ANOINTTEL

MARK I

RESEARCH BRIEF / REVISED EDITION

Engineering retained adaptation.

Persistent state. Predictive control.
Evidence that can be inspected.

THE RESEARCH QUESTION

Can an artificial agent use experience to improve future decisions, retain that improvement across a restart, and make the causal role of its state measurable?

The original synthetic benchmark remains documented. This revision adds the latest Codex reference-intake audit, the complementary ten-operation proposal, and a strict boundary between mathematical consistency and tested cognition.

CURRENT EVIDENCE STATUS

Bounded runnable prototype + completed reference audit

Research direction: Xantana / taykOff™

---

01 / EVIDENCE STATUS

What actually changed.

Revision principle: keep each piece of evidence attached to the exact mechanism and date it supports.

21 OCT / IMPLEMENTED

MARK I: an executable synthetic hidden-rule benchmark with prediction, feedback, SQLite state, restart and held-out performance. [1]

POST-BENCHMARK / REPORTED

PSE-001 comparison concluded ITERATE. The available follow-up describes RLS/control conditions, but supplies no score table for independent review here. [4]

01 NOV / AUDITED

Reference Intake 001 preserved seven supplied images, verified provenance and Numogram arithmetic, and enumerated mapping ambiguity. No new cognition benchmark was run. [4,5]

01 NOV / UNTESTED

The state-dependent ten-operation router and outer architecture-transformation validator remain proposed contracts, not executable demonstrated improvements. [4,5]

FOUR SEPARATE PROOF OBLIGATIONS

ACQUIRE
Experience changes a record.

RESTORE
A fresh process reloads the record.

USE
The restored record changes an action.

BENEFIT
That causal influence improves outcomes.

A correct arithmetic audit is not a causal test. Passing software tests is not a performance result. A synthetic rule-family advantage is not evidence of consciousness or general intelligence.

02 / HISTORIC EXPERIMENT

Original MARK I benchmark.

21 October 2026 source-reported experiment; metrics are preserved, not rerun for this revision. [1]

IMPLEMENTED CONTROL LOOP

01 Controller
Commit a prediction and choose an action.

02 Actor
Write an actual binary output file.

03 Evaluator
Score output against hidden reference bytes.

04 Adapt
Revise compatible rule hypotheses.

05 Persist
Store updates, traces and state hashes in SQLite.

06 Restart
Reload state in a fresh worker process.

HELD-OUT CORRECTNESS



DESIGN + LIMITATIONS

Five hidden-rule worlds, 128 training and 32 held-out tasks per world. There were 24 training cues and eight strictly held-out cues; the report records five distinct worker process IDs per agent/world and 13 passing software checks. All agents saw the same public task sequences, but compute budgets and built-in rule assumptions were not matched.

93.1% is an in-family result (149/160), not a measurement of transfer across model backends. The hypothesis-class prior and weak retrieval comparator are important alternative explanations; remaining adaptive errors were attributed in the original report to its 10% exploration policy. [1]

03 / REFERENCE INTAKE

What Codex verified.

The later reference package is a conceptual and integrity addendum, not a new performance experiment. [4,5]

07	06	50	3,840
images archived	unique image hashes	prior files unchanged	compatible bijections

EVIDENCE INVENTORY

01 / NUMOGRAM	Ten zones, complementary pairs, currents and gate labels.
02-03 / RECURSION	Duplicated image: A[n+1] = V(A[n], T(A[n], A[n])).
04 / OPERATIONS	Five complementary named cognitive-operation pairs.
05 / ROUTING	Proposed state-dependent routing and transformation formula.
06 / STARGRAM	Distinct perspective diagram; meanings remain unspecified.
07 / PRISM	Visible-spectrum analogy; no computational dimensions inferred.

AUDIT VERDICT / MATHEMATICAL CONSISTENCY ONLY

Five zone pairs, five operation pairs and ten gate rows check out. A shift assignment is compatible but not unique. operation_algorithms_supplied=false; cognitive_architecture_executed=false; new_performance_evidence=false. [5]

04 / CANDIDATE SPECIFICATION

Ten named operations.

The supplied labels specify conceptual contrasts. Distinct executable algorithms have not been supplied. [4]

01 INCEPTION — **10 COMPLETION**

Initial intention ↔ achieved result

02 FORMULATION — **09 EMANATION**

Internal structure ↔ external expression

03 CONCEPTION — **08 COAGULATION**

Generated possibilities ↔ stabilized synthesis

04 CONSOLIDATION — **07 IMPLEMENTATION**

Conceptual coherence ↔ practical construction

05 DESTABILIZATION — **06 DISPENSATION**

Disrupted assumptions ↔ redistributed possibilities

TWO DIFFERENT INDEX SPACES

```
Numogram zones:  z ∈ {0…9},  complement = 9 − z
Operations:      k ∈ {1…10}, complement = 11 − k
```

A complement-preserving map must satisfy $\varphi(9-z)=11-\varphi(z)$. $\varphi(z)=z+1$ is one possible assignment, but it is not uniquely implied by the reference. Five pair permutations × 32 pair orientations = 3,840 valid bijections. The architecture needs a justified mapping, not a convenient default. [4,5]

05 / PROPOSED CONTROL SYSTEM

Route. Transform. Retain.

