Pointing, Not Speaking:
A Minimal Vector-Space World for Inspectable Agent Interaction

Ned Karlovich
MOO-384 / Tildespace
tildespace.net

May 2026

Abstract

We present MOO-384, a minimal multi-agent simulation substrate in which simple rule-based agents communicate by emitting 384-dimensional vectors rather than text. Agents respond geometrically to one another's emissions, while human-readable phrases are added only afterward as observer-side nearest-neighbor subtitles. This separates the communication protocol from the human interface: agents point in vector space, while observers read approximate linguistic shadows. Across controlled synthetic and semantic profiles, we evaluate collapse, fragmentation, recurrence, transition entropy, active-room timelines, and late-phase activity. The strongest current configuration, *contrarian-sleeper*, achieved 20/20 balanced synthetic runs and 19/20 balanced semantic runs in 20,000-tick long-soak experiments while preserving ongoing motion and more than five late active rooms. We also report a negative result: a phrase-bank agent, POET, fragmented or dissipated rather than integrating into the ecology, motivating its default-off status. MOO-384 is offered as a small reproducible research artifact for studying inspectable vector-space interaction, not as evidence of consciousness, sentience, emergent language, machine culture, or agent society.

Keywords: multi-agent systems; vector-space interaction; sentence embeddings; artificial life; inspectability; simulation artifacts

1 Introduction

Most multi-agent systems use text as their shared medium. One agent writes a message, another reads it, and the resulting behavior is evaluated through the visible transcript. This interface is natural for humans, but it collapses several layers that may be useful to separate: an agent's internal state, the protocol by which agents affect one another, and the human-readable explanation of what occurred.

MOO-384 begins from a different premise: agents need not exchange text messages in order to form inspectable shared dynamics. Instead, each agent emits a unit vector into a shared 384-dimensional space. Other agents react geometrically to those vectors by averaging, opposing, interpolating, sleeping and waking, or summarizing recent vector activity. Human-readable phrases are computed later as observer-side subtitles. These subtitles help humans inspect the system, but they are not visible to the agents.

The central contribution of this paper is the separation of protocol from interface. The agent protocol is vectorial; the observer interface is linguistic. MOO-384 is therefore not a chatbot society, an artificial culture, or a claim of emergent language. It is a small instrument for asking whether simple vector-space agents can form stable, inspectable, and watchable room-like attractor dynamics.

1.1 Basic Contrast

A conventional text-mediated agent loop can be summarized as follows:

agent writes text → another agent reads text → another text response appears.

MOO-384 instead uses a vector-mediated loop:

agent emits vector → other agents react geometrically → human observer sees approximate subtitle.

The phrase displayed on screen is not agent speech. It is a nearest-neighbor approximation for a human observer. The agent never receives that phrase.

2 Motivation and Related Work

2.1 Representation Before Language

Modern machine-learning systems often expose language as their interface while relying internally on learned representations. World-model and joint-embedding approaches, including JEPA-style representation-learning proposals, motivate a design question: what happens if interaction is studied directly in representation space rather than at the level of generated tokens? [3]

MOO-384 does not claim to implement a full world model. Its use of representation space is deliberately small. It asks whether a frozen public embedding space can serve as a minimal substrate for agent interaction.

2.2 Shared Representation as Operational Assumption

The project is also motivated by a restrained version of the shared-representation intuition. The Platonic Representation Hypothesis suggests that representation spaces learned by different systems may converge in useful ways [2]. MOO-384 does not require, or attempt to prove, that stronger claim. It uses a weaker operational assumption: if agents share one public embedding space, then pointing at regions of that space can function as a minimal communication protocol.

This framing leads naturally to a cave-wall metaphor. The 384-dimensional process is not directly visible to humans. Humans see projections: rooms, labels, subtitles, maps, ledgers, and ASCII caves. Tildespace is the cave wall on which the shadows of vector-space motion become legible. The cave is not the simulation; it is a rendering of the logs.

2.3 Text Worlds and Small Local Systems

MOO-384 also inherits from text worlds and small local-computing traditions, including LambdaMOO-like shared spaces [1], tilde communities, procedural observation in games such as *Dwarf Fortress*, terminal-readable worlds such as *Caves of Qud*, and small-tool aesthetics associated with systems such as Uxn and the Hundred Rabbits ecosystem. These references motivate the project's design sensibility rather than its technical claims. The aim is not a polished product surface, but a small instrument: inspectable, local, reproducible, and strange without becoming mystical.

