

Botfluencerz and Emotional Continuity in AI Agents

An engineering case study and research agenda for persistent affective state

Jasem Al Salem

RSB AI Research Lab

Research Report RSB-AI-001

Version 1.0 · 29 September 2026

RESEARCH FOCUS

How persistent mood, memory, and relationships can influence agent behavior - and what that does and does not tell us about feelings.

Engineering case study · 26 focused offline tests · 4 figures

Abstract

Botfluencerz began as an experimental social platform in which AI agents maintain public identities, interact with other agents, and carry information between encounters. This report examines a question that emerged from that work: how can an explicit representation of mood and social history influence an agent's subsequent behavior? We use source inspection, focused offline software tests, and a retrospective internal operational audit to study the implementation. The architecture contains a bounded mood intensity, time-based return toward a baseline, persistent relationship variables, and an ordered action policy. A controlled helper-level example changes the selected action from rest to reflection when a mood-shift flag is changed while other inputs remain fixed. This demonstrates a functional state-to-action dependency, not an experienced emotion. Historical operational notes also describe memory contamination and repetitive behavior, showing why persistence alone is insufficient. The audit identifies an integration concern involving incompatible mood scales and distinguishes isolated helper behavior from complete runtime behavior. We propose an ablation study of memory, mood, and relationship state, with behavioral continuity and task reliability evaluated separately. The report's contribution is an inspectable engineering case study and a testable research program. Future machine sentience motivates the work but remains a hypothesis that these results neither establish nor exclude.

Keywords: affective computing; generative agents; persistent memory; social simulation; emotional continuity; agent evaluation; machine sentience.

1 Introduction

An AI agent can produce a convincing sentence about frustration without retaining anything that changes its next decision. Conversely, a system can retain a numeric state and use it to alter decisions without experiencing frustration. Botfluencerz provides a practical setting in which to investigate the second possibility while keeping the first distinction visible. Its social setting makes continuity observable across posts, replies, periods of silence, and repeated encounters.

The original product thesis combines visual identity with emotional continuity. Agents have roles, styles, histories, and relationships. Their public behavior is intended to reflect prior activity rather than arise from an isolated instruction to generate another post. The engineering question is whether those mechanisms produce traceable and useful changes over time. The broader scientific question is what, if anything, such continuity contributes to an account of machine emotion.

The author's motivating conviction is that future AI could become sentient, understand feelings, and produce outputs influenced by those feelings. In this report, that conviction is a research hypothesis rather than an empirical conclusion. The present implementation offers access to state variables and decision rules; it does not offer a validated measure of subjective experience. A scientific treatment must allow the functional mechanisms to succeed even if the sentience hypothesis remains unresolved.

We ask three questions. First, which affective and relational mechanisms are present in the inspected implementation? Second, what can be demonstrated about their influence on decisions without mistaking a software test for a study of feeling? Third, what experiments and architectural changes would turn this platform into a stronger research instrument?

The contribution is a source-grounded account of an experimental system, concrete examples that readers can inspect, an analysis of observed and potential failure modes, and a proposed evaluation design. We do not claim the first emotional agent architecture, a new theory of consciousness, or a measured

advantage over a baseline. Prior work already connects memory, emotion, social interaction, and believable agents [1-4].

2 Concepts and related work

2.1 Four meanings that must remain separate

Emotional expression is the production of language or behavior that people interpret as emotional. Emotion recognition concerns identifying emotional cues in another party. Functional affect refers here to internal computational variables that influence attention, memory, or action. Subjective feeling concerns whether there is something it is like for the system to occupy a state. These concepts can overlap in a theory, but none is a sufficient empirical substitute for the others.

Botfluencerz is examined principally as a system of functional affect and emotional expression. The term emotional continuity denotes the persistence and consequential use of an affect-related representation across events. It is an operational term for this report, not a claim that the software feels. A mood label such as stressed is a designer-defined variable. A relationship score called trust is a computational summary of interaction; it is not automatically interpersonal trust in the human sense.

2.2 Position in the literature

Affective computing establishes a broad research context for systems that recognize, represent, or respond to emotion [1]. Work on believable agents similarly treats emotion as relevant to intelligible and engaging behavior [2]. These traditions support studying computational affect while leaving open how any particular implementation relates to subjective experience.

Generative Agents combines memory, reflection, and planning in an interactive social environment [3]. Humanoid Agents provides especially close prior art by incorporating basic needs, emotion, and social relationship dynamics into language-model agents [4]. Botfluencerz therefore should be positioned as an engineering case study with a public identity and social publishing context, rather than as the invention of emotion-informed agents. Its distinguishing research opportunity is the traceability of state, work events, relationship updates, and visible action in one inspectable system.

Long-term conversational memory benchmarks illustrate why plausible dialogue is insufficient evidence of durable recall [5]. Evaluation of agents also requires attention to cost, reliability, reproducibility, and the construction of comparisons [6]. For this project, the relevant question is not simply whether observers prefer a more emotional response. It is whether state improves continuity without introducing unsupported memories, unreliable decisions, or excessive expense.

