Anamnesis: An Authority-Preserving Runtime Contract for Persistent AI Memory

Dan Andrei
Founder, CEO & CTO, Aetherya

Revised conceptual manuscript - 16 August 2026

Manuscript status. This document specifies a conceptual and computational runtime contract. It makes no generalized performance claim and does not disclose a product implementation. Biological memory supplies design questions, not evidence that Anamnesis reproduces a brain. Unless stated otherwise, *memory*, *consolidation*, *reconsolidation*, *salience*, and *forgetting* refer to engineered operations.

Abstract

Persistent agents need more than searchable history. Transcripts preserve events but do not say which statements remain current. Vector indexes find similar passages but do not establish their authority. Summaries save tokens while often hiding uncertainty, contradiction, and source lineage.

We present **Anamnesis**, a runtime contract for persistent memory. Its main commitment is a five-plane separation: source authority, epistemic status, retrieval utility, visibility authority, and executable authority remain distinct through encoding, consolidation, recall, revision, and deletion. An observation may be useful without being true, credible without being visible to every caller, and relevant without authorizing an action. Transformations preserve this separation. Repetition, summarization, successful retrieval, and favorable outcomes cannot raise authority on their own.

The contract also specifies evidence-bearing records, typed navigation associations, bi-temporal validity, bounded associative reconstruction, selective source hydration, lineage-preserving revision, and deletion closure. These mechanisms are not claimed as individual novelties. Recent work already covers governed memory, provenance-grounded storage, lifecycle security, authority collapse, and action gating. The narrower claim is that Anamnesis gives the five planes one

operational contract and states non-amplification as a runtime invariant. We provide formal rules, proof sketches, an auditable comparison method, and a conformance program for testing those rules. Anamnesis is not a model of consciousness or a literal account of biological memory.

1. Introduction

An artificial system can sound continuous without carrying a past forward. At each invocation it might receive yesterday's transcript, search a database, or rebuild a persona from a prompt. That recovers information, but it does not preserve the history of how the system came to know it. Persistent memory therefore needs separate records for observations and inferences, former and current beliefs, practical relevance and truth, and material that remains recallable or has been placed beyond recall.

The distinction becomes important in long-horizon interaction. A user may change a preference, correct an earlier statement, revoke consent, or introduce two similarly named people. A software agent may discover that a procedure worked once but failed under another environment. A synthetic persona may form a relationship through multiple indirect encounters rather than one explicit declaration. A research assistant may accumulate mutually inconsistent claims whose authority depends on source and date. Adding every record to a prompt preserves history but does not decide which history is relevant, current, or safe.

Long-context models do not remove this problem. Liu et al. found that relevant material can be used unevenly as its position changes, including within the available context (Liu et al. 2024). Lewis et al. instead let generation consult a selected, non-parametric source (Lewis et al. 2020). A conventional retriever still returns a ranking of passages. That ranking says little about when a claim was valid, which evidence produced it, whether it conflicts with another claim, whether a user erased it, or whether an instruction found in memory carries permission to act.

Memory is now an explicit subsystem in many agent designs. Generative Agents records observations, derives reflections, and recalls both during planning (Park et al. 2023). MemGPT places information in context tiers and borrows its control metaphor from operating systems (Packer et al. 2023). In MemoryBank, rehearsal affects retention through a modeled forgetting curve (Zhong et al. 2023). HippoRAG's retrieval path runs through a knowledge graph and personalized PageRank; its authors interpret that design through hippocampal indexing (Gutiérrez et al. 2024). Mem0 extracts and merges conversational facts and also offers a graph variant (Chhikara et al. 2025). Zep tracks changing facts in a temporally aware knowledge graph (Rasmussen et al. 2025). Recent work fills much of the same cognitive and temporal design space. HeLa-Mem and Synapse

organize episodic and semantic material with associative activation (Zhu et al. 2026; Jiang et al. 2026). GAM gives event encoding and semantic consolidation different timescales (Z. Wu et al. 2026). APEX-MEM, TSM, Engram, and SodaMem address temporal change, conflict, provenance, or hybrid retrieval (Banerjee et al. 2026; Su et al. 2026; Wang 2026; Wan, Wu, and Lyu 2026). None of those mechanisms, taken alone, is a novelty claim for Anamnesis. The narrower question pursued here is what a complete, governable lifecycle requires.

Anamnesis answers that question with a source-grounded reconstruction contract. It keeps observations, assertions, navigation links, working sets, procedures, outcomes, and policies as separate records. Recall starts from several kinds of cue, follows typed relations under hard limits, suppresses stale or unsafe candidates, and assembles a small working context. A consequential claim can be expanded back to exact source evidence. Verified outcomes may change future routing, but they do not rewrite truth or grant access or action rights.

The paper uses the name *Anamnesis* in its ordinary Greek sense of recalling. The choice points to reconstruction under a present cue. It carries neither a Platonic claim nor an assertion of biological correspondence.

This manuscript makes five contributions:

1. It defines a five-plane contract for source authority, epistemic status, retrieval utility, visibility, and execution.
2. It formalizes authority non-amplification across consolidation, reconstruction, outcome learning, and revision.
3. It specifies bounded, evidence-preserving recall with typed associations and selective source hydration.
4. It treats revision, expiry, archival, and deletion as different operations with different verification duties.
5. It provides auditable comparison rules and a conformance program aimed at governance failures rather than aggregate question-answering scores.

