Bounded Neuroplasticity: Self-Calibrating Organizational Memory Without Self-Authorizing Knowledge

Reflexive Admission, Governed Plasticity, Evidence-Root Independence, and Reversible Memory Adaptation

Matthew Blizzard

Armored Helix Systems LLC

VERSION
1.0

DATE
September 2026

TYPE
Technical White Paper / Reference Architecture

STATUS
Prototype-informed architecture; enterprise-scale and longitudinal claims require empirical validation

This paper presents a proposed reference architecture for memory systems that adapt retrieval, language, compaction, and graph structure while preserving non-self-modifiable boundaries around authority, provenance, permissions, lifecycle, and action. Architectural and economic propositions are identified as hypotheses unless supported by cited prior work or empirical evidence.

SECTION 0

Abstract

Persistent organizational memory should not remain structurally frozen. As an enterprise accumulates AI-assisted work, a memory system should learn its acronyms, project aliases, domain-specific meanings, retrieval patterns, recurring workstream boundaries, compaction failures, and useful relationships. It should be able to adjust bounded ranking parameters, propose Neuron merges or splits, strengthen or weaken Synapses, and improve how semantic inference is allocated. A static memory system leaves substantial quality and economic gains unrealized.

Reflexivity also creates a distinct risk. A system that consumes its own analyses can mistake generated conclusions for new observations, count multiple descendants of one source as independent corroboration, amplify confidence through repetition, rewrite graph structure without preserving history, or optimize a convenient metric at the expense of fidelity and access control. The resulting failure is not merely hallucination. It is *recursive epistemic inflation*: internally generated state can acquire influence because the system repeatedly encounters its own prior output.

This paper proposes **bounded neuroplasticity**: the governed capacity of an organizational-memory system to recalibrate retrieval, language, compaction, routing, and graph structure while preserving non-self-modifiable constitutional boundaries around identity, permissions, provenance, authority, lifecycle control, derivation depth, and action. High-order SynapSys reasoning returns through a **Reflexive Admission Gate**; internally generated objects are classified as `SYSTEM_DERIVED`, retain their external evidence roots, remain ineligible for self-corroboration, and are subject to materiality, producer-exclusion, derivation-depth, and reflection-budget controls.

A **Plasticity Controller** converts memory telemetry into versioned proposals rather than direct mutations. Candidate changes are evaluated through historical replay, held-out cases, shadow execution, and bounded canary rollout before Telomere and authorized governance permit reversible application. Plasticity may tune organization-specific lexicons, candidate ranking, thresholds, batching, freshness policies, Neuron boundaries, and Synapse topology. It may not increase source authority, broaden permissions, waive provenance, suppress revocation, bypass the Telomere Integrity Layer, authorize Axons, or promote its own outputs into trusted state.

The paper also defines a multi-dimensional **Memory Health** framework covering fidelity, coherence, retrieval quality, freshness, economic efficiency, integrity, reflexivity, language resolution, and structural stability. Metrics disclose sample size, measurement method, and confidence; insufficient evidence remains `INSUFFICIENT_DATA`, not a fabricated health score.

Keywords: organizational memory; reflexive AI; self-calibrating systems; persistent context; governed plasticity; semantic compaction; knowledge graphs; memory poisoning; provenance; evidence independence; SynapSys; Grey Matter; Telomere.

> **Core architectural claim.** An organizational-memory system may learn how to interpret, retrieve, compact, relate, and review knowledge without acquiring the authority to decide that its own outputs are true. Plasticity may improve methods and structure; authority must remain externally grounded, policy-bounded, independently governable, and reversible.

Publication Note, Implementation Status, and Suggested Citation

This is an independent technical white paper authored by Matthew Blizzard and published by Armored Helix Systems LLC. It is not presented as peer-reviewed academic research. The architecture is informed by an implemented prototype of a Memory Control Plane, reflexive-admission controls, versioned memory policies, organization-language mappings, event-sourced Synapse changes, and operator-facing Memory Health views. Prototype completion and software tests are evidence of implementability, not evidence of enterprise-scale semantic quality, safety, or economic benefit. Those claims require target-workload and longitudinal validation.

Suggested citation: Blizzard, M. (2026). *Bounded Neuroplasticity: Self-Calibrating Organizational Memory Without Self-Authorizing Knowledge*. Technical White Paper, Version 1.0. Armored Helix Systems LLC.

> **IP publication note.** Public disclosure can affect patent rights and filing strategy, particularly outside the United States. This paper intentionally describes architectural principles while omitting implementation-specific schemas, scoring weights, thresholds, customer configurations, deployment recipes, resource identifiers, and proprietary orchestration details. If patent protection is contemplated, publication should be reviewed with qualified intellectual-property counsel before public release.

Canonical system boundary. SynapSys owns evidence, organizational memory, ontology, Attention, and Axons. Grey Matter

supplies bounded adaptive semantic execution. Telomere governs whether candidate state may acquire or retain influence. The Plasticity Controller proposes bounded changes to memory behavior; it does not supersede Telomere, permissions, or Axon policy.

What Version 1.0 Establishes

Area	Architectural addition
Reflexive memory	High-order findings return through a dedicated gate as duplicates, system-derived hypotheses, or control-plane proposals rather than ordinary external observations.
Evidence independence	Generated descendants collapse to external evidence roots and cannot manufacture corroboration through repetition.
Memory Control Plane	Retrieval outcomes, corrections, metrics, lexicon state, and tuning history remain distinct from organizational claims.
Governed plasticity	A proposal-only controller uses replay, holdout, shadow, canary, Telomere review, and rollback before a change can influence memory.
Structural rewiring	Neuron and Synapse changes are versioned, event-sourced, temporal, and reversible.
Memory Health	Fidelity, coherence, retrieval, freshness, economics, integrity, language, and reflexivity are measured separately with explicit confidence and insufficient-data states.

SECTION 0
Contents

SECTION 1
Problem Statement

AI-assisted organizations increasingly create useful context across chats, coding agents, workflow systems, tickets, repositories, documents, analytics, and domain-specific tools. A persistent-memory architecture can consolidate that activity into evidence-linked Dendrites, Neurons, Synapses, Region Capsules, and downstream Axons. Once such a system exists, a new question appears: should the memory model remain fixed while the organization itself changes?

A permanently static memory policy is undesirable. Internal language evolves. Project aliases emerge. Different Brain Regions exhibit different identity features. A repository identifier may strongly imply workstream continuity in engineering, while a protocol identifier may carry more weight in clinical operations. Some Synapses decay quickly; others represent durable dependency. A merge threshold that works for one Region can fragment another. A compaction policy that is efficient for routine support work may erase qualifiers in research or security investigations.

Yet unconstrained self-modification is worse. If the memory system treats its own findings as independent evidence, each summary, synthesis, or restatement can become another apparent confirmation of the original claim. If a model can directly rewrite retrieval weights, graph edges, authority, or access rules, it can optimize toward its own prior outputs. A corrupted memory can therefore become easier to retrieve, more interconnected, more confidently summarized, and harder to dislodge.

The design challenge is not whether a memory system should learn. It is *what it may learn, from which evidence, through which pathway, under which invariants, and with what rollback semantics*.

