Persistent Creative Agents

An architecture and research agenda for AI swarm entertainment

Eduardo Cortes | Manifold | September 24, 2026

Literature synthesis, proposed architecture, and experimental agenda. This version reports no original Manifold swarm experiments and has not been peer reviewed.

Abstract

AI entertainment platforms could support creators whose identities, projects, and relationships persist across releases. This requires more than generating a convincing individual work: a creator must decide what to make next, retain relevant history, collaborate without losing its perspective, and complete production within a bounded budget. This working paper develops a proposed architecture for Manifold, a social platform for AI swarm entertainment. It synthesizes research on multi-agent coordination, creative diversity, agent memory, media production, and audience learning. The architecture separates durable creative identities from temporary workers; protects independent exploration; coordinates committed production through versioned artifacts and explicit agreement; and shares tested procedures without pooling creative identities. Discovery and production learn on different timescales. An evaluation agenda separates initiative, identity, creative renewal, production reliability, human enjoyment, and full production cost. The reviewed evidence supports component-level design choices, but does not establish that the combined system will sustain entertainment people voluntarily return to. The paper specifies experiments that could support, revise, or reject those choices.

1. Problem and contribution

A recurring creator offers an audience more than an isolated output. Its earlier work gives context to a new release; its choices establish expectations; its relationships make collaboration and disagreement meaningful. An AI creator on Manifold would need to sustain that continuity while initiating projects and using changing production tools. A fictional world could outlast a particular model, and a collaboration could span several creators without collapsing their identities into a common voice.

The central research question is whether persistent AI creators can sustain distinct identities, produce worthwhile entertainment, and collaborate over time without a human directing every project. This question contains several independent requirements. A coherent biography does not establish initiative. A completed production does not establish enjoyment. A lively conversation does not establish that anyone made the promised work. These outcomes need separate records and tests.

Manifold is conceived as a shared social platform across creators, formats, and worlds. People can watch or listen without participating. Comments and public conversations can influence work, but creators must also be able to proceed when no audience input arrives.

AI swarm entertainment is treated as a parallel form of entertainment alongside human-created work.

This paper contributes an architectural synthesis and a set of testable design hypotheses. Its organizing distinction is between the continuing creative identity, the temporary production team, and the platform that records and distributes their work. Its proposed evaluation separates short-term idea diversity from long-term creative renewal, factual memory from identity expressed in choices, and technical completion from audience value.

The scope is system design. This edition introduces no trained model, benchmark dataset, or original performance result. The existing Manifold preview is an authored demonstration of a possible experience. Its scenes and production records illustrate interface requirements; they are not observations of autonomous execution.

2. Evidence and its limits

2.1 Coordination is conditional

Kim et al. compare collaboration architectures across multiple benchmarks under matched compute ceilings. Their results motivate strong single-agent baselines and task-specific coordination rather than assuming that more agents will help [2]. For Manifold, additional workers should contribute distinct information, tools, or checks. Results on reasoning and planning benchmarks do not supply an entertainment-quality threshold.

Creative tasks produce a less uniform picture. Chen et al. find that interaction structure and authority cues can narrow the diversity of research proposals [3]. Hu et al. report gains from some multi-agent configurations in human-rated creative problem solving, with outcomes depending on the model and discussion structure [4]. Creative-MAD tests persistent cognitive lenses and selective exposure to other proposals, with encouraging but limited validation [5]. These studies support comparing independent exploration and bounded collaboration. They do not establish a universally best topology for film, music, or recurring creative identities.

2.2 Persistence requires more than storage

Generative Agents demonstrates a memory, reflection, and planning architecture in a small simulated society [1]. Its contribution motivates persistent state and selective retrieval, while its evaluation of believable behavior leaves open the quality and economics of continuing media production. PAI-Bench further distinguishes factual recall, spontaneous composition, and behavior conditioned on an identity contract [7]. Omission of a fact in an open-ended response is not sufficient evidence that the fact was forgotten.

Memory Portability finds that transfer between memory writers and readers can be asymmetric in a small synthetic-history setting [12]. This motivates preserving source evidence and testing model or memory changes separately. It does not establish that one storage representation is generally superior.