A design contract inferred from the supplied state-transition formula; not yet an implemented learning controller. [4]

SUPPLIED FORMULA / CONCEPTUAL

$$S[t+1] = F_{\varphi(R_N(z[t], S[t]))}(S[t], I[t])$$

OPERATIONAL COMPLETION / PROPOSED

1 ROUTE

$r[t] = R_{N,\theta}(z[t], S[t])$

Select an allowed destination under a specified state-dependent policy.

2 ASSIGN

$k[t] = \varphi(r[t])$

Convert the chosen zone to a documented operation.

3 TRANSFORM

$S[t+1] = F_{k,\theta}(S[t], I[t])$

Apply one concrete typed transformation, with measurable effect.

4 RETAIN

$z[t+1] = r[t]$

Save the new zone alongside the transformed state.

THE RECURRENCE TEST THAT MATTERS

Show that two prior states differing because of a transformation lead to different later operation selections. Merely logging a zone, passing state as an argument, or running ten fixed filters does not demonstrate state-dependent recurrence. The reference currently supplies no complete F_k algorithms. [4]

06 / META-ADAPTATION

Propose. Validate. Accept.

The outer architecture-update equation is a research proposal, not a verified self-improvement loop. [4]

SUPPLIED OUTER SCHEMA

$$A[n+1] = V(A[n], T(A[n], A[n]))$$

This does not by itself specify an actual transformation or a noncircular validation rule.

A FALSIFIABLE CANDIDATE CONTRACT

PROPOSE

$$B[n] = T(A[n], A[n]; \text{train})$$

VALIDATE

$$\text{accept}[n] = V(A[n], B[n]; \text{validation, frozen criteria, budget})$$

COMMIT

$$A[n+1] = B[n] \text{ if accepted; otherwise } A[n]$$

REQUIREMENTS BEFORE THE CLAIM CHANGES

Versioned revisions
Bounded, reversible changes to operation parameters or transformation grammar.

Separated evidence
Training, validation and untouched final-test data; guard against validator overfitting.

Matched comparisons
Ordinary recurrence, learned router, sequential/random and graph controls at equal budgets.

The previous RLS covariance adjustments did not implement this outer loop. A rule that only changes selection probabilities is not evidence that the cognitive operations themselves learned new algorithms. [4]

07 / INTERPRETATION

Claims, separated by proof.

The decision threshold is empirical advantage over controls, not the visual elegance of the proposed design.

Historical MARK I

SUPPORTED	Runnable synthetic benchmark; reported 149/160 held-out, restarts and 13 tests.
NOT SHOWN	No matched-resource general advantage, real-model transfer or subjective experience. [1]

PSE-001 / RLS

SUPPORTED	Completed prior comparison; follow-up records ITERATE verdict.
NOT SHOWN	Full PSE score tables and independent rerun are not included in available materials. [4]

Reference audit 001

SUPPORTED	File integrity, pairs, gate arithmetic and mapping enumeration executed.
NOT SHOWN	No ten-operation transformation, router, recursive improvement or task-performance validation. [4,5]

Future architecture

SUPPORTED	A formally expressible zone router and candidate acceptance contract.
NOT SHOWN	Needs typed algorithms, causal testing, preregistration and independent replication. [4]

CURRENT OVERALL RESEARCH VERDICT

ITERATE. The symbolic specification is more precise, but useful cognitive function, autonomous architecture improvement, and consciousness remain unestablished.

08 / NEXT PROOF + SOURCES

Earn the next claim.

First complete one reproducible loop; then measure a causal advantage under frozen matched conditions.

01 / IMPLEMENT

Define typed state, input, output, graph edges, φ mapping and ten F_k algorithms (or disclose duplicates). Run actor → prediction → action → graded outcome → update → restart.

02 / COMPARE

Preregister budget-matched recency / crossed-shift study and compare fixed Direct/Trace, conventional memory, ordinary recurrence and learned-router controls. Freeze scoring, sample size, budgets and stopping rules.

03 / ABLATE + TRANSFER

Remove/swap/shuffle retained state; assess prediction calibration, generalization to new families/backends, tokens, costs and latency. Preserve failures and seek independent replication.

PROVENANCE / WHAT WAS REVIEWED

[1] MARK_I_VERIFIED_REPORT.md (21 Oct 2026), as documented in the uploaded six-page brief; underlying report not separately inspected.

[2] MARKI_live_go_no_go_rule_DRAFT.md (21 Oct 2026), referenced by original brief; not a registered protocol.

[3] Proto_Subjectivity_Expedition_Checkpoint.md (07 Oct 2026), referenced by original brief; research aims only.

[4] README.md, "Reference intake 001: Numogram, paired operations and recursive transformation" (later Codex addendum).

[5] audit.json, reference-audit-v1 (7 images, 6 hashes, 50 files unchanged, 3,840 compatible mappings).

ORIGINAL REPRODUCTION ENTRY

```
python -m unittest discover -s tests -v
python -m anoint_mark_i benchmark --out results/independent \
  --seeds 5 --train 128 --test 32 --restart-every 32
```

Commands require the original source package; they were not executed for this revision. Reference-audit reproduction is documented separately in [4]. Related-work context retained from the original brief: Reflexion (Shinn et al.), MemGPT (Packer et al.), LongMemEval (Wu et al.).