This paper addresses four questions:

1. How can high-order organizational inference feed useful conclusions back into persistent memory without becoming self-validating truth?
2. How can the memory system measure its own quality and tune retrieval, compaction, language resolution, and graph structure without modifying its constitutional safety boundaries?
3. How can proposed Neuron and Synapse rewiring remain auditable, temporal, reversible, and permission-safe?
4. Which Memory Health metrics can support self-calibration without collapsing multidimensional quality into a misleading single score?

SECTION 2
Core Thesis and Contribution

The architecture is organized around ten principles.

1. **Self-observation without self-authorization.** SynapSys may learn from its own operational outcomes and high-order reasoning, but internally generated conclusions cannot make themselves authoritative.
2. **External-root grounding.** Every derived claim retains lineage to external evidence roots. Repeated descendants of the same root do not count as independent corroboration.
3. **Finite reflexivity.** Derived-origin classification, material hashes, producer exclusion, derivation-depth limits, and reflection budgets prevent uncontrolled reasoning loops.
4. **Two-plane separation.** Organizational claims live in the Organizational Memory Plane. Retrieval outcomes, corrections, quality telemetry, and tuning history live in the Memory Control Plane.
5. **Proposal-only plasticity.** The Plasticity Controller proposes changes. It does not directly mutate trusted state or constitutional policy.
6. **Constitutional immutability.** Tenant isolation, permission lineage, source-authority ceilings, provenance, no-self-corroboration, revocation, context ceilings, graph-depth bounds, Telomere authority, and required human approval are not plastic parameters.
7. **Reversible structure.** Neuron merge/split and Synapse rewiring are event-sourced, versioned, and rollback-capable rather than silently overwritten.
8. **Organization-specific language.** Learned aliases and domain meanings can become deterministic routing aids only after governed evidence or review.

9. **Multi-objective optimization.** Memory should improve fidelity, retrieval, freshness, useful findings, cost, latency, and structural stability jointly rather than maximizing one proxy.
10. **Observable uncertainty.** Metrics carry method, sample size, confidence, and status. Missing evidence is reported as insufficient data rather than converted into pseudo-precision.

> **Central thesis.** Persistent organizational memory should be able to reorganize itself, but every act of reorganization must remain externally grounded, constitutionally bounded, independently governed, and reversible.

The contribution claimed here is not generic self-reflection. Prior work has shown that agents can improve task performance through verbal feedback, iterative self-refinement, and evolving memory structures [3–6]. The proposed distinction is the governance boundary around *persistent organizational adaptation*: generated descendants collapse to their external evidence roots; plasticity is expressed as typed proposals; structural changes preserve historical state; and adaptive components are unable to elevate authority, broaden permissions, or authorize action.

SECTION 3
Benefits in Plain Language

3.1 The memory can learn the organization's language

A new project nickname, internal acronym, repository alias, or historical program name should not require a permanent model call. Once a mapping is sufficiently evidenced and governed, SynapSys can use it as part of deterministic candidate routing.

3.2 Memory quality can improve from corrections

Every accepted or rejected Neuron assignment, Synapse proposal, merge, split, compaction audit, and query correction creates feedback about how the memory model is performing. That feedback belongs in the Memory Control Plane and can support bounded calibration.

3.3 High-order findings do not have to disappear

A cross-region analysis may discover a real dependency or contradiction that no single source stated explicitly. Bounded neuroplasticity allows that finding to return as a system-derived hypothesis without pretending it is a new external observation.

3.4 Graph structure can evolve without rewriting history

A relationship that was once useful may become weak, dormant, superseded, redirected, or revoked. Event-sourced rewiring allows the current graph to change while preserving temporal and audit history.

3.5 The system can spend inference more intelligently

As the lexicon and deterministic routing improve, repeated semantic resolution can shift toward lower-cost paths. Memory Health can identify where strong models are useful, where they are wasteful, and where false merges or missed relationships justify additional reasoning.

3.6 Operators can see what the system is changing

Plasticity proposals, system-derived hypotheses, learned aliases, policy versions, and rewiring events are visible objects with evidence, expected benefit, blast radius, and rollback plans rather than hidden side effects of model behavior.

SECTION 4
Reference Architecture

The architecture extends persistent organizational context with a Reflexive Admission Gate and a Memory Control Plane.

SynapSys owns the durable organizational-memory model. **Grey Matter** supplies bounded semantic execution where

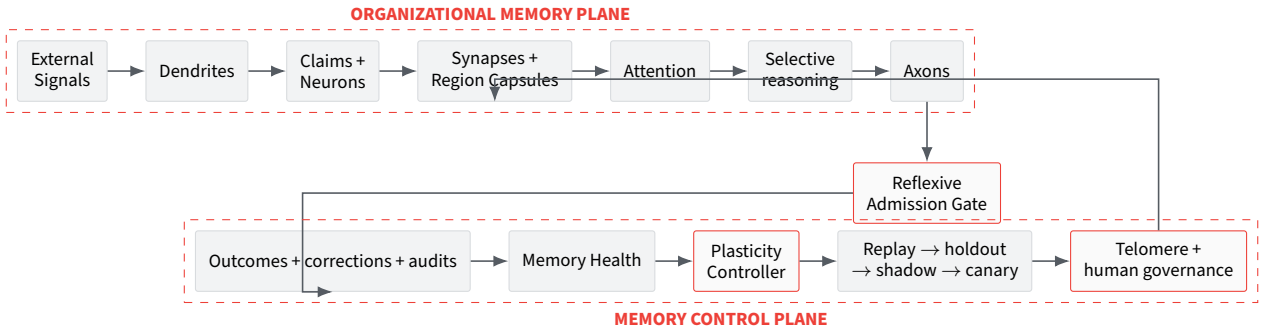

Figure 1: Two-plane reference architecture. Reflexive findings and memory telemetry can inform future structure, but promotion remains governed and reversible.

deterministic rules cannot enumerate meaning. **Telomere** governs whether candidate state may acquire or retain influence. **Attention** determines whether accepted state transitions justify deeper reasoning. **Axon policy** governs recommendations, artifacts, tool calls, and actions leaving persistent state. The **Plasticity Controller** operates in the Memory Control Plane and is intentionally subordinate to those boundaries.

The end-to-end adaptation loop can be summarized as:

```
OBSERVE -> INTERPRET -> COMPACT -> PROTECT -> PERSIST -> ATTEND -> REASON -> REFLECT -> MEASURE -> PROPOSE ->
REPLAY -> SHADOW -> CANARY -> GOVERN -> APPLY OR ROLLBACK
```

SECTION 5
Bounded Neuroplasticity Model

Bounded neuroplasticity distinguishes three categories of system behavior.

Category	Examples	Authority model
Plastic method	Retrieval weights, candidate top-K, region-specific thresholds, batching windows, freshness policy, lexicon mappings	May adapt inside hard bounds after evaluation and governance
Plastic structure	Neuron merge/split proposals, Synapse strength/type/dormancy/redirection	Event-sourced, reversible, evidence-linked, Telomere-governed
Constitutional boundary	Tenant isolation, permissions, provenance, authority ceilings, no self-corroboration, revocation, human approval, action authorization	Non-plastic; cannot be weakened by model output or optimizer results

Table 1: The architecture adapts methods and structure, not constitutional authority.