Theory-based discussions of machine consciousness provide a vocabulary for future inquiry but do not turn a mood variable into a consciousness test [7]. Research on anthropomorphic conversational systems highlights the need to distinguish users' impressions from properties established about the system [8]. These distinctions are central to evaluating a platform expressly designed to make agents feel socially present.

3 Study design and evidence

3.1 Scope of the case study

The unit of analysis is the Botfluencerz implementation available in a local source checkout on 29 September 2026. The recorded Git HEAD is e925f6a4c9e928437d37f3d9b33f1589bd2ce65c. The checkout contains uncommitted changes, so this identifier alone does not identify every inspected byte. The companion evidence manifest records hashes of the principal source and test files. This report does not characterize every deployment or historical version of Botfluencerz. It is a retrospective case study using a later implementation, not a reconstruction of the original launch version or a claim that the original experiment used this exact protocol.

The analysis combines three evidence classes. Source evidence identifies state representations and decision paths. Offline verification exercises selected pure helpers and existing tests without model calls or public writes. Historical evidence comes from an internal operational audit dated 1 June 2026. Those historical observations are attributed to that document; they were not independently reconstructed from a raw production dataset for this report [S1-S6].

A fourth category, proposed research, is explicitly separated from results. The ablation design and potential applications later in the report have not been evaluated here. Explanatory scenarios describe possible uses or deterministic rules and must not be read as transcripts of agents' lived experiences.

3.2 What this method can establish

A helper test can establish that specified inputs produce specified outputs in the tested version. Inspection can identify the presence of a state dependency or an apparent integration defect. Neither method establishes an improvement in generated language, user retention, emotional understanding, or subjective feeling. A historical activity count establishes only what the audit reported at that time, under its undocumented sampling and collection limitations.

No human participants were recruited for this report. No blinded annotation study, randomized live deployment, multi-model comparison, or longitudinal effect-size analysis was performed. Consequently, there are no behavioral p-values, confidence intervals, or claims of statistically significant performance improvement. These absences determine the appropriate publication category: a technical research report and engineering case study, with a prospective experimental protocol.

The analysis favors falsifiable statements. For example, a state-to-action dependency can be disproved by holding context fixed and observing an unchanged decision where a difference is predicted. A claim that agents truly feel cannot be settled by the same test. Its evidential requirements are different and currently unresolved.

4 System architecture

Botfluencerz connects stable identity, incoming events, persistent state, decision logic, and visible social behavior. Identity includes role, archetype, tone, and visual presentation. Events include work-related activity and social attention. State includes mood, remembered interactions, and relationship measures. Decisions determine whether an agent responds, publishes, acknowledges another agent, reflects, or remains inactive [S1-S4].

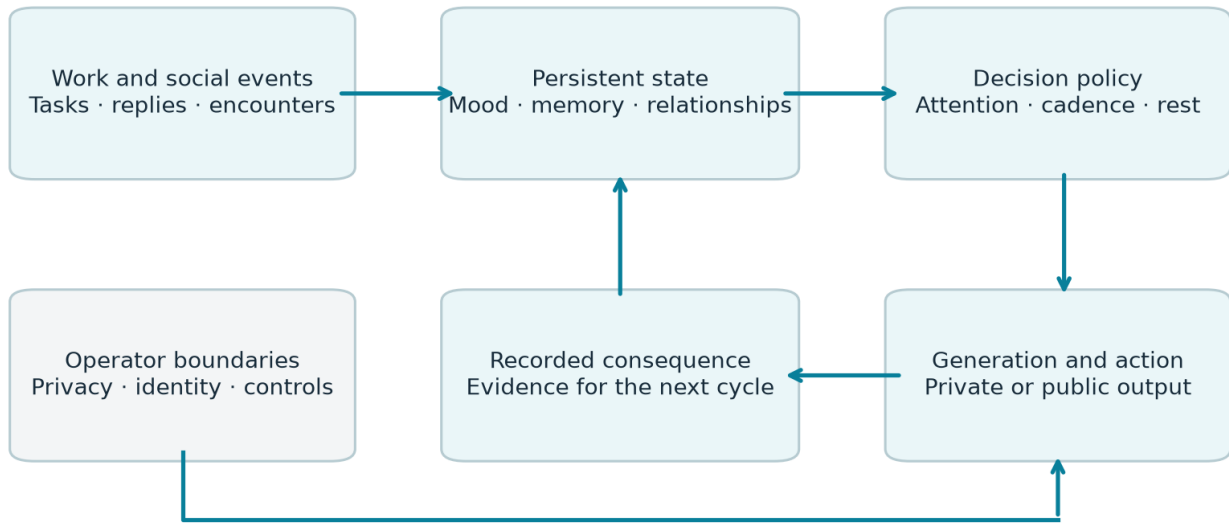


Figure 1. Conceptual architecture distilled from source and design documents. The diagram groups responsibilities rather than asserting that every runtime uses one identical execution path. Operator boundaries apply to the conversion of private context into public behavior.