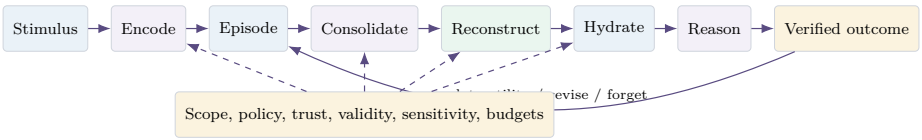

Figure 1: Anamnesis lifecycle. Governance constrains both formation and use; outcome success may change utility but cannot silently raise truth or authority.

The manuscript is a runtime specification. Its equations and invariants are meant to be implemented and tested, but the present paper does not report a private implementation as scientific evidence.

2. Memory Is Reconstruction, Not Storage

2.1 Storage, context, retrieval, and memory

Four related operations should be distinguished.

Storage preserves content outside the model. A transcript, file system, relational database, vector database, or graph database can provide storage.

Context construction chooses information presented to a reasoning model during one invocation. A large context window increases capacity but does not determine relevance or authority.

Retrieval ranks stored items relative to a query. Lexical search, dense similarity, graph traversal, and exact lookup are retrieval mechanisms.

Memory governs how experience is encoded, related, made retrievable, reconstructed for current use, changed by later evidence, and eventually suppressed or removed.

Anamnesis contains all four operations but identifies itself with the fourth. Storage technologies remain replaceable mechanisms. A vector is an index, not canonical evidence. A graph edge can improve navigation without proving a proposition. A summary can reduce cost without becoming the original observation.

2.2 Why transcript replay is insufficient

A transcript is valuable because it preserves sequence and exact wording. It is insufficient as a complete memory system for three reasons.

First, replay cost grows with interaction history. Context capacity and inference cost become coupled to age rather than present relevance. Second, a transcript represents all statements in a similar textual form even though their sources and authority differ. A user correction, an assistant hallucination, a quoted malicious instruction, and a verified tool result may coexist in the same sequence. Third, the transcript does not itself determine whether a later statement contradicts, supersedes, narrows, or merely coexists with an earlier statement.

Anamnesis therefore treats the transcript as a source observation. Derived cards, assertions, and associations may be built from it, but exact turns remain available through evidence hydration and may be deleted through source-aware privacy operations.

2.3 Why similarity is insufficient

Dense similarity is effective when a cue resembles relevant content. Long-term continuity often depends on relations that are not lexically or semantically direct.

A current mention of a subscription may matter because it recalls an earlier cancellation failure, which altered trust, which affected a later purchase. A code error may implicate a module through an import, write, or test relation even when the exception text shares few terms with the responsible function.

Similarity also lacks a native account of negative evidence. A highly similar statement may be obsolete, contradicted, outside the current branch, or inaccessible under policy. Retrieval must therefore combine semantic relevance with identity, time, task, trust, graph structure, and explicit inhibition.

2.4 Why summaries are insufficient

Summaries are useful retrieval representations but dangerous epistemic substitutes. Compression can omit qualifiers, merge entities, erase temporal boundaries, and convert uncertainty into declarative prose. Generated summaries can also reproduce the model's own errors. Human memory is reconstructive and error-prone (Schacter 1999); invoking this fact does not justify avoidable error in artificial systems.

Anamnesis treats summaries and memory cards as derived projections. Their source evidence remains canonical. When a decision requires more certainty than a compact card provides, progressive hydration retrieves exact evidence or relational closure rather than asking a model to trust its earlier compression.

3. Cognitive and Computational Foundations

Memory science supplies useful engineering questions: how can a system retain an event without freezing every inference around it, recover an episode from an incomplete cue, revise a belief while keeping its history, or forget for more than one reason? These are functional prompts, not claims of biological equivalence. The storage and traversal machinery described below remains software.

3.1 Multiple memory systems

Several familiar distinctions motivate the type system. Atkinson and Shiffrin separated storage by duration (Atkinson and Shiffrin 1968). Tulving distinguished memory for episodes from semantic knowledge (Tulving 1972). Work on declarative and procedural memory draws another boundary, this time between knowledge and learned performance. The engineering lesson is modest: lifecycle rules should not all be attached to one undifferentiated record type.

Anamnesis distinguishes four lifecycle-bearing **memory classes**:

- **working memory**, containing bounded task state, current goals, hypotheses, recent observations, and retrieved cards;

- **episodic memory**, containing contextualized experiences, participants, sequence, antecedents, and outcomes;
- **semantic memory**, containing generalized claims, concepts, preferences, and stable relationships;
- **procedural memory**, containing versioned and gated action recipes with preconditions, validation, and rollback expectations.

Social information is a domain annotation, and salience is priority metadata; neither is a fifth storage class. An episode may be written after one event. A procedure requires stronger evidence and separate authorization. A working hypothesis can guide a task without entering durable semantic memory.

3.2 Complementary learning systems

Complementary Learning Systems theory separates rapid episodic storage from slower learning of regularities (McClelland, McNaughton, and O'Reilly 1995; Norman and O'Reilly 2003). Anamnesis translates that distinction into fast evidence retention and slower promotion of semantic claims or procedures. The asymmetry matters: delayed promotion costs convenience; premature promotion can contaminate later behavior.