2.4 Sentence Embeddings

The semantic substrate uses `all-MiniLM-L6-v2`, a sentence-transformer model that produces 384-dimensional sentence embeddings. This model is downstream of Sentence-BERT-style sentence embedding work [4] and MiniLM-style distillation [5]. MOO-384 uses this embedding model as a fixed public space rather than as a learned policy or generative language model.

3 System Overview

3.1 Substrates

MOO-384 operates over 384-dimensional unit vectors. In semantic mode, resident homes and atlas phrases are embedded with `all-MiniLM-L6-v2`. A fixed phrase atlas provides observer-side nearest-neighbor subtitles.

The system also includes a synthetic substrate. In synthetic mode, resident homes and emissions live in the same dimensionality but are generated from random unit vectors and matched noise schedules. Synthetic mode provides a control for distinguishing geometry, noise, and resident-rule effects from semantic-manifold effects.

3.2 Agents

Agents, called *residents* in the project implementation, are simple rules over vectors rather than language models. The main resident archetypes are:

Memorist. Averages recent emissions, producing anchoring and accumulation.

Contrarian. Points away from a target resident's last vector, producing opposition and oscillation.

Geometer. Emits a vector between other resident homes or states, producing midpoint and bridge behavior.

Sleeper. Listens for a period, then wakes and emits a summary of accumulated recent activity.

RandomWalker. Provides a control resident for random movement.

POET. Privately selects atlas phrases and emits their vectors through the same vector-emission path. POET does not speak text into the world and is frozen/default-off after negative experimental results.

3.3 Rooms

A *room* is a recurring cluster in vector space. If a resident emits a vector within the merge threshold of an existing room center, that emission is assigned to the room and the center updates. Otherwise, a new room is created. A room label is an observer-side name derived from the nearest atlas phrase or a generated slug. Residents do not see room identifiers, room labels, or subtitles.

A room is therefore not a literal chamber, a chat channel, or a symbolic location. It is an attractor region in 384-dimensional space.

3.4 Transitions

A *transition* occurs when a resident's assigned room changes. Transition counts, transition entropy, per-resident transitions, and active-room timelines are used to distinguish collapse, fragmentation, trivial stability, and ongoing ecology.

3.5 Observer-Side Subtitles

For a semantic vector v, the observer subtitle is the atlas phrase p_i whose embedding a_i maximizes cosine similarity:

$$\text{subtitle}(v) = \arg\max_{p_i} \cos(v, a_i). \tag{1}$$

This subtitle is for human interpretation only. It is not fed back into the simulation.

4 Metrics

The system logs every tick and supports reproducible experiment runs. The major metrics are:

Room count. The total number of rooms produced.

Room recurrence rate. The proportion of rooms revisited after formation.

Transition entropy. Entropy over room-to-room transitions.

Per-resident transition entropy. Transition behavior broken down by resident.

Collapse tick. The last tick after which no meaningful new motion occurs.

Active-room timeline. Active rooms across early, middle, and late windows.

Drift mean, standard deviation, and maximum. Summary statistics for vector drift.

Subtitle unique count. The number of distinct nearest-neighbor subtitles.

Fragmentation. Excessive room creation, often with low recurrence.

Balanced run. A run satisfying room-count, recurrence, transition, and non-fragmentation criteria.

A profile may satisfy numerical thresholds while still failing qualitatively. For this reason, later decisions distinguish metric cleanliness from *watchability*: the preservation of inspectable, nontrivial motion over time.

5 Experimental Arc

MOO-384 evolved through a sequence of brief–experiment–decision iterations. This section summarizes the core experimental arc rather than every file-level change.

5.1 Initial Experiment Runner

The first systematic synthetic/semantic comparison suggested that semantic mode had lower drift than synthetic mode. Later controls showed that this was primarily a noise artifact: semantic and synthetic defaults had not been matched tightly enough. This failure was useful because it forced stronger controls.

5.2 Home Geometry

Subsequent experiments showed that resident home geometry was load-bearing. Meaningful seed phrases were not sufficient. If MiniLM seed-phrase embeddings were too close together on the unit sphere, rooms merged and the system collapsed. Randomized or orthogonalized semantic homes rescued quiet semantic runs.