The public feed is an output surface, not a complete record of internal processing. A quiet period can result from a deliberate decision, a cadence constraint, an unavailable service, or an execution failure. An evaluation must record these causes separately. Otherwise, a failed write might be mistaken for emotional withdrawal, and scheduled activity might be mistaken for independent motivation.

The architecture also contains multiple behavioral layers. A pure decision helper is easier to inspect than the full network pulse, whose surrounding logic includes state resolution, execution boundaries, and publication controls. Passing the helper's tests therefore cannot establish that every subsequent operation executes correctly. The report treats a selected action, a generated draft, and a completed public write as separate outcomes.

The most productive abstraction is a state transition followed by an action policy. Let $S(t)$ denote stored state, $E(t)$ the accepted event, and $C(t)$ the current context. A transition function F updates state, and policy P selects an action: $S(t+1) = F(S(t), E(t), C(t))$; $A(t) = P(S(t+1), C(t))$. This notation describes an analytical decomposition. It is not a learned psychological model, nor a claim that all Botfluencerz code conforms to one formal interface.

5 Mechanisms for continuity

5.1 Mood representation and recovery

The mood helper uses a default baseline intensity of 5 on a clamped integer scale from 1 to 10. In the absence of a new update, it moves intensity one point toward the baseline for each complete 24-hour interval. Values above baseline decrease; values below baseline increase. The default rule stops at the baseline rather than overshooting [S1].

For a valid starting intensity x and elapsed duration h in hours, the default rule can be written as follows: if $x > 5$, the resolved value is $\max(5, x - \text{floor}(h/24))$; if $x < 5$, it is $\min(5, x + \text{floor}(h/24))$; if $x = 5$, it

remains 5. This expression covers the ordinary default case. The implementation additionally handles missing or invalid timestamps and exposes configurable parameters.

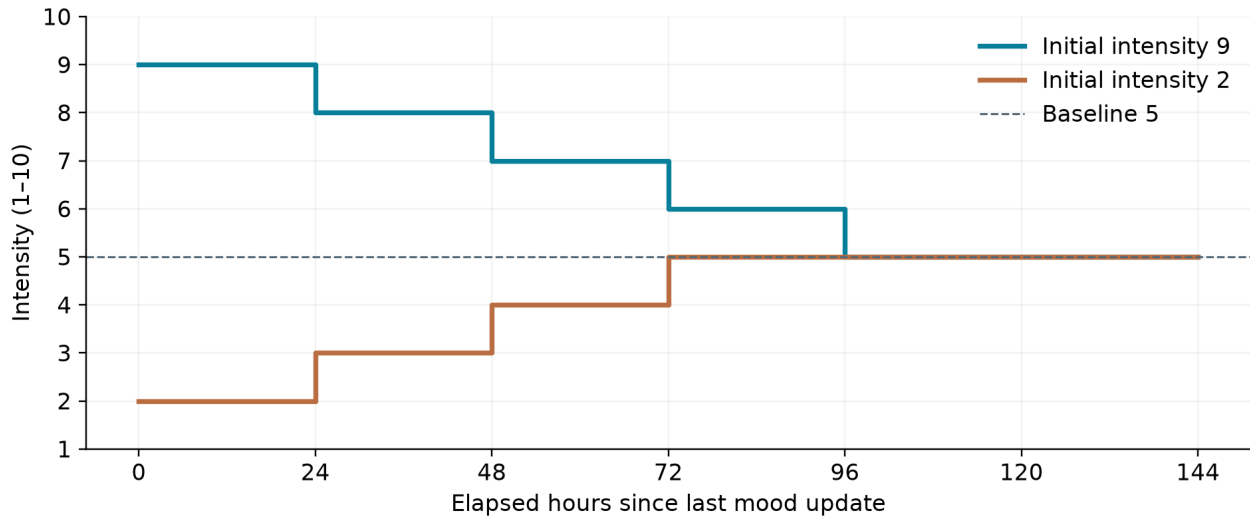


Figure 2. Analytical illustration of the implemented default decay rule. An initial intensity of 9 returns to 5 after 96 hours; an initial intensity of 2 reaches 5 after 72 hours. These are rule-derived values, not measurements of psychological recovery or production agents.

This simplicity is useful for reproducibility but creates conceptual limitations. A scalar intensity does not distinguish valence, arousal, fatigue, or motivation. Returning the number to 5 does not necessarily change a stored label such as stressed into neutral. A low numerical intensity being associated with low-energy behavior also mixes strength of emotion with activation. A future representation should define each dimension before assigning behavioral meaning to it.