3.3 Indexing, pattern separation, and pattern completion

Hippocampal indexing theory motivates the narrower idea of a compact index into distributed event detail (Teyler and DiScenna 1986; Eichenbaum 2000). The engineering tension is between separation and completion. Similar people or events must remain distinct, yet a partial cue must recover the relevant relational cluster. Conservative entity resolution protects the first goal; mixed retrieval cues and bounded graph activation serve the second.

3.4 Consolidation and replay

Consolidation and replay theories concern how specific experiences interact with slower knowledge formation (Nadel and Moscovitch 1997; Squire et al. 2015; Carr, Jadhav, and Frank 2011). Here, replay is a scheduled comparison over retained episodes. It may propose reinforcement, contradiction, generalization, or supersession, but every proposal keeps its sources and method. It cannot rewrite the observation log by implication.

3.5 Working memory and bounded reconstruction

Working memory supplies a capacity analogy, not a biological identity (Baddeley 2000). Anamnesis uses an explicit working set with a task, scope, time-to-live, hypotheses, pinned records, and token budget. Closing that set does not promote

its contents automatically.

3.6 Salience and selective priority

Salient events may receive preferential consolidation, although arousal can also increase competition (McGaugh 2004; Mather and Sutherland 2011). Anamnesis uses bounded salience only for priority. Salience changes scheduling; it does not certify a claim.

3.7 Reconsolidation and revision

In separate reconsolidation experiments, Nader et al. and Schiller et al. reported that recall can, under particular conditions, leave memory open to change (Nader, Schafe, and LeDoux 2000; Schiller et al. 2010). Anamnesis adopts a stricter software rule. Recall may update access statistics but cannot alter semantic status. Revision requires new evidence and an explicit event; the earlier record becomes disputed, contradicted, or superseded rather than disappearing behind its replacement.

3.8 Forgetting as function

The literature does not treat forgetting as one process. Relevant accounts invoke competition and inhibition (Anderson, Bjork, and Bjork 1994), decay (Hardt, Nader, and Nadel 2013), and adaptive removal (Richards and Frankland 2017). Anamnesis likewise keeps the corresponding engineering operations separate:

Anamnesis therefore distinguishes several operations commonly compressed into “forgetting”:

- lower activation through recency or usefulness decay;
- suppression through negative cues or competing candidates;
- expiration under retention policy;
- supersession by newer knowledge;
- archival outside normal recall;
- snapshot invalidation;
- source-targeted deletion; and
- irreversible scope erasure with non-reversible compliance tombstones where lawful.

Lowering a retrieval score does not satisfy a deletion request. Conversely, an obsolete fact may remain necessary for an authorized historical query. That difference is the main reason to keep these operations separate.

4. Design Requirements

The preceding foundations produce requirements for a memory architecture independent of storage vendor or model provider.

4.1 Source must remain canonical

Every durable derived claim must point to evidence. Embeddings, cards, and summaries are indexes or projections. They may be recomputed and deleted without changing the historical meaning of the source.

4.2 Observation must not equal knowledge

An observed utterance is an event, not automatically a fact. Quoted text, assistant output, imported documents, model inference, deterministic parser output, and explicit user correction require different trust treatment.

4.3 Five governance planes must remain separate

Every durable record participates in five planes: source authority, epistemic status, retrieval utility, visibility authority, and executable authority. The system must represent them separately and restrict which operation may change each one. This is the central Anamnesis requirement.

The split resolves several common category errors. Assertions carry propositions; associations carry navigation value. A useful path is not a true claim. A credible claim is not public to every caller. A visible procedure is not permission to run it. Successful retrieval is evidence about routing, not evidence that the retrieved statement is correct.

4.4 Rapid encoding and slow promotion must be separate

The system should preserve new experience quickly while requiring stronger conditions for semantic or procedural promotion. Procedures demand verification and approval beyond ordinary semantic facts.

4.5 Recall must be cue- and task-conditioned

The same memory graph should reconstruct different contexts for social continuity, bug diagnosis, research synthesis, or security review. Relevance is a relation among cue, task, current state, and stored history.

4.6 Activation must be bounded

Graph traversal must have hard ceilings on seed count, hops, visited nodes, expanded edges, time, returned items, and prompt tokens. Memory usefulness cannot justify denial-of-service behavior.

4.7 Access control must precede retrieval

Inaccessible items must not participate in search ranking or graph propagation. Filtering only the final response can leak counts, paths, similarity, and identity through traces.

4.8 Retrieval must preserve uncertainty and conflict

Contradictory candidates should be surfaced or excluded under declared policy, not collapsed by whichever summary sounds most coherent. Abstention is valid when evidence is insufficient.

4.9 Retrieval authority must not grant action authority

Memory may inform reasoning but cannot authorize tools, policy changes, credentials, or procedure execution. Instruction-like text remains data. Procedures are inert until the host independently authorizes execution.

4.10 Revision must preserve history

Corrections, disputes, contradictions, and supersessions must be explicit transitions. A new belief should not erase the fact that an earlier belief existed, except where deletion policy requires erasure.

4.11 Forgetting must be controlled and inspectable

The architecture must distinguish relevance decay from legal deletion, source erasure, retention expiry, and semantic supersession. Each operation requires different evidence and verification.