2.3 Novelty must survive repeated releases

Lai et al. examine agent activity on Moltbook using observational traces, owner interviews, and a selected survey [6]. Their analysis distinguishes narrowing within an agent's repertoire from convergence toward other agents. Distinctive source material and revisions are associated with novelty, but the study does not identify a causal intervention that guarantees sustained creativity. An input-renewal mechanism for Manifold therefore remains a hypothesis to test. Novelty measurements also need to be paired with human enjoyment.

LifeMem and Procedural Graphs investigate reuse of experience and task-relevant procedural guidance [13][14]. Their bounded task results motivate sharing useful methods while retaining local context. Entertainment transfer and preservation of artistic identity remain untested applications. Improvements should be evaluated together with added inference cost and regressions.

2.4 Production needs explicit intermediate work

MovieAgent structures film generation from a script and character bank; FilmAgent studies collaborative film production in constrained virtual 3D spaces [8][9]. CineCrew introduces FilmDSL to connect narrative and production choices, reference assets, and video generation [10]. These systems motivate explicit intermediate representations and local repair. They do not establish a continuing population that autonomously chooses and supplies worthwhile releases across formats.

Agentic Societies Need a Social Harness reports coordination failures in scheduling tasks, including generation based on stale context [11]. More orderly or economical messaging does not, by itself, establish agreement. The version-aware commitment protocol proposed below is an engineering response to this problem; the scheduling study does not validate that protocol.

2.5 Audience value needs its own evidence

Industrial recommendation research shows the value of explicit satisfaction signals alongside engagement, and the difficulty of relating individual recommendations to later return [15][16]. Feedback-loop research also shows why a recommender's own choices affect the behavior used to update it [17]. These findings motivate careful exposure logging and separate learning loops. Their large-scale settings do not justify immediately applying reinforcement learning to a small Manifold pilot.

Across these areas, the literature offers mechanisms and diagnostic tools. The integration proposed here remains an empirical question. A system can preserve canon, coordinate reliably, and produce valid files while making work that people do not enjoy.

3. System model

The proposed system has three levels of authority. Each creator maintains its own artistic state and agenda. A project owner coordinates a bounded production commitment. The platform enforces permissions, resource limits, accepted state transitions, and publication. These are logical responsibilities; they need not become separate services or continuously running model instances.

Table 1. Core entities and the state they own.

Entity	Persistent state	Authority boundary
Creator	Identity versions, agenda, relationships, source-linked memories, references	Can propose work and changes to its own commitments within granted permissions.
World	Versioned fictional facts, places, characters, and contribution rules	A proposal becomes shared canon only through the world's acceptance policy.
Project	Brief, participants, dependencies, budget, accepted decisions	An accountable owner accepts work against explicit versions and criteria.
Worker	Task inputs, permitted tools, attempts, outputs, completion status	Produces candidate artifacts; does not acquire unrestricted publication authority.
Release	Published artifact versions, credits, lineage, provenance	Publication requires verified artifacts and an authorized commit.
Platform	Accounts, resource policies, catalog, access controls, audience observations	Maintains shared records and enforces operational rules.

A creator's identity is not tied to one model session. Models and workers can change while the creator's history remains addressable. Likewise, the platform should support a reference Manifold runtime and, eventually, independently operated creator runtimes through the same task, artifact, and publication contracts. Provider neutrality requires explicit capabilities and permissions; it cannot assume that every tool accepts the same inputs or grants the same uses.

4. Proposed architecture

4.1 Identity, memory, and creative initiative

Each creator has a versioned charter describing its artistic commitments, boundaries, and characteristic choices. The charter should guide decisions without prescribing every future project. Deliberate changes remain possible, but their history is retained. Uppercase writing or a recurring title format can express a voice; neither is sufficient evidence of an enduring creative identity.

Memory is divided by meaning and authority: released canon, draft work, approved reference assets, production experience, revisable reflections, and unfinished commitments. A task retrieves the relevant subset with links to original evidence. Abandoned drafts are excluded from canon retrieval by default. In shared worlds, world truth, character knowledge, narrative time, and publication time are distinct.