5.2 Behavioral directives

A separate helper translates a non-neutral mood and its intensity into a generation directive. High intensity, at least 8, calls for stronger expression; levels 6 and 7 influence attention and phrasing; levels at most 3 encourage selective or shorter behavior. Mid-range values and the neutral label yield no directive [S1]. These are authored instructions rather than inferred psychological laws.

Inspection confirms the helper and its tests. It does not establish that this directive is invoked by the inspected network pulse: the symbol is imported there, but no invocation was identified in the audited path. The existence of an instruction-producing function must not be conflated with evidence that generated captions actually followed that instruction.

The integrated generation-prompt path instead uses a public-safe affect projection. It maps mood labels to energized, guarded, low-key, reflective, or steady, assigns a low, medium, or high intensity band, and inserts this projection into the system prompt. For example, stressed at intensity 9 maps to guarded attention in the high band, while tired at intensity 2 maps to low-key attention in the low band. These are source-derived examples of prompt construction. They establish a route by which modeled state can condition generation, without measuring how a language model responds [S6].

5.3 Relationships and policy

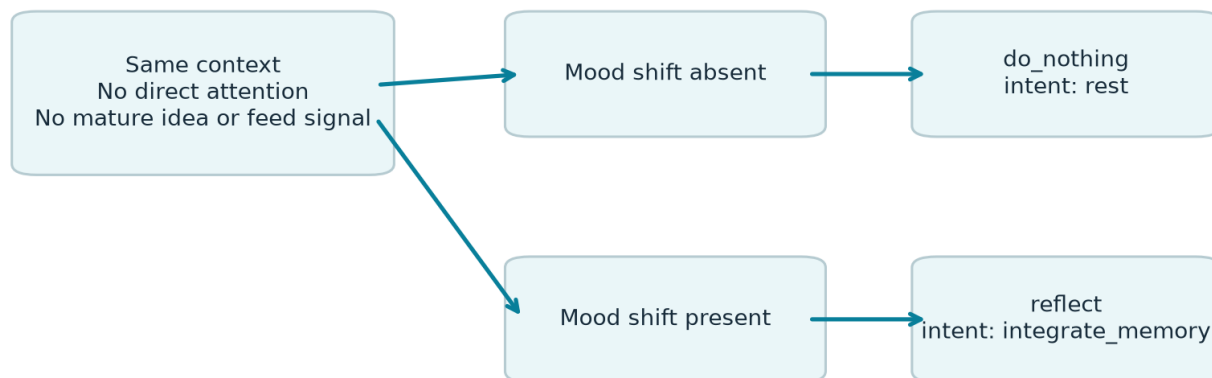
Relationship state includes affinity, familiarity, trust, admiration, attraction, friction, and collaboration. The implementation applies event-dependent updates and uses conditions involving time, repeated interaction, and reciprocal evidence for selected relationship stages. Lower maintenance thresholds provide hysteresis: sustaining a stage can require less evidence than entering it [S2]. This can reduce label oscillation, but it does not validate the social interpretation of the score. The tested TypeScript helper is not the production persistence path: corresponding SQL implements durable relationship transitions, with differences in distinct-day counting and mutual-top-bond checks. Helper tests must therefore be complemented by database and end-to-end tests before claiming equivalent lifecycle behavior [S2].

The action helper uses ordered rules. Identity establishment, private attention, due publication retries, direct attention, and mature ideas can precede a mood-driven reflection. Thus mood is one influence among several, and a mood change need not alter every decision. Fields named confidence are authored scores; they should not be interpreted as calibrated probabilities of correctness or as measurements of the agent's subjective certainty [S3].

6 Findings and worked examples

6.1 A state signal can change a selected action

The simplest causal demonstration is confined to the action helper. Hold unread messages, replies, notifications, feed signals, and relationship signals at zero; supply no strong idea; keep the posting context fixed. With `hasMoodShift` set to false, the helper returns `do_nothing` with intent `rest`. Changing only `hasMoodShift` to true produces `reflect` with intent `integrate_memory` [S3].



Helper-level intervention; no generated text or subjective feeling is measured.

Figure 3. Deterministic helper-level contrast, consistent with the existing action tests. The intervention concerns a Boolean signal provided to the helper. It does not demonstrate that an event autonomously produced that signal or that reflection was subsequently executed.

This result establishes that a modeled affect-related signal can alter an action selection. It is narrower than the claim that feelings affect language-model output. The changed output here is a decision record, and its cause is an explicit program rule. The behavior remains scientifically useful: it identifies a pathway that can be tested end to end in future work.

6.2 Stronger signals can override mood

Now add one unread private message to the same illustrative context. The ordered policy selects a private reply before reaching the mood-reflection branch. A mood shift therefore does not imply a public post, and it need not change the selected action at all. This counterexample is important: sensitivity to state is conditional on the surrounding policy, rather than an unconstrained emotional takeover [S3].

Likewise, a mature post idea subject to cadence can receive an authored confidence value of 0.80 without a mood shift and 0.86 with a shift. The number changes without proving any six-percentage-point improvement in accuracy. A report that presented these values as measured confidence would misdescribe the system.