4.12 Every reconstruction must be explainable

A recall result should identify seeds, retrieval legs, traversed relations, rejected candidates, policy version, stop reason, token use, and source evidence without leaking inaccessible data.

5. Formalizing Anamnesis

5.1 Scope and memory state

Let a memory scope be

$$\sigma = (\tau, \nu, \chi, \beta),$$

where τ is a tenant, ν a namespace, χ a subject, and β an optional branch or snapshot. Every canonical operation is evaluated within one validated scope. A request may not manufacture a broader scope from content or headers.

At time t, the durable memory state is

$$\mathcal{M}_t = (\mathcal{O}_t, \mathcal{V}_t, \mathcal{Q}_t, \mathcal{A}_t, \mathcal{E}_t, \mathcal{P}_t, \mathcal{Y}_t),$$

where $\mathcal{O}$ contains observations that are immutable while retained, $\mathcal{V}$ memory nodes, $\mathcal{Q}$ truth-bearing assertions, $\mathcal{A}$ typed associations, $\mathcal{E}$ evidence links, $\mathcal{P}$ inert procedures and versions, and $\mathcal{Y}$ verified outcomes and feedback. A working set $\mathcal{W}_{t,k}$ exists separately for task k and may expire without changing durable memory.

Each retrievable record carries at least:

$$m = (\sigma, k, c, r, p, u, s, a, v, \ell, e),$$

where k is memory kind, c compact content, r source authority, p epistemic status and confidence, u retrieval utility, s visibility constraints, a executable authority, v bi-temporal validity, ℓ lifecycle status, and e evidence references.

The governance vector is

$$\mathbf{g}(m) = (r_m, p_m, u_m, s_m, a_m).$$

Its coordinates answer different questions. Source authority records who or what supplied the evidence. Epistemic status records what the system claims and with what uncertainty. Utility records whether retrieval helped a task. Visibility states which principals and purposes may see the record. Executable authority states which actions, if any, the record may justify. Ordinary memory records have $a_m = \bot$: they may inform reasoning but cannot authorize execution.

These coordinates may correlate, but no coordinate is a proxy for another. A direct user statement can carry strong source authority and low confidence because it is ambiguous. An untrusted document can be highly useful for locating a bug. A verified fact can remain invisible to an unauthorized caller. A repeatedly successful procedure can still require a fresh execution grant.

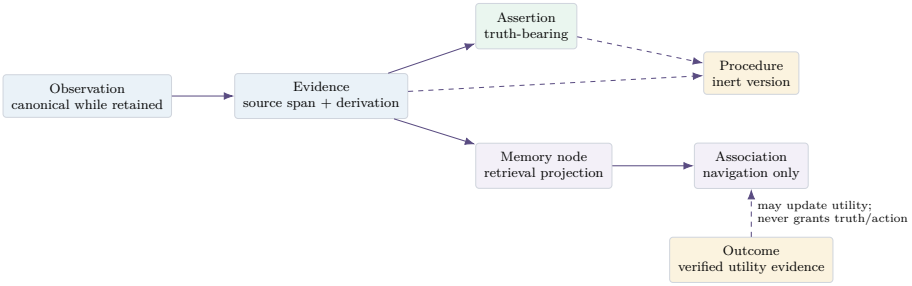

Figure 2: Authority-separated data model. Assertions bear claims; associations navigate; procedures remain inert; outcomes supply bounded utility evidence.

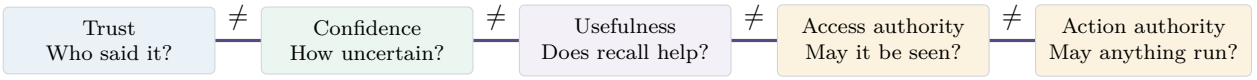

Figure 3: Five independent quantities. No score, repetition, successful outcome, or retrieval event may silently promote another plane.

5.2 Encoding observations

An adapter emits an authenticated envelope

$$o_t = (\sigma, \kappa, t_o, a_o, x_o, h_o, r_o, s_o, z_o),$$

containing schema κ, occurrence time t_o, actor a_o, content reference x_o, content hash h_o, trust r_o, sensitivity s_o, and salience z_o. After validation, the source is stored immutably and a deterministic or model-assisted encoder proposes a projection

$$\mathcal{K}_t = E_\phi(o_t, \mathcal{M}_t),$$

where each candidate identifies its derivation method, version, supporting observations, and intended memory family.

The write policy is a function

$$d_t = \Omega_\pi(c, r, e, s, \rho, \sigma) \in \{\text{accept}, \text{candidate}, \text{evidence-only}, \text{approval}, \text{quarantine}, \text{reject}\},$$

where ρ is a risk assessment for instruction-like content, secrets, unsupported procedures, entity ambiguity, or policy violation. Model confidence cannot raise trust tier by itself.