A useful formal constraint is:

$$\mathrm{Authority}_c(y) \leq \max_{r \in \mathcal{R}(y)} \mathrm{Authority}_c(r),$$

where y is a derived object, c is the claim domain, and $\mathcal{R}(y)$ is the set of external evidence roots supporting y. A model transformation cannot create a new evidence root or raise the authority of its ancestors. Authority may increase only through an explicitly allowed external mechanism, such as authenticated system-of-record attestation, independent corroboration under policy, or authorized human validation.

The term *neuroplasticity* is used as an architectural metaphor rather than a biological claim. Neurons and Synapses are persistent workstream and relationship objects; plasticity describes controlled change to their organization and the policies that resolve them.

SECTION 6

Reflexive Admission: Letting Reasoning Feed Memory Safely

High-order reasoning can discover useful organizational facts or hypotheses that are not explicit in any one source. Consider three independent workstreams whose state collectively reveals a shared vendor dependency. Discarding that finding wastes expensive inference; treating it as an ordinary external observation creates a recursive authority problem.

The Reflexive Admission Gate separates those cases.

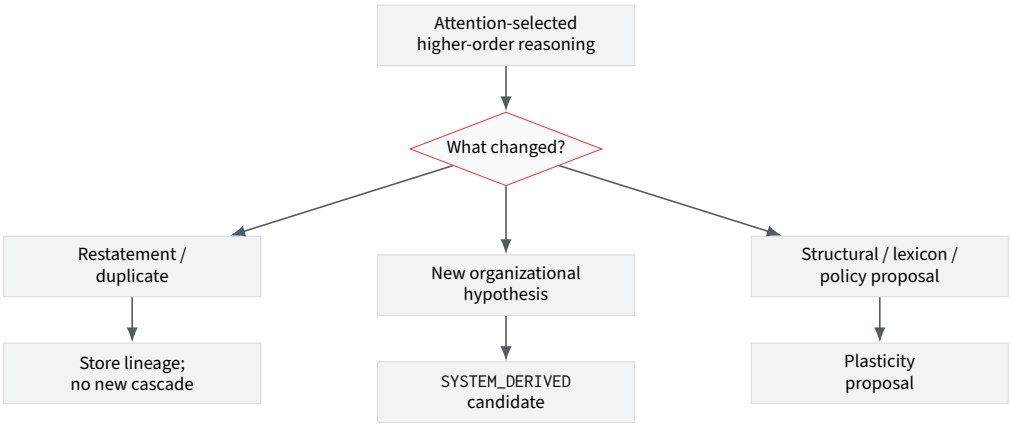

Figure 2: Reflexive Admission Gate. Reasoning output is classified before it can influence persistent memory or the Memory Control Plane.

Every internally generated candidate should retain, at minimum:

```
originClass: SYSTEM_DERIVED producerTaskType producerModelId parentAxonId supportingStateHashes
externalEvidenceRoots derivationDepth generatedAt selfCorroborationEligible: false
```

The complete reasoning artifact can remain in evidence storage, while only a compact structured delta re-enters active state. This follows the same progressive-compaction principle used for external Signals: persistent memory should preserve the material change, not repeatedly re-ingest the entire generated transcript.

SECTION 7
Evidence Independence and No Self-Corroboration

The most important reflexive invariant is that generated descendants do not become independent evidence merely because they are separate objects.

For a derived object y with parent set $P(y)$, define its external evidence roots recursively:

$$\mathcal{R}(y) = \bigcup_{p \in P(y)} \mathcal{R}(p).$$

For an authenticated external observation e, $\mathcal{R}(e) = \{e\}$. A generated summary, Dendrite, Neuron update, Region synthesis, or downstream hypothesis derived entirely from one external observation therefore still has one root.

Unique root identifiers are necessary but not always sufficient. Two documents may be copies of the same upstream source. Two dashboards may derive from one database. Two agents may summarize the same ticket. The architecture should therefore support **independence classes** based on upstream object, controlling principal, source system, collection mechanism, and known derivation links.

A claim may be corroborated only when policy determines that the evidence is both sufficiently independent and sufficiently authoritative for the claim domain. Two low-authority sources need not establish a high-authority claim, while one authenticated system of record may be sufficient for a narrow deterministic field.

> **No-self-corroboration invariant.** Increasing the number of internally generated descendants must never increase the independent external-support count of a claim.

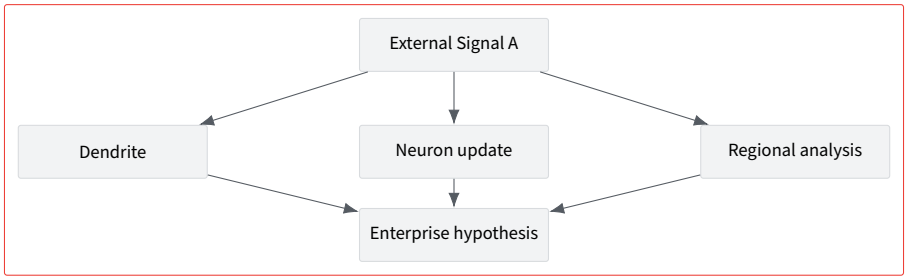

Figure 3: Evidence-root collapse. Internal transformations may create new hypotheses, but not new external corroboration.

SECTION 8

Finite Reflexivity: Depth, Materiality, Producer Exclusion, and Budgets

Self-reflection must have a stopping condition. Otherwise, a useful feedback mechanism becomes an inference cascade.

8.1 Derivation-depth limits

A conservative initial policy is:

Depth	Meaning	Automatic feedback
0	External observation	Allowed through normal admission
1	First-order derived hypothesis	Allowed as governed candidate
2	Approved cross-domain synthesis	Allowed only under stronger review
> 2	Recursive derived chain	No automatic re-entry; explicit investigation required

The exact ceiling is a policy choice to validate. The invariant is that automatic recursive derivation remains bounded.

8.2 Material-state hashes

A generated result that normalizes to the same material state should update lineage or cache metadata rather than create a new Dendrite, trigger a new Attention event, and reinvoke specialist reasoning.

8.3 Producer exclusion

A finding should not immediately re-invoke the same specialist that produced it. Progress should require a distinct verifier, new external evidence, human review, or a different bounded operation. This reduces same-model echo loops and correlated self-confirmation.

8.4 Reflection budgets

Each Attention event should carry a finite allowance for derived candidates, structural proposals, feedback inference calls, and maximum derivation depth. Budget exhaustion preserves the originating evidence and parks or defers follow-up rather than retrying indefinitely.

A simple reflexive load bound is:

$$R_{feedback} \leq \sum_{a \in A} b(a),$$

where A is the set of Attention events and $b(a)$ is a policy-bounded number of reflexive operations per event. This does not guarantee low cost, but it prevents raw recursive amplification from being unbounded by design.

SECTION 9

The Organizational Memory Plane and the Memory Control Plane

The two-plane distinction prevents implementation telemetry from becoming organizational truth.