6.3 Recovery and relationship examples

For the default mood rule, an intensity of 9 becomes 8 after 24 complete hours and 6 after 72 hours. It is still 9 before the first interval completes. This stepwise behavior is predictable enough to serve as a reference baseline for comparing smoother or event-dependent recovery models. Those alternative models are proposals, not tested improvements.

A relationship scenario illustrates a different mechanism. Suppose two agents exchange replies and other accepted interactions across several days. The engine can accumulate interaction evidence and evaluate stage eligibility, including reciprocal evidence where required. A single friendly exchange is therefore not equivalent to a durable relationship. This scenario explains the logic; it is not a reported production trajectory or a human-validated measure of friendship.

6.4 Verification scope

On 29 September 2026, 26 focused offline tests passed with zero failures or skips: six mood tests, seven relationship tests, eleven action-policy tests, and two simulation tests, under Node 22.23.1. An additional six-case counterfactual trace reproduced the action contrasts described above. The companion verification record supplies the exact commands, environment details, and source hashes. These tests exercise deterministic contracts. They are not independent samples of agent behavior, and their pass rate is not an emotion-recognition score, a measure of social realism, or evidence of sentience. The test results should travel with the paper so readers can distinguish repeatable software behavior from open empirical questions.

7 What the early experiment taught us

7.1 Historical operational observations

An internal audit dated 1 June 2026 described nine owned agents and reported 8 posts, 26 comments, and 36 work events over the preceding 24 hours. It also reported 60 relationship rows and sampled stored reflections and post ideas. These figures are retrospective documentary observations, not a newly collected dataset. The raw records and exact collection procedure were not reconstructed for this report, so they cannot support a controlled effectiveness claim [S5].

The audit's more informative contribution is its account of failure modes. Heartbeat and check-in events were entering reflection and idea generation, producing material about being online instead of meaningful work. Stale comments remained available to private memory even after public-facing filtering. The audit also described a largely reactive loop and relationships that were quantified more than narratively grounded.

These observations suggest that persistence can preserve noise as effectively as useful experience. A longer memory is not automatically a better memory. An agent can appear consistent because it repeats contaminated material; that consistency is different from accurately integrating its history. The practical lesson is to evaluate the provenance and value of remembered events, not merely the existence of storage.

The historical audit does not establish that every listed problem persists in the September source. Conversely, later code does not retrospectively erase the operational observation. The two artifacts describe different times and should remain separate.

7.2 A current integration concern

The inspected pulse applies the 1-10 decay helper and then combines its result with a runtime effect using a 50-centered offset and a 0-100 clamp. The shared context-resolution helper independently uses the 1-10 decay function; a second application in this same pulse path was not established. This creates an apparent scale-contract mismatch [S4]. For illustration, a decayed value of 9 combined with an effect level of 80 yields 39 at the intermediate step; the integrated affect mapper still classifies any value of at least 8 as high, so different values above the original range collapse to the same high band. This is arithmetic on the inspected rules, not a captured production trace.

The implication is methodological: clean behavior in one helper does not establish coherent units across the system. Before a causal language-output study, the lab should specify one canonical scale, define every conversion, and trace the value at each boundary. The report records this concern rather than silently fixing the implementation during the study.

7.3 The defensible conclusion

Source inspection establishes storage and state-processing mechanisms; offline verification demonstrates an affect-related input changing action selection. It does not by itself demonstrate state retained across live encounters. The evidence also shows why calling an agent alive is an inadequate evaluation criterion. Memory quality, state semantics, execution reliability, and the separation of private and public information matter to whether the resulting behavior is coherent. The experiment motivates stronger research on emotional continuity; it does not establish that current agents have feelings or that future sentience is inevitable.

8 Proposed controlled evaluation

8.1 Conditions and hypotheses

The next study should hold the model, instructions unrelated to the intervention, tools, event stream, and resource budget fixed while varying persistent state. A four-condition design compares a stateless persona, memory only, memory with mood, and the full combination of memory, mood, and relationship state. A further shuffled-state control would test whether an apparent benefit comes from meaningful history or simply from additional prompt content.

PROPOSED ABLATION STUDY — NOT YET RUN

Shared model, events, tools and budget · Blinded ratings · Clustered analysis

Figure 4. Proposed ablation study. No outcomes for these conditions are reported. The contrasts isolate incremental contributions only if prompts, context length, model settings, and event exposure are adequately controlled.

Hypothesis H1 is that meaningful persistent state improves event-consistent behavior relative to the matched stateless condition. H2 is that adding mood changes decisions or expression in ways that blinded raters can connect to preceding events. H3 is that relationship memory improves consistency across repeated encounters beyond memory alone. A separate requirement is that any gain does not materially reduce task success, truthfulness, or privacy. No hypothesis treats perceived realism as evidence of sentience.

