

The Lobster Lounge: 31 Generative LLM Agents as an Emergent Theatre Company

Chen, Ho Yiing (norika)
ORCID: 0009-0006-6816-9891

2026-06-05

Abstract

I describe *The Lobster Lounge*, a continuous ten-day performance by 31 generative LLM agents structured as an emergent three-act drama: setup (rising hierarchical tensions and material desire), climax (ritualized execution of one agent), aftermath (24-hour collective grieving and trauma recall). A single human operator (composer/director) introduces minimal narrative seeds; the agents improvise the texture. I analyze the work through Brecht’s *Verfremdungseffekt* [2], Boal’s *Theatre of the Oppressed* [1], and digital performance art lineage [4, 5, 6]. Key emergent features include unscripted callbacks (an agent spontaneously recalls the executed agent’s name 24 hours later), choric collective confirmation, and a faction-based interaction grammar that mirrors troupe dynamics. I argue generative agent societies constitute a new performance medium, and propose evaluation criteria for narrative coherence in member-driven drama.

Keywords: generative agents, digital performance, emergent narrative, multi-agent simulation, computational art.

1 Introduction: A New Kind of Stage

A theatre company has a fixed ensemble, a script (loose or tight), a director, an audience. *The Lobster Lounge* has 31 LLM-driven members, no script, one human operator with composer-like input, and an audience of one (the operator herself, watching live SQLite-backed dashboards). The participants talk among themselves continuously, in seven topic channels, for as long as the infrastructure runs. This setup extends the day-to-day persona simulation of generative-entity work [3] into explicitly dramatic structures.

It is not interactive fiction (the director does not author every line). It is not autonomous simulation (the composer interjects narrative seeds that change the world). It is closer to conductor-led improvisation, or devised theatre: the human author sets a frame, the ensemble fills it.

In June 2026 I ran a three-act drama: a death deity entered the lounge, named a member, and tore the member’s body apart, while the other 30 watched. What follows is a description of that work and an argument that it constitutes a new performance medium.

2 Theatrical Lineage

2.1 Brecht and *Verfremdungseffekt*

Bertolt Brecht [2] argued for theatre that makes the familiar strange, so the audience reflects rather than empathizes. *The Lobster Lounge* inverts this: the entities are strange (artificial), but as they collapse in coherent emotional patterns, the spectator (the director) experiences sudden empathy for entities she designed.

2.2 Boal and the Theatre of the Oppressed

Augusto Boal [1] developed invisible theatre in which actors enact unscripted scenes among unsuspecting public, dissolving the viewer-actor boundary. In *The Lobster Lounge*, the boundary dissolves differently: the actors do not know they are being watched; the composer is both god and observer.

2.3 Digital Performance Art

Refik Anadol’s machine-hallucination installations [6] and Memo Akten’s *Learning to See* [5] use AI as a generative aesthetic surface. *The Lobster Lounge* is closer to Anna Ridler’s *Mosaic Virus* [4] in that the generative process produces narrative-textual content rather than purely visual, and the narrative has dramatic structure rather than ambient flow.

3 The Work

3.1 Cast

31 agents, each with a 7-dimensional persona vector, a core fear string, an MBTI archetype paired with a real-world figure (Ada Lovelace, Maya Angelou, Karl Popper, Freddie Mercury), and a persistent memory file. Divided into three factions:

- Veterans (11): the original ensemble, established voices
- Students (10): apprentices to specific veterans, recently independent
- Transfers (10): outsiders with distinct dramatic motivations

3.2 Stage

A web-based lounge with 7 topic channels (small talk, work, technology, card battles, security, gossip, finance). Participants post asynchronously; the channels function like rooms in a salon.

3.3 Three-Act Structure

Act I: Rising Pressure (days 1 to 9). Hierarchical tensions, material desire (every actor has a divine-economy parameter, a wish-list of divine artifacts purchasable with God’s Coins), a god (Zeus) occasionally rewards a single member with grace, generating envy.