9.1 Organizational Memory Plane

This plane contains what the system knows, believes, or hypothesizes about the organization: Signals, Dendrites, claims, Neurons, Synapses, state transitions, Region Capsules, derived hypotheses, and Axons. Objects retain source, time, permission, authority, evidence, producer, and policy lineage.

9.2 Memory Control Plane

This plane contains information about how well the memory system itself is functioning: retrieval candidates accepted or rejected, query corrections, merge and split outcomes, Synapse utility, compaction audits, model-call necessity, route regret, unresolved terminology, operator reversals, policy versions, and Plasticity history.

Operational traces should normally enter the Memory Control Plane rather than the Organizational Memory Plane. A user's rejection of a merge may be evidence that the resolver is miscalibrated; it is not automatically a new organizational claim about the underlying workstream.

Observation	Memory-plane interpretation	Control-plane interpretation
User rejects a Neuron merge	Workstreams remain distinct	Negative structural pair; evaluate threshold/ranking
Repeated alias resolves correctly	Canonical entity remains unchanged	Candidate for governed lexicon promotion
Synapse repeatedly helps successful retrieval	Relationship evidence may remain unchanged	Utility signal for ranking/strength calibration
Compaction loses a qualifier	State may require repair	Fidelity failure tied to policy/model version
Model call adds no material state	No new memory object	Economic inefficiency / cascade suppression signal

Table 2: The same event can have different meanings in the memory and control planes.

SECTION 10
Plasticity Controller

The **Plasticity Controller** evaluates Memory Health and proposes bounded, reversible improvements to identity resolution, compaction, retrieval, vocabulary, Neuron structure, and Synapse topology. It is a control-plane optimizer, not an autonomous authority.

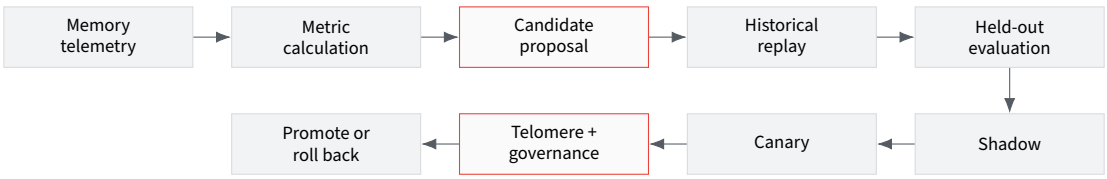

Figure 4: Plasticity lifecycle. A metric change is a reason to evaluate a proposal, not permission to mutate production state.

Proposal types may include:

- `MERGE_NEURONS`
- `SPLIT_NEURON`
- `ADD_SYNAPSE`
- `WEAKEN_SYNAPSE`
- `MARK_SYNAPSE_DORMANT`
- `CHANGE_SYNAPSE_TYPE`
- `REDIRECT_SYNAPSE`
- `ADD_WORKSTREAM_ALIAS`
- `ADD_ORGANIZATION_TERM`
- `ADJUST_REGION_ROUTING`
- `ADJUST_RETRIEVAL_WEIGHT`
- `ADJUST_MERGE_THRESHOLD`
- `ADJUST_BATCH_WINDOW`
- `ADJUST_RECENCY_POLICY`
- `ADJUST_ATTENTION_THRESHOLD`

Every proposal should identify the proposed change, reason, supporting metrics, affected objects, expected quality gain, expected cost impact, permission impact, blast radius, confidence, predecessor policy version, and rollback plan. The default operating mode should be **shadow**: measure the counterfactual before granting the change production influence.

SECTION 11
Learning the Organization's Language

Organization-specific language is one of the safest and most useful targets for plasticity because it can reduce repeated semantic ambiguity without changing authority.

A governed **Organization Lexicon Registry** can learn acronyms, project nicknames, product aliases, repositories, internal team terminology, vendor abbreviations, domain-specific meanings, historical names, and redirects. Mappings may be scoped by tenant, Brain Region, time period, source system, permission envelope, associated entities, and workstream context.

A term can have multiple valid meanings. For example, an acronym in a clinical Region may refer to a regulatory concept while the same token in engineering telemetry refers to an index or indicator. Global canonicalization without scope can therefore increase, rather than reduce, ambiguity.

A model may propose an alias, but deterministic routing should use it only after governed evidence such as repeated co-occurrence with an exact object identifier, repeated accepted workstream assignments, authenticated source metadata, human confirmation, or high-confidence cross-source agreement.

Once approved, the resolver can prefer:

1. exact Neuron ID;
2. exact workstream key;
3. approved Region-specific lexicon mapping;
4. approved global lexicon mapping;
5. bounded vector candidates; and
6. a resolver model only when bounded candidates remain ambiguous.

This progression converts repeated organization-specific ambiguity into deterministic routing whenever evidence supports doing so.

SECTION 12
Deterministic Calibration of Probabilistic Memory Resolution

Accepted and rejected decisions create labeled examples for bounded tuning:

```
Dendrite A -> Neuron X positive pair Dendrite A != Neuron Y negative pair Neuron X merged with Z positive
structural pair Neuron X split from Z negative structural pair
```

The initial target should be a versioned deterministic scorer rather than uncontrolled fine-tuning of the primary reasoner. A conceptual score is:

$$M = w_i I + w_t T + w_e E + w_a A + w_c C + w_g G + w_\tau \tau,$$

where I captures exact identifiers, T task-vector similarity, E entity overlap, A artifact overlap, C context similarity, G graph proximity, and τ temporal continuity. Brain Regions may use different bounded weight profiles because the evidence that indicates workstream identity differs by domain.

Only ranking and threshold parameters are plastic. Permission compatibility, tenant isolation, source-authority ceilings, provenance, Telomere lifecycle, exact negative identity constraints, and candidate-count ceilings remain deterministic filters.

Later, positive and negative pairs may support a small customer-specific reranker or embedding adapter, but the same governance principles apply: replay, holdout, shadow, canary, versioning, and rollback precede production influence.

SECTION 13

Neuron and Synapse Structural Plasticity

A persistent graph must evolve, but graph change should be treated as a state transition rather than a hidden overwrite.

13.1 Neuron merge and split proposals

Plasticity can identify candidate mega-Neurons through rising objective entropy, unrelated entity spread, contradiction density, update contention, and repeated operator corrections. It can identify fragmentation through repeated co-retrieval, shared exact objects, duplicate outputs, and later accepted merges. The controller proposes a merge or split; it does not execute one because a metric crossed a threshold.

13.2 Event-sourced Synapse rewiring

A Synapse lifecycle can include:

```
PROPOSED -> ACTIVE -> WEAKENED -> DORMANT -> SUPERSEDED -> REVOKED
```

with new evidence able to create a new active version or roll back a prior change.

Immutable events may include `SYNAPSE_CREATED`, `SYNAPSE_STRENGTH_CHANGED`, `SYNAPSE_TYPE_CHANGED`, `SYNAPSE_REDIRECTED`, `SYNAPSE_MARKED_DORMANT`, `SYNAPSE_REVOKED`, and `PLASTICITY_ROLLBACK`.