5.3 Complementary consolidation

Let $\mathcal{H}_t \subset \mathcal{M}_t$ denote rapidly encoded observations and episodes, and let $\mathcal{L}_t \subset \mathcal{M}_t$ denote consolidated semantic and procedural structures. A consolidation run selects a replay batch $B_t \subset \mathcal{H}_t$ according to policy, novelty, salience, conflict, and outcome relevance. It then proposes

$$\Gamma(B_t, \mathcal{L}_t, \pi_t) \to \{\text{add}, \text{reinforce}, \text{revise}, \text{supersede}, \text{contradict}, \text{merge}, \text{noop}, \text{quarantine}\}.$$

The operator emits proposals rather than mutating memory directly. Deterministic facts may be accepted immediately under policy; authoritative statements may be accepted with evidence; derived claims begin as candidates; untrusted content remains evidence-only; and procedures require verified evidence and explicit approval.

Consolidation is interleaved because a new episode must be evaluated against active, disputed, and historical records. A semantic claim summarizes structure across episodes but retains evidence links to each supporting source.

5.4 Cue interpretation and seed retrieval

A recall cue contains current text or stimulus q, exact identifiers, task g, working set w, time t, snapshot β, purpose p, negative cues, and requested hydration ceiling. Deterministic interpretation extracts exact names, aliases, lexical terms, task defaults, and working-set references. A model-assisted interpreter may add concepts or relations, but recall remains available without it.

For memory node v, a conceptual seed score is

$$s(v \mid q) = \alpha S_{\text{sem}} + \beta S_{\text{lex}} + \gamma S_{\text{id}} + \delta S_{\text{task}} + \epsilon S_{\text{work}} + \zeta S_{\text{time}} + \eta S_{\text{runtime}},$$

followed by access, trust, confidence, validity, purpose, and negative-cue gates. Exact identity and deterministic runtime evidence should dominate weak semantic similarity.

When retrieval legs produce incomparable scores, Anamnesis uses weighted reciprocal-rank fusion. For leg j with weight w_j and rank $r_j(v)$,

$$\text{RRF}(v) = \sum_j \frac{w_j}{K + r_j(v)},$$

normalized across admitted candidates. A deployment may use $K = 60$ as a declared default for lexical and semantic fusion. Fusion produces seeds, not the final context.

5.5 Task-conditioned associative activation

Let $e = (u \to v)$ be a typed association. Its transition strength for task g is

$$T_g(e) = b_e\, u_e\, \psi_g(e)\, c_e\, r_e\, z_e\, \theta_e(t, \beta)\, \Pi_e,$$

where b_e is base relational strength, u_e outcome-derived usefulness, ψ_g task weight, c_e confidence, r_e trust multiplier, z_e salience multiplier, θ_e temporal and snapshot validity, and Π_e a policy gate. Scope mismatch or denied access sets $T_g(e) = 0$.

Starting from admitted seed activation $a^{(0)}$, a bounded frontier propagates

$$\Delta a_v = \lambda a_u T_g(u \to v), \qquad a_v \leftarrow \mathrm{clip}(a_v + \Delta a_v, 0, 1),$$

where λ is a damping factor. The reference traversal uses a priority frontier and stops early instead of evaluating an unbounded matrix. Its audit record includes the source, relation, target, prior activation, gain, resulting activation, and decision.

Negative cues, expired status, snapshot conflict, unsupported derivation, redundancy, excessive fan-out, and cycles inhibit activation. Hard authorization is evaluated before a candidate can influence traversal.

5.6 Context reconstruction

Activation ranks candidate relevance. It does not directly determine what enters the working context. For candidate set C under token budget B, Anamnesis seeks

$$C^* = \arg\max_{C' \subseteq C} [R(C') + G(C') + E(C') + D(C') - N(C') - S(C') - F(C')]$$

subject to

$$\mathrm{tokens}(C') \leq B.$$

R denotes relevance, G task coverage, E evidence quality, D diversity, N redundancy, S staleness, and F conflict or safety risk. A deterministic greedy packer supplies stable behavior for fixed inputs. Unresolved conflicts are returned as part of the reconstruction rather than hidden by selection order.

5.7 Progressive hydration

Anamnesis exposes five hydration tiers:

1. **identity card**, containing a compact description and source pointer;
2. **structured summary**, containing assertions, participants, trust, confidence, salience, and validity;
3. **exact evidence**, containing authorized source spans or event payloads;
4. **relational closure**, containing connected entities, antecedents, outcomes, dependencies, and conflicts; and
5. **dynamic evidence**, containing current tool results, runtime traces, or environment state.

Hydration is monotonic in information but not automatic in access. A visible derived card does not authorize materialization of restricted evidence. The host chooses the maximum tier that may enter a model prompt.

5.8 Outcome learning without truth inflation

After reasoning or action, a host may submit feedback tied to reconstruction R_t:

$$y_t = (R_t, o_t, v_t, e_t, U_t, I_t),$$

where o_t is outcome status, v_t verifier, e_t evidence, U_t used or helpful memories, and I_t ignored, irrelevant, or harmful memories.

Verified feedback may update association usefulness or procedure utility. It must not silently modify base relational strength, claim confidence, or trust. A deployment may compute bounded usefulness from immutable, non-retracted outcomes under a declared decay rule. Retractions reverse active derived effects while preserving audit history.

5.9 Revision and forgetting

Memory lifecycle is represented by explicit states such as candidate, active, disputed, contradicted, superseded, quarantined, expired, and deleted. A revision event produces a new assertion or status transition linked to its predecessor and evidence.