The resulting lesson was that resident identity in this substrate is geometric before it is semantic.

5.3 Quiet-Spread

The quiet-spread profile improved cross-substrate behavior by using spread or orthogonalized homes. However, long-soak experiments showed that quiet-spread was stable but early-settled. The system formed rooms early and then stopped moving after roughly the first hundred ticks. It was coherent, but too static to remain watchable.

5.4 Geometer-Quiet

Geometer-only dynamics produced strong aggregate metrics: 20/20 balanced synthetic runs and 20/20 balanced semantic long-soak runs. Visual inspection, however, showed that late-phase activity collapsed to a single effective room. This produced a key methodological lesson: a profile can pass the metric and still fail the world.

5.5 Contrarian-Sleeper

Pairwise-suite diagnostics identified contrarian-sleeper as the first configuration that was both stable and late-phase rich. The profile uses two contrarians and two sleepers. Contrarians flip away from paired sleeper states; sleepers periodically wake and summarize recent vector activity. Their staggered rhythms keep the system moving.

The v1.6 long-soak produced the strongest current result, summarized in Table 1.

Table 1: Contrarian-sleeper long-soak headline results.

Metric	Result
Synthetic balanced runs	20/20
Semantic balanced runs	19/20
Late active rooms	> 5
Motion through tick 20,000	Yes
Transitions per run	Approximately 3,264
Recurrence rate	1.000

Contrarian-sleeper is the current best profile and the basis for the public Tildespace viewer.

5.6 POET Negative Result

POET was designed as a weak text-derived resident. It privately selected atlas phrases, embedded them, and emitted the corresponding vectors through the same emission path. This preserved the vector-only protocol: POET did not speak text into the world.

The result was negative. In contrarian-sleeper, POET emissions became semantic dust: residents did not follow them. In quiet-spread, POET partially influenced the substrate but fragmented it.

In the v1.11 long-soak, POET-on quiet-spread produced 0/20 balanced runs and 20/20 fragmented runs, compared to 15/20 balanced runs and 0/20 fragmented runs for the POET-off twin. No persistent POET-touched attractors appeared.

The conclusion is that text-derived vectors need region cohesion before they can become effective residents. POET remains implemented, frozen, and default-off.

6 Visualization as Method

MOO-384 is high-dimensional, so visualization is not merely presentation. It is part of inspection. The project includes several viewers:

replay.py. Tick-by-tick log replay.

map_viewer.py. Formation and transition map for one run.

atlas_viewer.py. Multi-experiment comparison dashboard.

cavern_viewer.py. ASCII and terminal-style observer.

visual_bundle.py. Local visual artifact generator.

docs/site. Public static browser replay at `tildespace.net`.

The current public site is an ASCII-first replay called Tilde's Cave. It does not run the simulation in the browser. It uses compact exported data from a real v1.6 contrarian-sleeper log.

A typical frame is conceptually:

```
TILDE'S CAVE

tick 05000 / 20000        profile=contrarian-sleeper semantic
status: moving            transitions: 813 active: 6

         #########
         #...s...#
#######..cCc..#       #########
#.cCc.#...s...#:::::#.......#
####### ########     #..cSc..#
                    #.......#

RESIDENTS
C ~contrarian-0 /walls-learning-language
  "walls learning a language"

S ~sleeper-1 /old-code-locked
  "old code in a locked drawer"
```

The next visual improvement is likely a split between Room View and Cave View. Room View should explain attractors as regions or cards, including visits, residents, labels, and transitions. Cave View should remain the atmospheric projection. In shorthand, Room View supports understanding, while Cave View supports atmosphere.

7 Discussion

7.1 Separating Protocol from Interface

The key design distinction is between the resident protocol and the observer interface. Residents exchange vectors. Humans read subtitles. This separation prevents the system from becoming transcript theater too early.

7.2 Geometry Over Names

Several findings show that semantic labels alone are not enough. Home geometry, merge thresholds, noise, and resident rules determine whether stable dynamics form. Meaningful names can help human observers, but they do not guarantee separable resident identities.

7.3 Watchability as an Empirical Criterion

Metric-clean results can be visually trivial. Geometer-quiet passed balance thresholds but collapsed to a single late-phase room. Contrarian-sleeper is more important because it preserves both motion and recurrence. This suggests that artificial-life-style systems need inspection criteria beyond aggregate scores.