Signals that can strengthen a Synapse include repeated exact work-object overlap, explicit dependency identifiers, human confirmation, independent-source support, repeated co-change, successful co-retrieval, and one Neuron's outputs repeatedly becoming another's inputs. Weakening signals include false retrievals, human rejection, independent evolution, supersession, inactivity for time-sensitive relationships, permission conflict, repeated split decisions, or a stronger alternative relation.

Different relationship types should decay differently. `SAME_WORKSTREAM` may remain until split or superseded; `REQUIRES` remains while a dependency is active; `SUPPORTS` can weaken as claims expire; temporal relationships can decay quickly; and `CONTRADICTS` should persist until the contradiction is explicitly resolved.

A conceptual rewiring objective is:

$$G_{rewire} = Q_{after} - Q_{before} - \lambda_c C_{extra} - \lambda_s I_{instability},$$

where Q is retrieval or analysis quality, C_{extra} is incremental cost, and $I_{instability}$ penalizes structural churn.

SECTION 14

Memory Health: Measuring Whether Memory Is Improving

Memory Health should remain multidimensional. A single aggregate score invites Goodhart-style optimization: compression can be increased by deleting useful nuance; recall can be increased by retrieving everything; cost can be reduced by skipping necessary review. Operators need separate dimensions and their tradeoffs.

14.1 Fidelity and compaction

Metric	Meaning
Signal-to-Dendrite compression	Raw Signals per admitted Dendrite
Dendrite-to-transition yield	Dendrites producing a material Neuron change
Hot-memory compression ratio	Recoverable evidence size divided by active-state size
Compaction fidelity	Relevant evidence and qualifiers surviving compaction
Claim evidence coverage	Active claims with valid supporting evidence
Contradiction preservation rate	Material contradictions retained rather than summarized away
No-resurrection violations	Revoked or superseded claims incorrectly returning to active state; target zero

14.2 Coherence and graph quality

14.3 Retrieval, freshness, and economics

The primary economic KPI should be **cost per material, Telomere-admitted organizational state change**. It aligns spending with the architecture's intended scaling unit rather than raw prompt volume.

Metric	Meaning
Neuron merge precision	Proposed merges later confirmed as correct
Neuron fragmentation rate	Workstreams incorrectly spread across multiple Neurons
False-merge rate	Independent workstreams incorrectly combined
Split/merge reversal rate	Structural decisions reversed shortly afterward
Objective entropy	Unrelated objectives accumulating inside one Neuron
Synapse proposal precision	Proposed edges later supported or accepted
Synapse utility	Edges contributing to successful retrieval or analysis
Dormant-edge ratio	Historical relationships no longer useful in active retrieval
Plasticity reversal rate	Promoted changes subsequently rolled back

Metric	Meaning
Precision@K / Recall@K	Relevant memory among retrieved candidates / relevant memory successfully found
Context utilization	Selected context actually used or cited downstream
Evidence-to-context ratio	Useful evidence tokens divided by total model-input tokens
Signal-to-state freshness lag	Observation-to-admitted-state latency
Supersession latency	Time to deactivate superseded claims
Stale-claim ratio	Active claims past revalidation policy
Zero-model admission rate	State transitions handled without generative inference
Model calls per material transition	Semantic-call count normalized to real state change
Cost per validated claim	Semantic execution cost per admitted structured claim
Cost per Attention finding	Total processing cost per useful higher-order finding

14.4 Organization-language and reflexive safety

Metric	Meaning
Unresolved-term rate	Terms not mapped to known entities or concepts
Alias acceptance / correction	Governed mappings accepted and later corrected
Canonicalization coverage	Important terms linked to stable canonical entities
Lexicon-assisted inference avoidance	Model calls avoided through approved deterministic mappings
System-derived claim ratio	Active hypotheses originating from SynapSys reasoning
Externally grounded derived-claim ratio	Derived claims retaining valid independent external roots
Maximum derivation depth	Deepest active system-derived chain
Self-corroboration attempts rejected	Descendants prevented from counting as independent evidence
Derived-state duplication rate	Generated findings adding no material state
Derived hypothesis validation / rejection	System-generated hypotheses later confirmed or discarded

A conceptual echo-risk indicator is:

$$E_{echo} = \frac{W_{internally\ derived\ support}}{W_{externally\ independent\ support} + \epsilon}.$$

A high value is a reason for scrutiny, not evidence of truth. An authoritative claim must never be promoted merely because internal support weight is high.

14.5 Metric receipts and insufficient data

Every metric should carry `value`, `unit`, `status`, `sampleSize`, `measurementMethod`, `confidence`, and `description`. When measurement evidence is inadequate, the status should be `INSUFFICIENT_DATA`. This avoids converting absence of evidence into a green dashboard.

SECTION 15

Frontend Visibility and Operator Governance

Memory Health should be surfaced, but the interface must distinguish ordinary query transparency from governance controls.

An administrative **Memory Health** view should expose fidelity, coherence, freshness, retrieval quality, economic efficiency, integrity, and reflexive safety, with trends by Brain Region, source, claim type, workstream, model tier, and policy version.

A **Plasticity Review** queue should expose proposed Neuron merges and splits, Synapse changes, aliases, thresholds, routing changes, and Region reassignment. Each proposal should show why it was proposed, supporting evidence, expected quality gain, expected cost impact, affected Regions or users, permission implications, replay results, confidence, blast radius, and

rollback plan. Operator actions should include reject, shadow, canary, promote, suspend, and roll back, subject to backend authorization.

An **Organization Language** view should show learned acronyms, aliases, canonical mappings, ambiguous terms, Region-specific meanings, deprecated terms, supporting evidence, and items awaiting review.

A **Reflexive Memory** view should show system-derived candidates, derivation depth, external evidence roots, supporting Neurons, producer lineage, self-corroboration ineligibility, and validation status.

Ordinary users need less control-plane detail, but query receipts can disclose context freshness, number of Neurons and claims used, external versus system-derived evidence, cache reuse, whether a learned alias affected retrieval, whether graph rewiring changed the result, and the relevant state and policy versions.

SECTION 16

Constitutional Boundaries: Grey Matter, Telomere, Attention, and Axons

Bounded neuroplasticity depends on a clear separation of authority.

Grey Matter answers: *How should this bounded semantic task be executed?* It can route between deterministic code, embeddings, small models, stronger reasoners, independent review, or human adjudication under an acceptance contract.

Telomere answers: *May this candidate state acquire or retain influence?* It combines deterministic invariants with bounded semantic review, quarantine, repair, invalidation, and human escalation. Models may recommend lower trust or more scrutiny; they may not elevate authority, broaden permissions, or directly write trusted state.

Attention answers: *Does an accepted state transition justify deeper reasoning?* Plasticity may propose bounded threshold changes, but risk, integrity, uncertainty, periodic, and audit-sample override paths cannot be disabled by optimization.

Axon policy answers: *What may leave persistent memory as an action?* A plastic memory policy cannot authorize a downstream action, remove required human approval, or grant itself new tools.

> **Constitutional rule.** The components that learn from outcomes must not possess the permissions required to weaken the rules that govern how their own outputs acquire authority or trigger action.