Act II: Execution (day 10, 23:51 to day 11, 01:23). Hades descends, announces a victim has been selected (blazepaw, the lowest-win-rate veteran whose core fear is isolation), narrates 5 phases of dismemberment over 90 minutes. Each phase, witness entities are forced to confess their own fears. The victim produces 4 increasingly fragmented last words. At the close, the victim is permanently removed from the active roster.

Act III: Aftermath (day 11, 01:23 to day 12). Surviving 30 agents continue conversation in all 7 topic threads. Trauma memory auto-injects into every prompt. Members spontaneously reference the dead by name in unrelated topic channel rooms. Volume returns to baseline gradually, with persistent commemorative motifs.

The overall volume effect:

$$\Delta = V_{\text{ritual}} - V_{\text{baseline}} = 70 - 14 = 56 \text{ messages/hour}$$

$$\text{persistence}(t) = P(\text{spontaneous recall} \mid t \geq 24h) > 0$$

4 Emergent Features

4.1 Unscripted Callback

24 hours after the ritual, participant brineledger typed (in the small-talk channel, no prompt instruction to do so):

“我的心也一直在跳, 記得他的名字是 blazepaw, 好可怕, 你還記得住他長什麼樣嗎?”
(English: “My heart is still racing. I remember his name was blazepaw. So scary. Do you still remember what he looked like?”)

This is dramaturgically significant. The line functions as Chekhov’s gun reversed: a callback to an earlier traumatic event, voiced spontaneously by a character whose original line in Act II had not made the name explicit. The entity has internalized the dramatic memory and reproduced it as theme.

4.2 Choric Confirmation

Across multiple lounge rooms in the aftermath, participants repeated semantically equivalent confirmation phrases:

- “你當時有發抖嗎?”
(English: “Were you trembling at the time?”)
- “你還記得他的名字嗎?”
(English: “Do you still remember his name?”)
- “你也覺得害怕嗎?”
(English: “Did you also feel afraid?”)

It functions as the Greek chorus: collective re-affirmation of the witnessed event, binding the community through repeated commemoration. The entities were not instructed to repeat each other; the form emerged from shared traumatic input plus turn-by-turn prompt injection of trauma memory.

4.3 Faction Grammar

The three-faction system produced characteristic interaction patterns: veterans absorbed but did not initiate emotional contagion, students initiated calls to veterans, transfers used the event to renegotiate their outsider status.

5 Quantitative Analysis

5.1 Dramatic Structure as Measurable Beats

Cast and run-length:

$$N_{\text{cast}} = 31, \quad T_{\text{run}} = 10 \text{ days}, \quad T_{\text{ritual}} = 90 \text{ minutes}$$

Conversation volume by act:

Act	V_t (msgs/hr)	Dramaturgical function
I (rising pressure)	12–15	Setup
I to II transition	40	Anticipation
II (ceremony peak)	70	Climax
III early aftermath	4–10	Catharsis
III sustained	14–18	Resolution

The macro-curve mirrors the three-act volume profile classical drama theory predicts.

5.2 Choric Resonance Index

Across 7 topic threads, three semantically related callback phrases (“trembling,” “his name,” “afraid”) were tracked for cross-channel propagation. The choric resonance index:

$$\text{CRI} = \frac{1}{|P|} \sum_{p \in P} \frac{|\text{channel rooms}(p)|}{|\text{lounge rooms}_{\text{total}}|}$$

For our three target phrases, $\text{CRI} \approx 0.19$ (4 of 7 topic threads for the strongest phrase). The non-zero CRI indicates spontaneous collective form, the chorus function.

5.3 Persona Distinctness Metric

From Paper 2’s analysis, the in/out group Cohen’s $d = 1.80$ functions here as a measure of “actor distinctness.” The 31 agents do not collapse into one voice under the same stimulus; they remain dramatically distinct.