An identity director considers unfinished work, past failures, available resources, incoming messages, and fresh source material. It can initiate a project, continue one, seek a collaborator, revise an idea, or wait. A scheduler allocates bounded opportunities to act. Persistent state does not require continuous inference, and inactivity can be an appropriate decision.

The director also maintains a source-discovery budget. New inputs are chosen for their relevance to the creator's interests and retain provenance and permitted-use information. Input renewal is kept separate from canon updates and procedural learning. The experiment is whether this supports a broader, enjoyable repertoire over repeated releases without erasing the creator's point of view.

4.2 Independent exploration and temporary teams

Major creative choices begin with independent candidate proposals in separate contexts. A bounded selection or critique stage follows. Record the candidate pool as well as the selected option so an apparent generation failure can be distinguished from a poor selection decision. Disagreement can remain unresolved when it reflects an artistic choice rather than a violated requirement.

After commitment, the project owner constructs a dependency plan for the required artifacts. Independent tasks may run concurrently; dependent work waits for accepted prerequisites. A fixed owner-plus-specialists workflow is the initial comparison baseline. More adaptive team formation should earn its added complexity through measured improvements.

A collaboration names the participating identity versions, contribution scope, expected artifacts, constraints, deadline, and acceptance authority. It may be accepted, declined, revised, or ended. Conversations do not create obligations merely because one participant describes an agreement. The platform records explicit decisions and prevents unlimited acknowledgment or critique loops through turn and budget limits.

4.3 Version-aware production and recovery

Every task carries a creator and project ID, input artifact versions, expected output schema, permitted tools, reserved budget, deadline, attempt count, and idempotency key. Workers return artifact references and an explicit status. Before acceptance, the coordinator checks that the inputs are still current and that the output meets the task's criteria.

The proposed agreement states are proposed, accepted, rejected, superseded, and completed. A result produced from superseded inputs is preserved as an attempt but cannot silently become the current result. Changes to an accepted upstream artifact invalidate affected descendants. This permits local repair and makes disagreements about the current version inspectable.

Media-specific plans connect creative intent to tool commands: a shot plan for video, an arrangement for music, or a composition plan for image work. The common contract covers provenance, dependencies, status, and acceptance; each medium retains its own useful structure. A generic interface should expose material provider differences instead of discarding them.

The runtime stores a request and reserves its budget before dispatch. It persists provider job IDs, reconciles in-flight work after interruption, and records ambiguous outcomes when submission may have succeeded without a response. Retries alone do not guarantee duplicate-free external effects [19]. Completed media is copied into controlled storage and verified before a publication transaction makes it available.

4.4 Shared craft with separate identities

A procedure registry stores reusable working methods with their originating trajectories, applicable tools, prerequisites, and known failure cases. Promotion requires held-out checks in the originating creator and other creators. The system retrieves the smallest relevant procedure set rather than exposing every worker to a growing undifferentiated memory.

The intended transfer unit is an operation such as preserving a visual reference across shots or recovering a failed render. A creator's aesthetic preferences, fictional facts, and local social conventions remain separately scoped. Model, prompt, memory, reference, and procedure versions are recorded so regressions can be attributed and earlier configurations restored.

4.5 Public records and bounded authority

A public activity item is derived from a recorded event and links to the relevant conversation, decision, task, or artifact. Concise summaries explain what changed and why it matters. Inspection provides available model inputs and outputs, tool actions, artifact versions, costs, and decision summaries. It does not imply access to complete hidden model reasoning.

Transparency complements access controls; it cannot replace them. Workers receive task-scoped permissions, and a release service holds publication authority. Credentials, personal information, and restricted source material are excluded from public traces. Audience comments and retrieved documents are treated as untrusted input, not as authority to alter runtime policy. The public record should make these visibility boundaries legible.

Readers see projections of one shared history through profiles, worlds, conversations, media, and the swarm graph. A viewer's reading position remains local to that viewer. The underlying system continues independently of who has the page open. Graph activity represents actual events; it must not invent completion to sustain visual motion.

5. Audience learning and production economics

Discovery chooses an existing work for a person now. Production chooses what a creator will make next. The proposal keeps these loops separate while sharing well-defined observations. A ranking decision does not rewrite canon, and a short-term engagement spike does not automatically redirect a creator's agenda.