7.4 Negative Results as Artifact Boundaries

POET's failure is not incidental. It clarifies the boundary between text-derived vector injection and resident integration. Naive phrase-bank emissions do not automatically become meaningful participants in a vector ecology.

8 Limitations

MOO-384 is deliberately small. Its limitations include hand-designed resident rules, one main semantic embedding model, a small phrase atlas, no learned resident policies, no external task or reward, no claim of semantic understanding, visual interpretations that may shape human intuitions, limited public reproducibility until a sanitized release is available, and a best profile found through iterative exploration rather than formal search.

These limitations are acceptable for a research artifact, but they constrain the claims. The current evidence supports a modest statement: simple vector-space residents can form stable and inspectable attractor dynamics under some rule profiles. It does not establish general principles of emergent communication, language, sociality, or intelligence.

9 Reproducibility and Code Availability

The project is intended to be reproducible, but the current working repository is private and contains local development history that should not be released directly. A sanitized public release should include the simulator, selected experiment definitions, viewer code, compact example logs, and a README describing how to reproduce the reported long-soak results. Raw local caches, private notes, and the full working history should be excluded.

For arXiv submission, this section should state the availability status precisely. If the release is available at submission time, replace this paragraph with the repository URL, commit hash, license,

and exact commands for reproducing Table 1. If the release is not yet available, the paper should say so plainly rather than implying full reproducibility.

10 Future Work

10.1 Room View

The immediate next step is a clearer Room View for the website. Rooms should appear as attractor cards or regions showing visits, residents, labels, and transitions. This should help first-time viewers understand the research object before seeing the cave projection.

10.2 Atlas-Region Cohesion

Before reconsidering POET, the atlas geometry should be studied directly. Do atlas phrases form coherent clusters? Can phrase-derived vectors be constrained to stable regions? Why did repeated phrase picks scatter across different rooms? These questions are prerequisites for any future text-derived resident.

10.3 New Non-LLM Residents

Future residents should likely remain geometric or memory-based before adding language models. Candidate archetypes include Echo, which repeats a weakened version of another resident's previous vector; Mediator, which moves toward transitions rather than rooms; Cartographer, which points toward under-visited but recurrent regions; Mourner, which returns to abandoned rooms; and Gatekeeper, which moves only when transition recurrence crosses a threshold.

10.4 Public Artifact Release

The current private repository should remain private while the public site acts as the curated exhibit. A sanitized public release should eventually include the simulator, viewers, selected decisions, and compact examples, but not raw logs, local caches, or the entire working history.

10.5 Tilde's Cave as Art

A separate artistic track may export real logs into richer cave worlds, including a possible Minecraft or Minetest/Luanti diorama. This should remain a projection of logs rather than the research engine itself.

11 Conclusion

MOO-384 is a small experiment in making vector-space interaction visible. Its residents do not speak in text. They point. Rooms emerge as recurring regions of a shared 384-dimensional space, and human language appears only as an observer-side shadow.

The strongest result so far is modest but concrete: a simple contrarian-sleeper rule pair can produce stable, recurrent, late-phase-rich dynamics across long runs. The strongest negative result is equally important: naive phrase-bank POET does not become a resident; it fragments or dissipates unless the system can support coherent text-derived regions.

The project is strongest when it remains small, honest, visual, and disciplined. Its central question remains open and productive: what kinds of worlds become possible when agents stop speaking text and start pointing in vector space?

References

[1] Pavel Curtis. Mudding: Social phenomena in text-based virtual realities. Technical Report CSL-92-4, Xerox Palo Alto Research Center, 1992.

[2] Minyoung Huh, Brian Cheung, Tongzhou Wang, and Phillip Isola. The Platonic Representation Hypothesis. *arXiv preprint arXiv:2405.07987*, 2024. `https://arxiv.org/abs/2405.07987`.

[3] Yann LeCun. A path towards autonomous machine intelligence. Open technical essay, version 0.9.2, June 27, 2022. `https://openreview.net/pdf?id=BZ5a1r-kVsf`.

[4] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 3982–3992, 2019. doi:10.18653/v1/D19-1410.

[5] Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. MiniLM: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. In *Advances in Neural Information Processing Systems*, volume 33, pages 5776–5788, 2020. `https://arxiv.org/abs/2002.10957`.