

PALIMPSESTUS RESEARCH AGENDA

Provenance-Faithful Autonomous Systems

Stephen Weber | Founder & Principal Researcher, 1791 Technologies LLC | Los Angeles, CA

Can an autonomous system change what it believes and does without losing the path that warrants the change?

Palimpsestus is an independent research program testing whether provenance, warrant, inherited state, uncertainty, and accountable transition can be architectural constraints on adaptive systems. Its central research bet is that **provenance-preserving control constraints may produce behavioral regularities that survive changes of implementation and substrate**. *That is a hypothesis, not an established result.*

FROZEN FOUNDATION

- **Open Spec v1.1.3 / C1** is the frozen reference for current experiments.
- **Specification -> conformance -> implementation** are kept separate so compliance and empirical utility are not conflated.
- **No-Ad-Hoc-Rescue**: failures cannot be repaired by silently changing the frozen architecture after seeing outcomes.

EVIDENCE SO FAR

- **ARC-AGI-2**: frozen conformant implementation passed **192/192 semantic obligations** and **160/160 derivation edges**. Confirmatory work supported predictive Fold geometry; controller endpoint efficacy was null; operational-horizon preservation remains unresolved.
- **ARC-AGI-3 controlled Crucible**: on preregistered official **Is20**, Palimpsestus repeated disproven actions **0/250** times vs. **35/250 (14%)** for a matched learner. **su15** showed no separation; neither architecture achieved level progress; goal-directed utility remains unresolved.
- **ARC-AGI-3 official evaluation**: a Kaggle submission completed with **publicScore = 0.08**. The sealed record has not yet established attribution of that score specifically to Palimpsestus agent execution, so no stronger hidden-task performance claim is currently warranted.

WHAT REMAINS UNRESOLVED

- **Substrate independence**: do independent implementations converge on a common kernel?
- **History sensitivity**: can identical present beliefs warrant different futures because their evidence histories differ?
- **Compression**: how much history can be discarded without discarding future warrant?
- **Goal-directed utility**: can epistemic gains coexist with effective action?
- **Comparative advantage**: what does Palimpsestus produce that Active Inference, NARS/AERA, RL, or provenance systems do not?
- **Domain recurrence**: are cross-domain similarities genuine invariants or post-hoc analogies?

NEXT DISCRIMINATING EXPERIMENTS

- **1. Extract and freeze the cross-substrate kernel**. Remove ARC-specific machinery from independently developed ARC-2 and ARC-3 realizations and identify the smallest common spec-derived implementation.
- **2. Compare architectures prospectively**. Palimpsestus vs. Active Inference vs. NARS/AERA vs. a matched conventional learner under identical observation, action, memory, and resource budgets. Use cases where the architectures should make different predictions before execution.
- **3. Test independent domains without changing the ruler**. Candidate Crucibles include autonomous synthetic worlds, aerospace autonomy, and provenance-intensive distributed systems. **The domain must consume Palimpsestus; Palimpsestus must not be rebuilt to fit the domain**. Failure is a valid result.

COLLABORATORS SOUGHT Formal methods & mathematical systems | autonomous agents / Active Inference / cognitive architectures | distributed systems & provenance | experimental methodology & independent evaluation | aerospace, robotics, simulation, games & persistent synthetic worlds

Belief in Palimpsestus is not required. Competent disagreement is particularly useful.

Research stance: What does the architecture predict? What survives substrate changes? What does it cost? What do competing approaches do better? Where does it break?

Open Spec: github.com/thepalimpsestus/palimpsestus-spec | **Conformance:** github.com/thepalimpsestus/palimpsestus-conformance