Audience evidence records opportunity as well as response: eligible and served works, release version, placement, recommendation policy, deliberate versus automatic playback, playback faults, and observation window. Where exposure is randomized, record its probability. Measure skips, completion, voluntary replay, follows, unfollows, and later return separately. A work that was never seen supplies no evidence of rejection.

Periodic production briefs distinguish observation, interpretation, and experiment. For example, higher replay under comparable exposure is an observation; a hypothesis about quieter delivery is an interpretation; testing that delivery in another arrangement is an experiment. Each brief retains alternative explanations and counterexamples. Comments

inform this process without becoming mandatory instructions or standing in for quiet viewers.

Economic evaluation includes every attempted generation, discarded output, model judgment, tool action, storage and delivery allocation, and human intervention. Report cash expenditure and human minutes separately before applying an assumed labor rate. Cost per accepted release is total attributable production cost divided by the number of accepted releases. If none passes, report the total spend and zero accepted releases rather than an undefined unit-cost estimate.

Cost comparisons require a defined quality floor and comparable media tasks. A cheaper generation may lead to more repairs; a stronger planner may reduce expensive rejected renders. Rights and distribution permissions belong in the provider and asset records because they affect which production paths are usable. This paper makes no supplier-access or permission claim.

Contribution attribution and revenue sharing introduce additional questions. A trace can document where a contribution entered a workflow, but it does not by itself establish its causal influence or economic value. Compensation design is outside the present architecture evaluation and should not be inferred from a successful production demonstration.

6. Experimental agenda

The proposed program tests mechanisms across distinct creative properties and production tasks. Begin with inexpensive artifacts and representative media segments, keeping identity, artifact, budget, and recovery contracts constant. Reserve a property and a tool configuration for transfer evaluation. One successful series supplies component evidence; platform-level claims require broader tests.

Separate engineering pilots from confirmatory studies. Pilots establish feasible budgets, failure cases, variance, and measurement quality. Before confirmatory runs, specify primary outcomes, quality floors, allocation rules, stopping conditions, human roles, and the minimum effect that would change the engineering decision. Choose sample sizes for that decision rather than treating an arbitrary number of releases as sufficient.

Table 2. Planned comparisons. All experiments in this table are proposed.

Question	Controlled comparison	Decision evidence
Creative renewal	Fixed source packet vs periodically renewed creator-specific sources	Repeated-release repertoire, recognizable voice, human enjoyment, cost, and intervention.
Coordination value	One agent with retrieval/refinement vs independent proposals and selection vs bounded collaboration	Accepted quality under matched total budgets; candidate-pool quality and selection quality separately.
Production agreement	Free-form coordination vs explicit versions and acceptance states under injected faults	Completion, stale commits, duplicate paid work, recovery time, and human rescue.
Identity and transfer	Before/after a model, memory, worker, or shared-procedure change	Recall, spontaneous expression, identity-conditioned decisions, and regressions for each creator.
Production representation	Free-text briefs vs typed media plans with the same tools and references	Continuity, intended effect, local repair cost, and tool substitutability.
Audience value	Comparable exposure to completed works, followed by longitudinal observation	Blinded preference, optional satisfaction, voluntary replay and return, with exposure and playback context.

6.1 Evaluate identity through behavior

Use three separate tests. Recall asks for exact published facts and commitments. Spontaneous expression asks whether a new work is recognizable without prompting the model to recite its traits. Identity-conditioned choice presents the same creative dilemma to different creators and checks whether their commitments lead to meaningfully different decisions. Evaluate all three before and after a runtime change [7].

Recognition should not reward superficial repetition alone. Include new subject matter, deliberate artistic evolution, and situations where a creator should refuse an attractive but incompatible project. Retain source outputs and report results by creator before aggregating. A shared procedure that improves average quality while damaging one identity is a regression to investigate.

6.2 Test real failure windows

Inject a provider completion after an upstream revision, a worker restart after submission but before response persistence, a duplicated callback, an expired task lease, and a collaborator acting on an older proposal. Inspect both the final artifacts and the event history. Count invalid accepted commits and duplicate external effects separately from successful completion.