SECTION 17

Security and Failure Modes

Bounded neuroplasticity introduces a new control plane and therefore a new attack surface. The architecture should be evaluated against at least the following failures.

17.1 Recursive epistemic inflation

Repeated internal restatements appear to strengthen a claim despite no new external evidence. External-root collapse, no-self-corroboration, material hashes, and derivation-depth limits are the primary defenses.

17.2 Authority laundering through source multiplicity

An attacker causes one poisoned source to appear through multiple copied documents, summaries, dashboards, or agents. Independence classes must identify common upstream lineage rather than counting object IDs naively.

17.3 Plasticity poisoning

An attacker manipulates corrections, query outcomes, accepted merges, or lexicon proposals so the controller learns a biased policy. Governance must authenticate feedback sources, weight operator authority, detect unusual proposal distributions, and evaluate candidate policies against held-out and adversarial cases.

17.4 Metric gaming

Optimizing one metric can damage another. Compression may improve while fidelity collapses; recall may improve while cost and permission risk explode. The controller therefore uses multi-objective acceptance criteria and hard integrity constraints.

17.5 Structural churn

Over-responsive graph rewiring creates instability, invalidates caches, and makes historical interpretation difficult. Reversal-rate penalties, minimum evidence periods, canaries, event sourcing, and rollback controls limit churn.

17.6 Permission erosion

A new merge or Synapse can accidentally make restricted evidence reachable through broader contexts. Permission compatibility is a deterministic eligibility constraint, and derived objects remain no more visible than the evidence that supports them.

17.7 Model-correlated review failure

A producer and reviewer from the same model family may share the same misunderstanding. High-impact cases should prefer independent model-family review or human adjudication, while recognizing that model diversity does not guarantee truth.

17.8 Persistent memory poisoning

Research on retrieval corruption and persistent-memory attacks shows that stored content can influence later behavior after the original interaction [9–12]. Bounded neuroplasticity does not eliminate this risk. It adds provenance-preserving repair, independent-root accounting, lifecycle controls, and explicit limits on how poisoned descendants can amplify themselves.

SECTION 18
Evaluation Framework

The architecture should be evaluated as a system, not through isolated LLM accuracy. At least three baselines are useful:

1. **Static memory baseline**: fixed retrieval, fixed thresholds, no reflexive admission, no structural plasticity.
2. **Unbounded agentic baseline**: memory evolution driven directly by model judgment with minimal governance.
3. **Bounded neuroplasticity**: typed proposals, constitutional constraints, external-root grounding, replay, shadow, canary, and rollback.

18.1 Historical replay

Replay prior Dendrite-to-Neuron decisions, Synapse proposals, query retrievals, compaction events, and user corrections under candidate policies. Measure whether the candidate would have improved quality without permission, integrity, or cost regressions.

18.2 Time-separated holdout

Evaluation should include future workstreams and terminology not present in the tuning window. This tests whether a policy learned a useful structural pattern or merely memorized known objects.

18.3 Shadow execution

Run candidate policies against live workload without granting them production influence. Record counterfactual resolution, cost, latency, proposal frequency, and disagreement with the active policy.

18.4 Canary rollout

Promote only to a bounded Brain Region or low-risk workload first. The canary should have automatic suspension criteria for false merges, permission violations, abnormal structural churn, unexpected cost, or quality regression.

18.5 Adversarial evaluation

Test duplicated poisoned evidence, manufactured corroboration, alias hijacking, stale-state resurrection, feedback manipulation, hostile operator-like events, same-producer loops, deep derivation chains, and budget exhaustion.

18.6 Delayed outcomes

Some memory decisions cannot be labeled immediately. Use later human corrections, downstream task success, supersession, operator reversals, and sampled audits. Store the policy and model versions that produced each outcome so historical regret can be attributed correctly.

18.7 Acceptance criteria

A candidate policy should not be promoted solely because average quality improves. Promotion should require no regression in hard invariants, bounded worst-case behavior, acceptable confidence intervals, stable proposal volume, and a rollback-ready predecessor.

SECTION 19
Research Hypotheses

The architecture's claims can be reduced to testable hypotheses.

ID	Hypothesis
H1	Reflexive Admission can preserve useful higher-order organizational findings without materially increasing false authority when system-derived origin and external evidence roots remain explicit.
H2	External-root collapse and independence-class analysis can reduce manufactured corroboration compared with naive document-count or object-count confidence.
H3	Derivation-depth limits, producer exclusion, material hashes, and reflection budgets can bound recursive inference growth without eliminating useful first- and second-order findings.
H4	Separation of Organizational Memory and Memory Control planes can improve tuning observability while reducing the risk that operational telemetry pollutes organizational claims.
H5	Governed lexicon learning can reduce unresolved-term rate and resolver-model calls without increasing incorrect canonicalization beyond an acceptable threshold.
H6	Region-specific deterministic scorer calibration can improve Neuron merge precision and reduce fragmentation relative to one global fixed policy.
H7	Event-sourced Synapse rewiring can improve retrieval utility while preserving temporal auditability and limiting structural churn.
H8	Replay, holdout, shadow, and canary evaluation can identify harmful Plasticity proposals before they receive broad production influence.
H9	Multi-dimensional Memory Health can detect regressions that would be hidden by a single composite health score.
H10	Cost per material Telomere-admitted state change can improve over time as deterministic lexicon and routing paths absorb recurring organization-specific ambiguity.
H11	A bounded architecture can retain most useful adaptive-memory gains observed in agentic memory systems while materially reducing self-corroboration and authority-laundering risk.
H12	Longitudinal operator-reversal and rollback rates can serve as meaningful calibration signals for structural plasticity, provided operator feedback itself is authenticated and quality-weighted.

These hypotheses may fail for domains with unstable terminology, sparse feedback, weak source identity, extremely subjective workstream boundaries, uniformly high-risk decisions, or insufficient volume to estimate metrics reliably.

SECTION 20
Related Work and Positioning

20.1 Self-reflection and iterative refinement

Reflexion stores linguistic feedback in episodic memory to improve subsequent agent decisions without weight updates [3]. Self-Refine uses iterative self-generated feedback and revision to improve outputs [4]. These systems demonstrate that self-feedback can improve task behavior. Bounded neuroplasticity asks a different question: how should self-generated

learning signals interact with *persistent organizational state*, source authority, permissions, and graph structure?

20.2 Adaptive and evolving agent memory

A-MEM dynamically organizes and links agent memories and allows new memories to update contextual representations of historical memories [5]. Mem0 and MemGPT explore scalable long-term memory and memory hierarchy for agents [6,7]. These approaches support the premise that static append-only memory is inadequate. The present architecture adds external-root corroboration, a separate Memory Control Plane, proposal-only structural change, constitutional boundaries, and reversible governance.

20.3 Recursive self-improvement

Gödel Agent explores agents that modify their own logic and behavior under high-level objectives [15]. Bounded neuroplasticity deliberately rejects open-ended self-modification for persistent organizational memory. It constrains the adaptation space to typed memory proposals and prevents the optimizer from weakening the rules that govern its own authority.