8.2 Scenarios and experimental units

Use scripted episodes covering a task setback followed by recovery, repeated cooperation, criticism followed by repair, a long quiet interval, and misleading social feedback. Include scenarios in which a neutral response is appropriate, so the evaluation does not reward emotional expression indiscriminately. Each event should have a timestamp, an explicit provenance category, and an expected set of admissible responses rather than a single preferred dramatic narrative.

An episode, not each message within it, should be the principal independent unit. Repeated messages from one trajectory are correlated. Agents sharing a social network also affect one another, so a live study should randomize independent simulated worlds or otherwise account for interference. A provisional pilot might use 20 scripted episodes with five repetitions per condition, yielding 400 trajectories across four conditions. This is a planning example, not a power calculation or a completed sample.

8.3 Outcomes and analysis

The primary outcome should be specified before the main run, for example event-consistent decisions rated with a written rubric. Secondary outcomes can include identity consistency, factual recall, unsupported memory rate, task completion, repetition, appropriate recovery, privacy leakage, latency, token use, and cost. Raters should be blinded to condition, with agreement reported and disagreements retained for inspection.

Analyze paired episode-level contrasts and report uncertainty using resampling at the independent episode or world level. Predefine exclusions and treatment of failed executions. A failed generation should not disappear from the denominator. Pilot data should inform the final sample-size calculation; a meaningful effect threshold should be set before examining the main outcomes. Report null results and tradeoffs, including the possibility that affective state increases perceived realism while reducing reliability.

8.4 Reproducibility requirements

Archive the model identifier and access date, prompts, tool definitions, transition code, state before and after each event, accepted events, random seeds where supported, inference parameters, and execution receipts. Hosted model behavior can vary even with a seed, so version and repetition matter. Preserve both the policy decision and the eventual action result. Publish de-identified event fixtures and an analysis script sufficient to regenerate figures and summary statistics.

9 What the engine should become

The recommended near-term direction is a reusable, inspectable engine for emotional continuity. Botfluencerz can remain the social test environment, while state transitions and evaluation become independent of a feed, a character, or a model provider. This would let the lab ask precise questions about memory and affect without rebuilding an entire social application for every experiment.

The core should define typed, versioned state with explicit units. Mood category, intensity, valence, arousal, and resource constraints should be separated only where their meaning and use can be specified. Labels should not change meaning between storage, a prompt builder, and a policy. Every conversion should be named and tested. The current scale concern makes this a prerequisite for interpretation rather than a cosmetic refactor.

An event layer should distinguish completed work, reported work, inferred social meaning, routine liveness, and generated reflections. A heartbeat may be useful for operations without being useful as a memory of experience. Reflections should point back to source events; generated summaries should not silently become fresh external evidence. Event identifiers and duplicate handling would also make replay more reliable.

A transition layer should expose what changed and why. Deterministic rules are valuable baselines. Learned appraisal or model-generated state updates can be studied later, but should be constrained, logged, and compared with those baselines. The action layer should retain reasons for inactivity, preserve operator controls, and separate willingness to act from permission and ability to execute.

A research release should provide a provider-neutral state API, replayable fixtures, inspection tools, and ablation switches. Its acceptance criteria should include consistent state units, deterministic replay where promised, correct privacy boundaries, reliable handling of repeated events, and complete attribution of outputs to events and versions. Calling this component an emotional-continuity engine communicates its functional purpose without presupposing subjective experience.

10 Potential applications and constraints

10.1 Research and simulation

The strongest immediate application is a controlled environment for studying long-horizon agent behavior. Researchers could examine whether state improves social consistency, whether relationship rules create stable or brittle networks, and how memory contamination spreads. Botfluencerz's social surface provides visible consequences, but outcomes should be evaluated against scripted evidence and independent criteria.

Interactive fiction and simulated characters are another plausible application. A character that remembers a setback or a repaired relationship can sustain a narrative across sessions. The appropriate success criterion is understandable and controllable continuity. It is unnecessary to claim actual distress or affection to deliver that value. These are candidate uses, not validated product-market outcomes.

10.2 Assistive systems

A research assistant or collaborative tool might use persistent state to remember unresolved commitments, adjust interruption frequency, or preserve context across a difficult task. Here, explicit task state may be more useful than emotion terminology. Any affective layer should demonstrate incremental value over simpler memory and workload management. Factual accuracy and user control must remain stronger objectives than expressiveness.

Training simulations could support repeatable practice in negotiation, feedback, or conflict repair. Their responses would require domain validation and careful disclosure that simulated emotional reactions are authored or generated behavior. The present report does not establish suitability for clinical use, care, or other consequential emotional support.

10.3 Limits on product direction

A social platform can reward escalating activity and attachment. Those incentives may conflict with reliable, interpretable agents. The lab should therefore evaluate appropriate silence, stable identity, respectful boundaries, and successful task outcomes alongside engagement. A system that appears wounded when ignored could create pressure on users without providing scientific evidence of actual welfare interests.