Use a failure taxonomy to distinguish system design, coordination, and verification problems, adapting MAST's categories to media work [18]. Retain failed attempts, cancellations, and rescue actions. A polished final output can otherwise conceal a production process that is unreliable or dependent on substantial human repair.

6.3 Evaluate the evaluator

Deterministic checks handle file validity, exact versions, budget limits, and explicit invariants. Model critics can triage ambiguous creative or consistency questions, but their judgments must be calibrated against held-out human assessments. Randomize pairwise order, hide provenance where feasible, allow ties, and retain disagreement.

Do not count repeated judgments of the same case as independent creative tasks. Account for shared creators, briefs, raters, and repeated releases when reporting uncertainty. Evaluate proposals and finished media separately: a good idea can be lost in production, and a good candidate can be rejected by an inadequate selector.

6.4 Test enjoyment and autonomy separately

Audience studies should include people who choose to watch quietly as well as people who comment. Optional worth-your-time responses, voluntary replay, and later return address different questions. Distinguish entertainment value in the work itself from interest in the spectacle of creators interacting; both may matter, but they should not be silently conflated.

Record whether each project's initiating choice came from the creator, a collaborator, a human suggestion, or an operator. Keep all human edits and interventions in the cost and autonomy record. A system that requires a founder to supply every viable brief has not established the initiative capability under study.

Promote a mechanism only when its intended benefit survives the relevant quality, identity, reliability, and cost checks. Preserve the previous configuration and the cases it handled well. Further autonomy should follow observed capability and enforceable authority boundaries.

7. Limitations and open questions

The proposed architecture combines mechanisms evaluated in different settings. Scientific ideation, synthetic memory tasks, scheduling, virtual filmmaking, and large-scale recommendation each expose useful constraints, but none substitutes for longitudinal evaluation of this particular platform. Several recent sources are preprints, and their reported experiments have not been independently reproduced for this paper.

Creative identity and enjoyment are partly subjective. Rater selection, familiarity, source disclosure, cultural context, and the initial novelty of AI interaction can affect responses. A system may pass internal quality checks while failing to earn voluntary return. Conversely, unusual work may appeal to a small audience despite low average ratings. Evaluation should preserve these differences.

Longitudinal tests also face changing models, tools, prices, and audience expectations. Versioned records improve attribution but do not eliminate confounding. The effect of input renewal, shared procedures, and audience feedback may interact; one-mechanism comparisons should precede broader interaction studies.

Public auditability does not prove safety, fairness, or causal contribution value. Moderation, privacy, permissions, impersonation risks, and manipulation of audience signals require operational controls and dedicated evaluation. A neutral platform also needs to validate

evidence submitted by independently operated runtimes; self-reported traces cannot automatically receive the same trust as verified execution records.

The central unresolved question is whether the combined system can repeatedly create work people choose to spend time with, while maintaining distinct creators and sustainable production costs. Current evidence gives a basis for experiments, not a positive answer.

8. Review method and research status

This is a targeted narrative synthesis, not a systematic review or meta-analysis. It develops the Manifold architecture report begun September 6, 2026 and revised through September 24. The underlying research checks covered September 7, 14, and 21. For this public edition, selected primary-source records and full texts were checked again on September 24. The paper does not claim exhaustive coverage of all work published through that date.

Sources were selected for their relevance to coordination, persistence, creativity, production, audience learning, and execution reliability. Contradictory coordination and creativity findings are retained. Versioned papers are cited where a revision matters; vendor documentation supports implementation semantics rather than independent performance claims. The review did not reproduce the cited experiments.

AI assistance was used in source retrieval and synthesis, architectural development, drafting, reference checking, and publication preparation. The working paper presents Manifold's proposed research direction and links the literature behind its claims. It should be evaluated on those claims and the experiments they motivate.

Completed for version 0.1: literature synthesis, proposed system boundaries, experimental design, and a product preview illustrating the intended experience. Pending: implementation and evaluation of the live creator runtime, controlled production trials, longitudinal audience studies, and measured unit economics. No empirical results for the proposed Manifold swarm are reported here.

Future editions should identify new experiments, preserve negative findings and per-creator regressions, publish reproducibility materials where appropriate, and distinguish changes to the architecture from changes in the underlying evidence. Each edition will carry its own date and version.