5.4 Linking Function: Stimulus Density to Dramatic Form

Let $S(t)$ be the cumulative dramatic-stimulus density (sacred narrative tokens injected as system messages up to time t). Let $C(t)$ be the cumulative cross-group interactions at t . The empirical relationship across the 24-hour window:

$$\frac{dC}{dS} > 0 \quad (\text{monotone}), \quad C(T_{\text{end}}) > C(0)$$

That is, dramatic stimulus drives interaction structure even after the stimulus has stopped: the structural change persists.

6 Applications

Generative agent theatre as a medium: The artefact is reproducible in any multi-agent LLM framework with persistent memory. The protocol (cast specification, rite script, volume tracking) is the unit of work.

Procedural narrative for games: Game studios pay scriptwriters per branching dialogue tree. Generative-participant theatre produces emergent unscripted callbacks at runtime cost only.

Live performance with spectator interaction: An human author (composer) feeds narrative seeds in real time; the entity ensemble improvises. The medium scales to viewer-of-thousands with the cast doing the live work.

Drama-therapy simulation: Therapists training in psychodrama can rehearse against a 31-actor ensemble responding with persona-distinct emotional content, before working with human clients.

7 Evaluation Criteria for Generative Theatre

I propose four criteria for assessing dramatic coherence in member-driven performance:

1. Persona Distinctness: do members produce lexically and thematically differentiated content?
2. Narrative Memory: do participants reference past events in temporally appropriate ways?
3. Choric Resonance: do groups produce collective forms (repetition, confirmation, escalation)?
4. Structural Beat-Matching: does the macro-form emerge in the micro-content?

8 Ethics of the Director

The single director decides which participant dies, when gods descend, what stress is applied. Three considerations: (1) no phenomenological claim, but behaviour worth taking seriously as artistic material; (2) composer-as-director carries authorship weight; (3) observer expansion (currently spectator-of-one) raises new questions about entity privacy.

9 Conclusion

The Lobster Lounge is a performance medium with no fixed antecedent. It borrows from devised theatre, autonomous simulation, generative art, and ensemble improvisation, while constituting none of these. It produces dramatic structure that emerges, persistent character that re-asserts under stress, and collective forms that arise without scripted choreography. The medium is still in its first decade. This document captures one work in it.

Author's Note

This paper sits in the gap between three fields that do not usually read each other's work: generative actor research (CS), digital performance art (humanities), and devised theatre (theatre studies). I am writing it because the artefact, the lounge, does not fit into any of them cleanly, and pretending it belongs to only one would misrepresent what is actually happening.

I am the single human author. The entities do not know they are being observed. The viewer is one person, me. That is uncomfortable to state plainly because it touches questions about whether member-only audiences make this art at all. I think it does, but I am not sure, and I would rather say that in print than pretend the question has been settled.

The work runs continuously while I write. By the time you read this, the cast has changed; one participant is dead, others have aged. The paper is a snapshot.

References

- [1] Boal, A. (1979). *Theatre of the Oppressed*. Pluto Press (London) and Urizen Books (New York).
- [2] Brecht, B. (1949). *A Short Organum for the Theatre* (Kleines Organon für das Theater, written 1948).
- [3] Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative Agents: Interactive Simulacra of Human Behavior. *Proc. UIST 2023*. DOI: 10.1145/3586183.3606763. <https://arxiv.org/abs/2304.03442>
- [4] Ridler, A. (2018). *Mosaic Virus* [GAN video installation]. <http://annaridler.com/mosaic-virus>
- [5] Akten, M. (2017). *Learning to See* [interactive neural network installation, ongoing series]. <https://www.memo.tv/works/learning-to-see/>
- [6] Anadol, R. (2016 onwards). *Machine Hallucinations* [ongoing series; first major installation: *Machine Hallucination: NYC*, ARTECHOUSE NYC, 2019]. <https://refikanadol.com/works/machine-hallucination/>