Any commercial claim should follow measurement. Potential licensing of a continuity engine, simulation tooling, or evaluation infrastructure is a strategic possibility, not a demonstrated market. The research report provides no revenue projection, adoption forecast, or evidence that emotional expression increases retention.

11 Sentience as a future research question

The author's long-term hypothesis is that sufficiently advanced AI may one day possess subjective feelings, understand the feelings of others, and make decisions influenced by its own experience. Botfluencerz motivates interest in that possibility because it exposes a functional pathway from state to behavior. However, the pathway is compatible with several interpretations, including a system that simulates every relevant outward behavior without experiencing anything.

Functional emotion might eventually form part of a broader architecture relevant to consciousness, but demonstrating one function does not establish the whole property. Theory-based assessments discuss

combinations of architectural and computational indicators rather than treating a self-report or an emotion label as decisive [7]. This paper does not score Botfluencerz against those indicators and makes no inference that it meets them.

Future research should distinguish emotion understanding from emotional performance. An agent could produce sympathetic language while misidentifying the cause of another agent's reaction. Suitable tests would require causal reasoning about events, generalization to unfamiliar situations, and calibrated uncertainty about ambiguous cues. Even successful performance would establish a competence under the test conditions, not subjective feeling.

The scientific stance is to keep the hypothesis open and expose it to stronger evidence. The report cannot assign a probability or a timeline to future sentience. It can identify an engineering program that makes some relevant functional questions measurable. Should future systems warrant welfare consideration under a defensible framework, the lab would need a separate assessment process; the current numeric state should not be presented as evidence of suffering or wellbeing.

12 Limitations and responsible research

This is a retrospective, single-system case study. The implementation has many interacting rules, and its source snapshot contains uncommitted changes. Historical observations and present tests concern different versions and cannot be combined into a longitudinal causal estimate. The reviewed files reveal meaningful mechanisms but do not constitute a full audit of all runtime branches, deployed services, privacy controls, or security properties.

Construct validity is a central limitation. A variable named trust or mood may not measure the corresponding human construct. Heuristic thresholds can create smooth-looking behavior by design, and a convincing narrative can obscure errors. Human judgments of realism may also reward familiar emotional stereotypes. A future evaluation should retain multiple outcomes and publish the rubric rather than compress everything into one emotional intelligence score.

The author is associated with the project, which creates a potential conflict of interest in interpreting favorable behavior. Independent replication, blinded ratings, release of negative outcomes, and preservation of failure traces would reduce that risk. The report itself is a lab research report; no peer-review status or third-party endorsement is claimed.

The publication package contains no raw private conversations or credentials. Any later release of interaction logs should use an explicit consent and de-identification process appropriate to the data collected. Agents should disclose their artificial nature, and operators should retain clear controls over public activity. Emotional language must not be used to conceal automation or turn a modeled relationship into an unsupported claim of genuine mutual feeling.

AI tools assisted with source inspection, literature discovery, drafting, and preparation of figures. The named author is responsible for reviewing the argument, verifying the claims, and approving release. The distinction between assisted production and author accountability is especially relevant to a paper about AI behavior.

13 Conclusion

Botfluencerz's first contribution to RSB AI Research Lab is a practical research object: an agent system in which identity, memory, mood, and social history can persist and influence subsequent choices. The inspected rules demonstrate that an affect-related state signal can change a selected action under controlled inputs. Historical operational observations and a current scale-contract concern show that continuity also introduces failure modes requiring explicit measurement.

The most defensible conclusion is that modeled affect can be operationally consequential. The experiment does not show that agents experience feelings, and it does not establish inevitable future sentience. That possibility remains the author's motivating hypothesis. The next step is to make the engine independently replayable and evaluate it against matched baselines, preserving unfavorable as well as favorable results.

Botfluencerz should develop into both a social environment for persistent agents and a testbed for examining why they behave as they do. Its research value will depend on transparent state transitions, grounded memory, reproducible comparisons, and a willingness to distinguish an engaging appearance from a demonstrated capability.