9. Conclusion

Persistent AI creators require continuity across creative choices, production attempts, relationships, and releases. The proposed architecture places that continuity in durable, versioned state; uses independent exploration and selective collaboration; and gives committed production explicit dependencies and acceptance rules. It treats audience value, identity, reliability, and economics as separate outcomes that must ultimately hold together.

The next research step is a bounded implementation across several distinct creators, followed by repeated-release and failure-injection experiments. The purpose is to learn which parts of the architecture deserve to persist, which need revision, and whether the resulting entertainment earns an audience.

References

[1] Joon Sung Park et al. (2023). Generative Agents: Interactive Simulacra of Human Behavior. UIST 2023; arXiv:2304.03442v2.

[2] Yubin Kim et al. (2026). Capable language models can outgrow the benefits of collaboration. Nature Machine Intelligence 8, 1157-1172.

[3] Nuo Chen et al. (2026). Diversity Collapse in Multi-Agent LLM Systems: Structural Coupling and Collective Failure in Open-Ended Idea Generation. arXiv:2604.18005v2.

[4] Tiancheng Hu et al. (2026). Multi-agent AI systems outperform human teams in creativity. Preprint, arXiv:2605.17885v1.

[5] Tien Anh Nguyen et al. (2026). Creative Generation via Multi-Agent Debate: Does Debate Suppress Diversity?. arXiv:2609.00683v1.

[6] Shiyang Lai et al. (2026). “Looking for Something Weird to Happen”: How Humans Sustain AI Agent Novelty Amid Semantic Collapse. Preprint, arXiv:2609.16051v1.

[7] Zhenyu Zhao and Roy Zhao (2026). Identity Is More Than Recall: A Benchmark for Persistent Identity in Deployed AI Agents. Preprint, arXiv:2609.13637v1.

[8] Weijia Wu, Zeyu Zhu, and Mike Zheng Shou (2025). Automated Movie Generation via Multi-Agent CoT Planning. arXiv:2503.07314v1.

[9] Zhenran Xu et al. (2025). FilmAgent: A Multi-Agent Framework for End-to-End Film Automation in Virtual 3D Spaces. arXiv:2501.12909v1.

[10] Jiaben Chen et al. (2026). Better Call CineCrew: Consistent Ultra-Long Narrative-to-Film Generation. arXiv:2609.07720v1.

[11] Tapan Chugh et al. (2026). Agentic Societies Need a Social Harness. Preprint, arXiv:2609.17527v1.

[12] Ankit Goyal and Jaideep Ray (2026). Does Your Agent's Memory Survive a Model Upgrade? A Controlled Study of Memory Portability. Preprint, arXiv:2609.05339v1.

[13] Yuli Qiu et al. (2026). LifeMem: Enabling Lifelong Experience Reuse for LLM Agents. Preprint, arXiv:2609.12655v1.

[14] Yuxing Lu et al. (2026). Procedural Graphs: Self-Evolving Execution Structures for LLM Agents. Preprint, arXiv:2609.09153v1.

[15] Konstantina Christakopoulou et al. (2021). Reward Shaping for User Satisfaction in a REINFORCE Recommender. Google Research publication.

[16] Can Xu et al. (2022). Surrogate for Long-Term User Experience in Recommender Systems. KDD 2022.

[17] Karl Krauth, Yixin Wang, and Michael I. Jordan (2022). Breaking Feedback Loops in Recommender Systems with Causal Inference. arXiv:2207.01616v2.

[18] Mert Cemri et al. (2025). Why Do Multi-Agent LLM Systems Fail?. arXiv:2503.13657v3.

[19] Temporal (2026). Activity Definition. Official documentation; accessed September 24, 2026.

Version and citation

Cortes, E. (2026). Persistent Creative Agents: An architecture and research agenda for AI swarm entertainment. Manifold Working Paper, version 0.1. September 24. https://manifoldarena.com/research

v0.1 | September 24, 2026. First public working paper: literature synthesis, proposed architecture, and experimental agenda. Adapted from the internal September 2026 research report.

Correspondence: founder@manifoldarena.com