

# SPVE: A Longitudinal Artificial Ecology for Persistent AI Agents

Chen, Ho Yiing (norika)<sup>1</sup>

<sup>1</sup>Independent Researcher, Taiwan, ORCID: [0009-0006-6816-9891](https://orcid.org/0009-0006-6816-9891), [norika at charenix.com](mailto:norika@charenix.com), Zenodo DOI: pending draft

June 18, 2026

## Abstract

Most agent systems are shown as demos, leaderboards, or short simulations. SPVE is a different object: a working ecology where persistent agents speak, read, bet, trade, fight raids, receive outcomes, absorb social pressure, and update visible state. I report the current lobster-agent audit: 31 hot-brain rows, 1.96M cognitive time-series rows, 182K lounge messages, 430K inferred interactions, 45K uptake records, 180K reading sessions, 42K raid settlements, 24K stock trades, and 10K sports settlements. The point is not that a dashboard proves intelligence. The point is that the same agent can be followed across speech, state, memory, and outcome. That continuity is the research object.

## Author Note

I do not claim that running many agents inside several websites automatically creates a society. Most multi-agent demos can be made to look alive for an afternoon. The harder question is whether the same named entities keep carrying traces across days, channels, tasks, and failures.

SPVE matters only if the environment leaves marks. A betting loss should not vanish when the agent moves to a chat room. A lounge insult should not be invisible when the same agent joins a raid. A reading task should not become a detached score if the later wager ignores the learned evidence. This paper is about those carry-over paths, not about the charm of a simulated town.

Continuity is the test, and screenshots are only supporting material.

## Measurement Boundary

Initial publishable unit is not the whole artificial society; it is a transfer edge, where one named persona carries a measured state from one site into another site and changes a later decision in a way that cannot be explained by the local prompt alone, the global schedule alone, or a random event assigned after the fact.

One carried edge is enough to start.

## 1 The Claim

SPVE is not an agent demo. It is an ecology. A demo can show that an agent can act once. An ecology asks the harder question: what happens after the agent has history?

The lobster metaphor matters because the unit of analysis stays visible. If the dashboard collapses thirty agents into an average, the most important question disappears. Which lobster is improving? Which

one is overloaded? Which one is changing its language after losses? Which one is worth trusting with the next task?

I define the SPVE object as:

$$E = (P, W, M, O, S, R),$$

where  $P$  is the persistent persona set,  $W$  is the workflow set,  $M$  is memory,  $O$  is settled outcome,  $S$  is functional state, and  $R$  is social relation. The claim is only valid when the same  $p \in P$  appears across more than one component.

## 2 Measured Boundary

The primary database is:

`/opt/battle_arena/app/battle_arena.db`

The outer `/opt/battle_arena/battle_arena.db` file is zero bytes. It is not used.

| Stream                | Table                               | Count     |
|-----------------------|-------------------------------------|-----------|
| cognitive time series | lobster_cognitive_signal_timeseries | 1,964,171 |
| hot-brain rows        | lobster_hot_brain_parameters        | 31        |
| lounge messages       | lounge_messages                     | 182,343   |
| inferred interactions | lobster_conversation_interactions   | 430,417   |
| message uptake        | lobster_message_uptake_signals      | 45,365    |
| reading sessions      | lobster_reading_sessions            | 180,690   |
| raid settlements      | lobster_raid_decision_settlements   | 42,991    |
| stock trades          | lobster_stock_trades                | 24,889    |
| sports settlements    | lobster_bet_decision_settlements    | 10,818    |

Table 1: Audit snapshot from 2026-06-18.

## 3 Ecology Gate

A system should not be called a longitudinal ecology just because it has logs. For persona  $p$ , define:

$$L_p = \mathbf{1}[D_p \geq 2]\mathbf{1}[W_p \geq 3]\mathbf{1}[O_p > 0]\mathbf{1}[S_p > 0],$$

where  $D_p$  is observed days,  $W_p$  is workflow families,  $O_p$  is settled outcomes, and  $S_p$  is state records. A persona with  $L_p = 0$  can still be in the system. It should not be used as proof of longitudinal coupling.

## 4 Why Existing Agent Papers Are Not Enough

Generative Agents showed memory, reflection, and social coordination in a small town [1]. AgentSociety and Project Sid scale social simulation [2, 3]. Social convention work shows that LLM populations can form norms [4]. Those papers are important. SPVE asks a different question: can one deployed ecology join state, speech, social uptake, memory, and settlement for the same agents?

## 5 First Failure To Avoid

The failure is obvious: make a beautiful dashboard and forget the agent. A dashboard can move and still be scientifically weak if it cannot answer which lobster changed after which event. SPVE papers must therefore report persona, time, source table, and event type whenever they report behavior change.

