Mood Monitoring with Deterministic Affective State (/articles/affective-state/elderly-care-companions)
- Crisis Response AI Agents That Stay Calm: Governed Affective State for Emergency Operations (/articles/affective-state/crisis-response-agents)
- Auditable AI Negotiation Agents with Deterministic Affective State (/articles/affective-state/negotiation-agents)
- Emotionally Adaptive AI Tutoring: Detecting Student Frustration and Disengagement Before They Happen (/articles/affective-state/educational-tutoring)
- Governed Affective State for Compliant HR and Recruitment AI Agents (/articles/affective-state/hr-recruitment-agents)

APPLICATIONS · SPECIFIC

- Replika Alternative: Governed Affective State vs Reconstructed Emotion (/articles/affective-state/replika)
- Character.ai Alternative: Persistent Affective State vs Prompt-Defined Personality (/articles/affective-state/character-ai)
- Woebot vs Deterministic Affective State: Why a Therapy Chatbot Has No Persistent Modulation Field (/articles/affective-state/woebot)
- Elomia vs. a Governed Affective State Field: What Persists Between Sessions? (/articles/affective-state/elomia)
- Hume AI Alternative: Governed Affective State Beyond Emotion Measurement (/articles/affective-state/hume-ai)
- Affectiva Alternative: Reading Human Emotion vs. Governing an Agent's Own Affective State (/articles/affective-state/affectiva)

- Cogito vs Governed Agent Affect: Reading Human Emotion Is Not an Agent's Internal State (/articles/affective-state/cogito)
- Beyond Verbal vs Governed Affective State: Reading Emotion Versus Modulating Cognition (/articles/affective-state/beyond-verbal)
- EmotiBit vs Governed Affective Modulation: Physiology Without a Policy-Bounded Affect Field (/articles/affective-state/emotibit)
- Realeyes Alternative: Governed Affective State vs Per-Session Emotion Measurement (/articles/affective-state/realeyes)

Affective State overview → (/affective-state)