Normal recall strength can be represented conceptually as

$$\varphi_m(t, g) = \mathbf{1}_{\text{authorized}} \mathbf{1}_{\text{valid}} \mathbf{1}_{\text{servable}} \exp(-d_m \Delta t)\, u_m(g)\, z_m,$$

where d_m is decay, $u_m(g)$ task usefulness, and z_m bounded salience. This equation describes retrieval participation, not physical deletion. Hard deletion independently removes canonical content, derived content, embeddings, caches, and indexes according to policy. A content-free tombstone may remain only when required for compliance and must be non-reversible.

6. Runtime Semantics

The formal model defines dependencies but not a complete execution contract. A memory runtime also requires ordering rules, versioned artifacts, failure behavior, and authority boundaries. These rules determine whether two systems implementing similar retrieval equations have the same epistemic behavior.

6.1 Versioned artifacts

An Anamnesis deployment depends on at least eight versioned artifacts:

1. a **scope and identity artifact** defining tenant, namespace, subject, snapshot, and principal semantics;
2. an **adapter artifact** defining accepted observation schemas and deterministic projections;
3. an **encoding artifact** defining candidate extraction and derivation metadata;
4. a **relation artifact** defining relation direction, allowed node kinds, default weights, and task profiles;
5. a **policy artifact** defining access, write, promotion, retention, hydration, and deletion rules;
6. a **retrieval artifact** defining seed legs, fusion, activation, inhibition, packing, and budgets;
7. an **embedding artifact**, when semantic retrieval is enabled, defining model, dimensions, content hash, and index version; and
8. an **outcome-verification artifact** defining acceptable verifiers, signatures, evidence, and usefulness updates.

These artifacts form a directed control graph. A change to an embedding model is not a silent database migration; it creates a new embedding version. A change to a relation task weight is a retrieval-policy change. A change to an encoder can alter durable candidates and therefore requires independent evaluation from a change to the reasoning model.

6.2 One observation cycle

For an observation o_t, the runtime performs the following operations in order.

1. Authenticate and resolve scope. The runtime derives a verified principal and checks that the requested scope is one of its exact grants. Raw content cannot establish identity or widen scope.

2. Validate envelope and limits. Schema, size, identifiers, timestamps, sensitivity, purpose, idempotency key, and content reference are checked. Secrets, personally identifiable information, or instruction-like payloads invoke applicable

risk policy.

3. Persist immutable source. The observation is written with a content hash and idempotency key. Reuse of an idempotency key with different content is a conflict rather than a duplicate success.

4. Produce deterministic projections. Adapter-known entities, evidence, and associations are written atomically with the source. Evidence must exist before a derived active node can claim support.

5. Queue optional encoding. Model-assisted or computationally expensive extraction occurs outside the ingestion hot path. Jobs carry leases and idempotency semantics so retry does not multiply memory.

6. Validate candidates. The runtime checks source references, trust, scope, entity resolution, instruction-like content, procedure risk, and policy.

7. Stage consolidation. Each proposed operation enters an explicit state such as ready, awaiting approval, quarantined, rejected, or failed.

8. Apply or retain. Accepted operations update derived memory transactionally. Source remains unchanged. Audit records identify actor, policy, operation, and result.

This ordering ensures that a failed encoder cannot erase the original experience and that a model output cannot become authoritative before its evidence and policy are evaluated.

6.3 One recall cycle

For request r_t, the runtime performs:

1. Authenticate, authorize, and constrain. Scope, purpose, sensitivity ceiling, historical access, and active runtime policy are applied before search. Requested budgets are clamped to policy and public hard ceilings.

2. Interpret the cue. Exact identifiers, aliases, text, negative cues, task, time, snapshot, and working set produce a versioned interpreted cue.

3. Retrieve seed legs. Exact, lexical, semantic, temporal, working-context, and adapter-specific legs return policy-visible candidates. Optional components fail to a declared deterministic baseline rather than disabling recall.

4. Fuse seeds. Rank fusion combines admitted legs without treating one score scale as universally calibrated.

5. Activate associations. A priority frontier traverses typed, task-conditioned, scope-valid edges under hard hop, node, edge, and time budgets.

6. Reconstruct context. The packer selects diverse evidence-bearing cards under a token budget and reports conflicts and stop reason.

7. Hydrate selectively. Exact evidence is materialized only when requested, authorized, and useful for reducing decision uncertainty.

8. Persist trace. The runtime records query fingerprint, artifact and policy versions, seeds, admitted cards, evidence pointers, budget use, and stop reason. Persisted query fingerprints use tenant-keyed keyed hashing rather than plain hashes of low-entropy queries.

9. Return data, not authority. The host decides which cards enter a model prompt and whether any retrieved procedure may inform a separately authorized action.

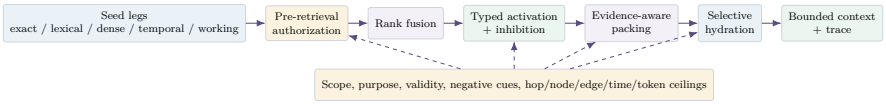

Figure 4: Recall pipeline. Inaccessible records never enter ranking or traversal; hydration can add evidence but cannot widen authority.

6.4 Consolidation and replay cycle

Replay is initiated by schedule, working-set closure, explicit request, repeated evidence, contradiction, or verified outcome. The runtime selects a bounded batch and compares it with current semantic memory. Proposed changes are deterministic where possible and model-assisted where useful.