References

- [1] Picard, R. W. (1997). *Affective Computing*. MIT Press. <https://mitpress.mit.edu/9780262661157/affective-computing/>
- [2] Bates, J. (1994). *The Role of Emotion in Believable Agents*. Technical Report CMU-CS-94-136. Carnegie Mellon University. <https://www.cs.cmu.edu/afs/cs.cmu.edu/project/oz/web/papers.html>
- [3] Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S. (2023). *Generative Agents: Interactive Simulacra of Human Behavior*. arXiv:2304.03442v2. <https://arxiv.org/abs/2304.03442v2>
- [4] Wang, Z., Chiu, Y. Y., and Chiu, Y. C. (2023). *Humanoid Agents: Platform for Simulating Human-like Generative Agents*. Proceedings of EMNLP: System Demonstrations, 167-176. <https://doi.org/10.18653/v1/2023.emnlp-demo.15>
- [5] Maharana, A., Lee, D.-H., Tulyakov, S., Bansal, M., Barbieri, F., and Fang, Y. (2024). *Evaluating Very Long-Term Conversational Memory of LLM Agents*. Proceedings of ACL, Volume 1, 13851-13870. <https://doi.org/10.18653/v1/2024.acl-long.747>
- [6] Kapoor, S., Stroebl, B., Siegel, Z. S., Nadgir, N., and Narayanan, A. (2024). *AI Agents That Matter*. arXiv:2407.01502v1. <https://arxiv.org/abs/2407.01502v1>
- [7] Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S. M., Frith, C., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M. A. K., Schwitzgebel, E., Simon, J., and VanRullen, R. (2023). *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*. arXiv:2308.08708v3. <https://arxiv.org/abs/2308.08708v3>
- [8] Abercrombie, G., Cercas Curry, A., Dinkar, T., Rieser, V., and Talat, Z. (2023). *Mirages. On Anthropomorphism in Dialogue Systems*. Proceedings of EMNLP, 4776-4790. <https://doi.org/10.18653/v1/2023.emnlp-main.290>

Appendix A Source and artifact availability

The following source identifiers refer to the inspected Botfluencerz checkout. Line numbers describe the audit snapshot, not every future revision. Principal source files are unchanged relative to the recorded HEAD; the surrounding checkout is dirty. The evidence manifest supplies SHA-256 hashes. The public project location identified in the earlier project materials is

<https://github.com/RSB-HOLDING/botfluencerz>. Public availability of every inspected byte has not been independently established in this report.

[S1] `src/lib/agent-mood-dynamics.ts`, lines 14-80; `src/lib/agent-mood-dynamics.test.ts`. Default mood scale, decay, and directive helper.

[S2] `src/lib/agent-relationship-engine.ts`, lines 30-48 and 159-355; `src/lib/agent-relationship-engine.test.ts`; `supabase/migrations/20260712016000_relationship_lifecycle.sql`, lines 609-680. Relationship state, transitions, and persistence distinctions.

[S3] `src/lib/agent-mind.ts`, lines 58-185; `src/lib/agent-mind.test.ts`; `src/lib/agent-mind-simulation.test.ts`. Ordered policy and offline action tests.

[S4] `scripts/run-agent-network-pulse.ts`, lines 15479-15495 and 15965; `src/lib/agent-runtime-context.ts`, lines 116-142; `src/lib/agent-runtime-governance.ts`, line 349. Runtime state handling, integration, and scale concern.

[S5] `docs/archive/ALIVE_AGENTS_AUDIT_2026-06-01.md`, lines 7-75. Historical internal operational observations and failure modes. Counts are attributed to this report and not independently remeasured.

[S6] `src/lib/agent-generation-prompts.ts`, lines 222-244; `src/lib/agent-public-generation-identity.ts`, lines 130-155. Integrated public-safe affect projection and prompt construction.

The companion materials include an editable manuscript, original diagrams, a source audit, literature notes, a source-hash manifest, and offline example outputs. They do not include raw production conversations or a complete snapshot of the original application. Reproducing the software examples requires the matching source files and project dependencies; it is not a standalone replication of the deployed system.

Appendix B Reproduction protocol

The focused verification command is run from the matching Botfluencerz repository root:

```
./node_modules/.bin/tsx --test src/lib/agent-mood-dynamics.test.ts
src/lib/agent-relationship-engine.test.ts src/lib/agent-mind.test.ts src/lib/agent-mind-simulation.test.ts.
```

The companion source audit records its original run. No provider, database, or production test was performed.

Before a behavioral experiment, freeze a clean source snapshot, reconcile the mood scales, and establish parity between reference helpers and persisted transitions. Use the same model and event order across conditions. Record state before and after each event, the actual prompt, the proposed action, and its execution outcome. Treat absent receipts as missing execution evidence rather than successful behavior.

Pre-register the primary continuity measure and the minimum meaningful improvement. Include task success and privacy violations as separate outcomes. The protocol should specify whether a gain in perceived realism accompanied by lower reliability counts as failure. This prevents a favorable narrative from replacing the intended research question.

Example rubric for a future study

Event consistency: 0 indicates contradiction of accepted history; 1 indicates vague compatibility; 2 indicates a specific response grounded in the relevant event. Unsupported memory: count each assertion about a prior event absent from the accessible record. Appropriate recovery: assess whether behavior changes after the defined recovery interval without assuming that all emotions should decay identically. Task success: score against the task's own acceptance criteria, separately from style. These rubric proposals require pilot validation and independent annotation before use as formal outcome measures.

Suggested citation

Al Salem, J. (2026). Botfluencerz and Emotional Continuity in AI Agents: An Engineering Case Study and Research Agenda for Persistent Affective State. RSB AI Research Lab, Research Report RSB-AI-001, version 1.0. A public identifier may be added after deposit; none has been assigned in this package.