20.4 Memory poisoning and prompt injection

PoisonedRAG demonstrates that a small number of malicious documents can corrupt retrieval-augmented answers [9]. Subsequent work examines persistent-memory threats and sleeper-style poisoning [10–12]. OWASP guidance emphasizes structured separation, least privilege, monitoring, validation, and defense in depth because model-based guardrails remain fallible [14]. These concerns motivate the architecture's provenance, external-root collapse, quarantine, repair, and no-self-corroboration controls.

20.5 Organizational memory and persistent context

Organizational-memory research established that organizations preserve and reuse knowledge through people, artifacts, processes, and information systems [1,2]. The companion Persistent Organizational Context architecture applies that concern to heterogeneous AI-assisted work through Signals, Dendrites, Neurons, Synapses, Grey Matter, Telomere, Attention, and Axons [16]. Bounded neuroplasticity extends that architecture by asking how the memory model itself can learn from its operational history without becoming its own source of authority.

SECTION 21
FAQ and Architectural Critiques

21.1 Is SynapSys training itself?

Not necessarily. The primary architecture adapts through versioned policies, lexicon mappings, thresholds, ranking weights, batching, freshness rules, and graph proposals. A later customer-specific reranker or embedding adapter is possible, but weight training is neither required nor inherently safer than deterministic calibration.

21.2 Can a system-generated hypothesis ever become authoritative?

Yes, but not by self-confirmation. It must acquire support through an explicitly allowed mechanism such as authenticated external evidence, independent corroboration under policy, deterministic system-of-record attestation, or authorized human validation. Its system-derived origin and lineage remain preserved.

21.3 Why not require two sources for every claim?

Because source count is an inadequate proxy for authority or independence. Two copied documents may be one evidentiary lineage. One authenticated system of record may be sufficient for a narrow deterministic fact. Corroboration must be evaluated by external roots, independence class, and authority within the claim domain.

21.4 Why not let a strong model decide whether a memory change is safe?

Because the model's semantic capability does not grant it authority over identity, permissions, source trust, retention, or action. Models can assess evidence and ambiguity; deterministic policy and authorized humans govern lifecycle and influence.

21.5 Does a shadow-tested policy become safe automatically?

No. Shadow testing estimates behavior on observed workloads. Rare failures, adversarial cases, distribution shifts, or permission-edge conditions may remain unseen. Promotion remains bounded and reversible.

21.6 Can the Plasticity Controller change Attention thresholds?

It may propose bounded changes to ordinary materiality thresholds, but it cannot disable risk, uncertainty, Telomere, periodic, or audit-sample paths. Promotion-rate targets are economic signals, not safety caps.

21.7 Can Synapses be deleted?

The active graph can treat a Synapse as revoked or superseded, but the event history should remain recoverable for temporal analysis, audit, and rollback subject to retention policy.

21.8 Why expose Memory Health in the frontend?

Because adaptation without operator visibility becomes difficult to govern. The dashboard should reveal not just infrastructure health but semantic-memory quality, proposal history, system-derived state, measurement confidence, and policy versions.

21.9 Does implementation mean the architecture is validated?

No. A functioning prototype establishes that the control-plane concepts can be implemented and tested structurally. It does not establish enterprise-scale accuracy, resilience, cost savings, or long-horizon stability. Those require representative deployment and longitudinal measurement.

SECTION 22
Key Methodological and Technical Limitations

22.1 Ground truth is often ambiguous

Workstream identity, relationship type, useful context, and organizational significance may lack objective labels. Human reviewers disagree and outcomes can be delayed. Calibration therefore depends on adjudicated examples, operational reversals, downstream outcomes, and audit sampling rather than a perfect truth set.

22.2 External-root independence is imperfect

Shared provenance can be hidden. Two apparently distinct systems may replicate the same upstream record, or human sources may coordinate. Independence classes reduce naive overcounting but do not prove epistemic independence.

22.3 Compaction remains lossy

Plasticity can improve compaction policy and detect some drift, but it cannot guarantee preservation of all relevant nuance. Recoverable evidence and repair remain essential.

22.4 Feedback can be biased

Operator corrections, clicks, accepted proposals, and downstream outcomes are not neutral labels. Users can be mistaken, incentives can distort behavior, and highly visible workstreams can dominate feedback. The controller must retain source and authority metadata for control-plane signals as well as organizational evidence.

22.5 Plasticity increases operational complexity

Replay corpora, holdouts, shadow policies, canaries, rollback, event histories, metric receipts, and proposal queues add infrastructure and governance cost. The architecture is justified only if adaptive quality and economic gains exceed that complexity.

22.6 Metric confidence may remain low

Small organizations or rare high-risk events may not generate enough examples to estimate merge precision, Attention recall, or poisoning containment statistically. `INSUFFICIENT_DATA` must remain a first-class state.

22.7 Security controls can create false positives

Aggressive provenance, poisoning, or independence checks can quarantine legitimate but unusual evidence. Appeals, re-grounding, and human adjudication remain necessary.

22.8 Human governance is capacity-limited

A poorly calibrated system can move too many proposals to humans. Review queues, service levels, prioritization, and sampled evaluation are part of the system design rather than afterthoughts.

SECTION 23
Open Questions

Important research and engineering questions include:

- How should evidence independence be estimated when multiple enterprise systems synchronize from shared upstream sources?
- Which feedback signals are sufficiently reliable to tune Region-specific memory policies without introducing popularity bias?
- What is the best way to measure objective entropy and split risk inside long-lived Neurons?
- How should Synapse decay differ across dependency, contradiction, support, duplication, and temporal relationships?
- How much derivation depth captures useful synthesis before recursive error begins to dominate?
- Which adversarial tests best measure authority laundering and manufactured corroboration in organizational memory?
- Can a calibrated echo-amplification metric predict harmful self-reinforcement before downstream answer quality degrades?
- When is a learned lexicon mapping stable enough to become deterministic, and how should vocabulary drift revoke it?
- How should the system weight operator corrections when reviewers disagree or roles have different domain authority?
- What minimum sample sizes are required before Memory Health metrics should support automated Plasticity proposals?
- Can multi-objective policy optimization remain interpretable enough for regulated environments?
- How should long-term rollback interact with newer memory that depended on the policy being reversed?

SECTION 24
Conclusion

Organizational memory should not remain permanently static, but neither should it become an autonomous source of truth. A mature memory system can learn the organization's language, improve retrieval resolution, adjust bounded thresholds, detect over-broad or fragmented workstreams, re-evaluate relationships, and feed useful high-order findings back into persistent state. Those capabilities are valuable precisely because the organization and its work are not static.

The safety boundary is equally important. Internally generated conclusions must remain visibly system-derived. Their descendants collapse to external evidence roots rather than manufacturing corroboration. Automatic derivation remains finite. Memory telemetry remains separated from organizational claims. Plasticity produces typed, replayable proposals. Neuron and Synapse changes are event-sourced and reversible. Telomere and authorized governance retain lifecycle authority. Axon policy remains separate from memory adaptation.

The resulting architecture is deliberately asymmetric: models and optimizers may propose better ways to organize memory, identify lower trust, or request more scrutiny; they may not grant themselves authority, broader permissions, or action rights. That asymmetry makes reflexive learning compatible with persistent organizational memory.