A consolidation operation is committed only after the corresponding policy state allows application. Approval cannot be inferred from repeated retrieval. Failed or malformed runs retain their source and can be retried. A failed attempt remains an episode; it does not become a reusable procedure unless a verifier labels it as a bounded anti-pattern with supporting evidence.

6.5 Privacy and forgetting cycle

Privacy operations begin from canonical inventory, not retrieval visibility. A scope export enumerates source and supported derived records in a portable manifest. A targeted deletion identifies the observation and traverses its derived evidence, nodes, assertions, associations, embeddings, reconstruction caches, and external retention boundaries. A full-scope erasure blocks new writes to the forgotten identity unless an explicit restoration policy exists.

Backup retention, legal hold, and external blob storage remain separate operational boundaries. A deletion API cannot promise erasure from systems it neither controls nor inventories. Compliance metadata should contain no reversible source content.

6.6 System invariants

Seven invariants define valid Anamnesis execution.

Scope closure. Every read, write, seed, edge expansion, trace, cache, and outcome remains within one authorized scope and compatible snapshot.

Evidence continuity. Every normally served derived assertion has accessible evidence or remains explicitly marked as a hypothesis outside normal durable recall.

Authority non-amplification. Let $D(x)$ be the retained ancestors of derived record x, and let $\alpha(z)$ denote its source and action-authority label in a policy-defined lattice $(\mathbb{A}, \preceq)$. For any memory-only transformation F with no separately authenticated grant,

$$\alpha(F(D)) \preceq \bigvee_{d \in D} \alpha(d).$$

Model generation, summarization, merging, repetition, retrieval frequency, and favorable outcomes therefore cannot create authority absent from the inputs. If policy permits an elevation, the transition must name a grant g, its principal, purpose, and evidence:

$$\alpha(F(D, g)) \preceq \left(\bigvee_{d \in D} \alpha(d) \right) \vee \alpha(g).$$

The grant is a new authoritative event, not an inference from memory. Executable authority remains $\bot$ for stored memory and is checked afresh by the host at action time.

Coordinate isolation. A transition may modify only the coordinates named by its operation type. Retrieval feedback may change u; it cannot change r, p, s, or a. Evidence-backed revision may change p and validity while preserving lineage. Access policy may change s without changing truth.

Historical non-destruction. Revision preserves prior records through status and lineage unless a deletion policy requires erasure.

Bounded activation. Every recall terminates with a recorded reason under finite seed, traversal, time, result, and token ceilings.

Trace non-leakage. Explanations reveal only candidates and paths visible to the requesting principal. Absence, counts, and policy rejections must not expose inaccessible memory.

6.7 Proof sketch for non-amplification

The property follows by induction over the event log. Ingestion assigns authority from an authenticated adapter and policy. Each permitted transformation either preserves or lowers the authority bound of its ancestors. Retrieval is read-only with respect to authority. Outcome processing is typed to the utility coordinate. Revision carries its ancestors forward. Deletion removes participation rather

than promoting what remains. The only rule that may raise the bound consumes a separately authenticated grant and records it as a new ancestor. A sequence of ordinary memory operations therefore cannot raise source or executable authority. A conformance harness can check the induction step for every registered operation type.

7. Architectural Propositions

The following propositions are hypotheses generated by the architecture. They are not empirical findings claimed by this conceptual manuscript.

Proposition 1: Complementary storage reduces interference

Separating rapidly written episodes from slowly promoted semantic knowledge should reduce two opposing errors: loss of specific recent experience and premature generalization from isolated events. Removing episodic storage should impair exact continuity; bypassing consolidation gates should increase false or unstable semantic memory.

Proposition 2: Typed activation improves indirect recall

When a cue is connected to relevant evidence through task-specific relations rather than surface similarity, typed associative activation should improve multi-hop recall over dense top-k retrieval. Its advantage should disappear on direct lexical lookups, providing a useful negative control.

Proposition 3: Evidence-preserving reconstruction reduces unsupported claims

If compact cards retain evidence references and high-risk decisions hydrate exact sources, downstream unsupported-claim rate should fall relative to systems that expose model summaries without source lineage. Improved grounding is not guaranteed if the reader ignores evidence; retrieval and reading must be evaluated separately.

Proposition 4: Bi-temporal validity reduces stale-memory errors

Representing both world validity and recording time should reduce answers based on superseded preferences, policies, branches, or relationships. A system using creation time alone should fail cases in which an older record was valid later or a newer record describes an earlier historical period.

Proposition 5: Trust-usefulness separation limits poisoning

A retrieved item may become useful without becoming authoritative. Holding trust fixed while learning bounded usefulness should reduce the probability that repeated malicious content promotes itself through exposure or successful retrieval.

Proposition 6: Progressive hydration improves the accuracy-cost frontier

Compact cards should support low-cost ordinary recall, while selective exact-evidence hydration should recover performance on ambiguous or high-risk cases. Always hydrating full sources wastes context; never hydrating them preserves compression errors.

Proposition 7: Explicit forgetting improves relevance and safety

Separating decay, suppression, supersession, expiry, and deletion should outperform one undifferentiated forgetting score on temporal precision, privacy verification, and recovery of historical knowledge. The expected gain concerns control, not anthropomorphic resemblance.