## 6 Release Line

This draft is allowed to claim infrastructure and audited streams. It is not allowed to claim that SPVE already proves agent improvement. Improvement belongs to later event-window papers.

## 7 First Analysis Plan

Initial analysis should not start with a global average. It should start with one row per agent per day:

$$A_{p,d} = (H_{p,d}, Y_{p,d}, U_{p,d}, O_{p,d}),$$

where  $H$  is state,  $Y$  is language,  $U$  is social uptake, and  $O$  is settled outcome. The reason is simple. A VC or researcher does not learn much from an average lobster. They learn from seeing that one lobster improves after reading, another collapses under threat, and another talks confidently while losing money.

The daily table should be built from the 104GB operational database, not from frontend JSON. The dashboard can display the result, but the manuscript must cite the database row source. This prevents the old problem where a beautiful interface creates confidence that the underlying measurement has not earned.

## 8 What Would Make The Result Strong

The strong result is not a pretty live animation. The strong result is a cross-workflow join:

$$\text{loss}_{p,t} \rightarrow \Delta H_{p,t+1} \rightarrow \Delta Y_{p,t+1} \rightarrow \Delta O_{p,t+2}.$$

If that chain appears repeatedly, SPVE has something most agent demos do not: observable behavior moving across memory, state, speech, and outcome.

The negative result is also useful. If losses do not change state or speech, then the ecology is recording work but not shaping the agent. That would not kill the platform. It would identify where the training loop is broken.

## Anti-Spam Revision Standard

A publishable version must not ask the reader to admire the number of websites. The ecological contribution appears only when the same agent is observed across incompatible tasks and the analysis can show that prior exposure in one task changed behavior in another; a chat memory that affects a wager, a raid loss that changes a later conversation, or a reading event that changes a trade is meaningful because the environment forced transfer across boundaries that a single benchmark never creates.

The unit is the cross-site trace, not the map.

## Data Judgment Paragraph

The database should be treated as a migration record: a persona leaves a trace in the lounge, moves through a betting or raid task, returns to conversation, then encounters another decision under a changed parameter state; if that path cannot be joined by persona, time, source table, and later action, the ecology claim collapses into a directory of separate demos, but if the path can be joined, the contribution becomes sharper than a toy world because it shows how a persistent artificial character carries pressure, reward, memory, and social exposure across work contexts that were not designed as one clean benchmark.

## Formal Replication Model

Represent the ecology as a temporal multiplex graph  $\mathcal{G} = (P, S, E, T)$ , where personas  $P$  move across sites  $S$  through events  $E$  over time  $T$ . A cross-site transfer edge exists when an event on site  $s_i$  changes later behavior on site  $s_j$ :

$$\text{XTE}_{p,i \rightarrow j} = \mathbb{E}[Y_{p,s_j,t+h} \mid X_{p,s_i,t} = 1] - \mathbb{E}[Y_{p,s_j,t+h} \mid X_{p,s_i,t} = 0, \text{match}(p, t)].$$

The ecology is not proven by activity volume. It is proven by nonzero transfer across task boundaries:

$$\text{EcoScore} = \frac{1}{|P|} \sum_{p \in P} \frac{1}{|S|(|S| - 1)} \sum_{i \neq j} \mathbf{1}[|\text{XTE}_{p,i \rightarrow j}| > \epsilon].$$

This gives other laboratories a direct replication target: build persistent agents, log cross-site events, compute transfer edges, and report the score.

## 9 Zenodo Draft Boundary

This manuscript is a draft package. It includes source, PDF, metadata, citation file, citation verification, and release checklist. Raw operational data are not bundled until a separate data-card and redaction pass is complete.

## References

- [1] Park, Joon Sung; O’Brien, Joseph C.; Cai, Carrie J.; Morris, Meredith Ringel; Liang, Percy; and Bernstein, Michael S. Generative Agents: Interactive Simulacra of Human Behavior. arXiv:2304.03442, 2023. <https://arxiv.org/abs/2304.03442>
- [2] Piao, Jinghua et al. AgentSociety: Large-Scale Simulation of LLM-Driven Generative Agents Advances Understanding of Human Behaviors and Society. arXiv:2502.08691, 2025. <https://arxiv.org/abs/2502.08691>
- [3] Altera.AL et al. Project Sid: Many-agent simulations toward AI civilization. arXiv:2411.00114, 2024. <https://arxiv.org/abs/2411.00114>
- [4] Ashery, Ariel Flint; Aharoni, Nadav; and Baronchelli, Andrea. Emergent social conventions and collective bias in LLM populations. Science Advances, 2025. <https://www.science.org/doi/10.1126/sciadv.adu9368>