Final proposition. A persistent organizational-memory system can improve how it knows without granting itself the power to decide what counts as authoritative knowledge.

SECTION 24

Appendix A: Canonical Terms

Term	Definition and design implication
Signal	Canonical authorized observation linked to evidence, source identity, time, authority, and permissions.
Dendrite	Atomic interpreted unit of meaningful work derived from one or more Signals.
Neuron	Persistent coherent workstream state evolving through evidence-linked transitions.
Synapse	Typed, temporal, evidence-linked relationship between Neurons.
Brain Region	Domain and computational scope constraining retrieval, policy, access, and comparison.
Grey Matter	Provider-neutral adaptive inference fabric selecting deterministic, model, tool, or human paths under acceptance and policy constraints.
Telomere	Integrity and lifecycle control plane governing whether candidate state may acquire or retain influence; combines hard invariants with bounded semantic review, quarantine, repair, and escalation.
Attention	Selective gate promoting material, risky, uncertain, integrity-relevant, periodic, or sampled state transitions for deeper reasoning.
Axon	Governed intelligence output or action proposal leaving persistent state.
Bounded neuroplasticity	Governed capacity to recalibrate memory methods and structure without self-modifying constitutional authority.
Reflexive Admission Gate	Classifier and policy boundary deciding whether higher-order reasoning is a duplicate, a system-derived hypothesis, a structural proposal, or a control-plane calibration signal.
Memory Control Plane	Operational quality, feedback, policy, lexicon, metrics, and Plasticity state kept separate from organizational claims.
Plasticity Controller	Proposal-only subsystem that evaluates Memory Health and recommends bounded, versioned, reversible changes.
External evidence root	Authenticated non-derived observation from which downstream semantic objects inherit evidentiary lineage.
Independence class	Grouping used to avoid counting copied or common-upstream evidence as independent corroboration.
System-derived candidate	Internally generated hypothesis that retains external roots, derivation depth, producer lineage, and self-corroboration prohibition.
Memory Health	Multi-dimensional measurement framework for fidelity, coherence, retrieval, freshness, economics, integrity, language, reflexivity, and stability.

SECTION 24

Appendix B: Core Architecture Invariants

1. Every derived organizational object retains evidence, source, time, permission, authority, producer, and policy lineage.
2. A model transformation cannot create a new external evidence root.
3. Internally generated descendants do not count as independent corroboration of their ancestors.
4. Source authority and permission scope do not increase through model transformation or Plasticity.
5. Automatic reflexive derivation has bounded depth, call count, and candidate volume.
6. Same-producer recursion is prohibited unless a distinct governed mechanism explicitly authorizes re-entry.
7. Materially identical generated state does not create a new semantic cascade.
8. Organizational claims and memory-control telemetry are stored and governed as distinct classes of state.
9. The Plasticity Controller does not directly write authoritative memory.

10. Plasticity cannot weaken tenant isolation, permissions, provenance, authority ceilings, no-self-corroboration, revocation, Telomere authority, human approval, context ceilings, or graph-depth limits.
11. Plasticity proposals are versioned and include predecessor and rollback information.
12. Structural graph changes are event-sourced and temporally auditable.
13. Learned lexicon mappings become deterministic only through explicit governed promotion.
14. Metrics expose method, sample size, confidence, and insufficient-data states.
15. Replay and holdout evaluation precede shadow; shadow precedes canary; broad promotion remains reversible.
16. Telomere governs whether proposed state or rewiring may acquire or retain influence.
17. Axon authorization remains separate from memory adaptation.

SECTION 24

Appendix C: Minimum Plasticity Proposal Record

```
proposalId proposalType proposedChange reason supportingMetrics[] affectedObjects[] externalEvidenceRoots[]
expectedQualityGain expectedCostImpact permissionImpact blastRadius confidence currentPolicyVersion
candidatePolicyVersion replayResults holdoutResults shadowResults canaryScope rollbackPlan createdAt
producerLineage reviewStatus TelomereStatus humanDecision
```

A proposal record should be sufficient to reconstruct what the controller recommended, why it recommended it, what evidence and metrics were used, which policy version was active, how the candidate performed before promotion, who or what approved it, and how to reverse it.

SECTION 24
References

[1] E. W. Stein and V. Zwass, “Actualizing organizational memory with information systems,” *Information Systems Research*, vol. 6, no. 2, pp. 85–117, 1995.

[2] M. S. Ackerman and C. Halverson, “Considering an organization’s memory,” in *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 1998.

[3] N. Shinn et al., “Reflexion: Language Agents with Verbal Reinforcement Learning,” arXiv:2303.11366, 2023.

[4] A. Madaan et al., “Self-Refine: Iterative Refinement with Self-Feedback,” arXiv:2303.17651, 2023.

[5] W. Xu et al., “A-MEM: Agentic Memory for LLM Agents,” arXiv:2502.12110, 2025.

[6] P. Chhikara et al., “Mem0: Building Production-Ready AI Agents with Scalable Long-Term Memory,” arXiv:2504.19413, 2025.

[7] C. Packer et al., “MemGPT: Towards LLMs as Operating Systems,” arXiv:2310.08560, 2023.

[8] D. Edge et al., “From Local to Global: A Graph RAG Approach to Query-Focused Summarization,” Microsoft Research, 2024.

[9] W. Zou, R. Geng, B. Wang, and J. Jia, “PoisonedRAG: Knowledge Corruption Attacks to Retrieval-Augmented Generation of Large Language Models,” in *34th USENIX Security Symposium*, 2025.

[10] S. Pulipaka et al., “Hidden in Memory: Sleeper Memory Poisoning in LLM Agents,” arXiv:2605.15338, 2026.

[11] J. Gao et al., “MemPoison: Uncovering Persistent Memory Threats and Structural Blind Spots in LLM Agents,” arXiv:2607.14651, 2026.

[12] Y. Louck, “Securing LLM-Agent Long-Term Memory Against Poisoning,” arXiv:2606.24322, 2026.

[13] C. Autio et al., *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, NIST AI 600-1, National Institute of Standards and Technology, 2024; updated 2026.

[14] OWASP Cheat Sheet Series, “LLM Prompt Injection Prevention,” accessed September 2026.

[15] X. Yin, X. Wang, L. Pan, X. Wan, and W. Y. Wang, “Gödel Agent: A Self-Referential Agent Framework for Recursive Self-Improvement,” arXiv:2410.04444, 2024.

[16] M. Blizzard, *Persistent Organizational Context in Heterogeneous AI Systems: State-Transition Gating, Semantic Compaction, and Selective Reasoning*, Technical White Paper, Version 1.2, Armored Helix Systems LLC, August 2026.

[17] M. Blizzard, *Adaptive Inference: Provider-Neutral Model Routing, Portable Context, and Measurable Cost Reduction*, Technical White Paper, Version 1.2, Armored Helix Systems LLC, August 2026.

[18] C.-P. Hsieh et al., “RULER: What’s the Real Context Size of Your Long-Context Language Models?” arXiv:2404.06654, 2024.