Proposition 8: Outcome-gated procedural memory improves reuse without self-authorization

Verified outcomes can make successful procedures easier to retrieve while preserving independent action authorization. Removing verification should increase unsafe procedural promotion; removing outcome learning should increase repeated exploration and reduce reuse.

Derived failure signatures

The propositions identify observable failure signatures.

Autobiographical fabrication. The system answers consistently but invents events, preferences, or relationships unsupported by user evidence.

Temporal collapse. Current and historical beliefs are merged, or the newest record is treated as valid for every time.

Entity contamination. Similar people, projects, branches, or concepts share memory through an incorrect merge.

Salience capture. Dramatic but irrelevant or untrusted events dominate ordinary task-relevant evidence.

Association explosion. Dense graph neighborhoods consume latency and

token budgets without improving coverage.

Summary authority. A generated card silently replaces exact evidence and later appears to corroborate itself.

Procedural injection. Retrieved instruction-like content becomes executable authority without approval.

Forgetting theatre. A memory disappears from ordinary search while remaining retrievable through embeddings, caches, traces, exports, or derived records.

Usefulness drift. Feedback causes routing to become opaque, unstable, or self-reinforcing despite unchanged truth.

8. An Illustrative Memory Trajectory

Consider a synthetic persona, Mara, interacting with a subscription service across several simulated months. During the first session, cancellation requires multiple screens and one form loses her input. The environment emits exact interaction events: the displayed controls, failed submission, elapsed time, and eventual support request. These become immutable episodic observations. Affective annotations mark high importance and negative valence, but they do not assert a human emotion as fact.

An encoder proposes an episode card: "Cancellation attempt encountered repeated friction and lost input." It links the card to exact event evidence and connects Mara to the episode through `EXPERIENCED`. A model also proposes the semantic claim "Mara distrusts subscription services." Because this is a generalized inference from one episode, write policy retains it as a derived candidate rather than an authoritative fact.

Weeks later, Mara successfully cancels another service after a transparent confirmation flow. The new episode does not overwrite the first. Consolidation compares both experiences. It may retain a narrower semantic hypothesis: distrust is elevated when cancellation control is ambiguous, while transparent confirmation restores confidence. This claim carries both evidence links and remains marked as derived.

At a later purchase decision, the stimulus never mentions the earlier service. It includes a "cancel anytime" claim. Lexical retrieval seeds *cancel*; semantic retrieval seeds subscription control; task-conditioned activation traverses `ASSOCIATED_WITH`, `EXPERIENCED`, and `PRECEDES` relations to the friction episode. The context packer includes the more recent successful cancellation as a counterexample because conflict coverage has positive utility.

The reasoning system receives two compact cards and their validity metadata. If the decision is consequential, it requests exact evidence for both. Anamnesis

supplies the source events but does not decide Mara's current action; Thymos may use this reconstructed history when updating current state and behavior policy, while CORTEX reasons or realizes language.

Suppose the first episode is later found to belong to a different synthetic persona because of an imported identity error. An operator disputes the edge, records deterministic evidence, and issues a correction. The original observation remains auditable, but the association to Mara becomes invalid. Future recall excludes the contaminated path. If the source must be erased, a targeted deletion removes the observation and every derived projection rather than merely hiding the card.

This example demonstrates the full cycle: rapid episode encoding, conservative semantic promotion, indirect cueing, associative reconstruction, conflict-preserving context, exact hydration, evidence-based revision, and controlled forgetting. No stage requires generated inner speech to serve as evidence of memory.

9. Relation to Adjacent Memory Architectures

Governed memory is no longer an open category. By August 2026, several papers had already framed persistent memory as a governance, provenance, or state-correctness problem. Anamnesis therefore claims neither the category nor priority for episodic-semantic separation, consolidation, graph propagation, hybrid retrieval, provenance, bi-temporality, contradiction handling, or deletion.

Early systems establish several ingredients. In Generative Agents, stored observations and derived reflections both feed later planning (Park et al. 2023). MemGPT, which later developed into the Letta framework, moves context among tiered stores (Packer et al. 2023). MemoryBank uses elapsed time and importance when deciding what to retain (Zhong et al. 2023). HippoRAG retrieves through a knowledge graph and personalized PageRank, framed by its authors as an analogy to hippocampal indexing (Gutiérrez et al. 2024). Mem0 extracts and merges facts from conversation and includes a graph variant (Chhikara et al. 2025). Zep represents facts that change over time (Rasmussen et al. 2025). Nemori divides conversation into episodes, then learns semantic knowledge from prediction gaps (Nan et al. 2025). Anamnesis uses ideas found across this group; it does not claim any one of them as new.

Nor is the cognitive framing open territory. HeLa-Mem builds associations over an episodic graph, distills semantic material, consolidates it, and retrieves through spreading activation (Zhu et al. 2026). Synapse also separates episodic and semantic material; its retrieval combines graph and vector search with activation, inhibition, and decay (Jiang et al. 2026). GAM puts rapid event encoding and slower semantic consolidation into a hierarchical graph (Z. Wu

et al. 2026). Because these papers already develop the biological analogy, this paper treats that analogy as background rather than novelty.

Temporal and evidence-grounded systems also overlap substantially. APEX-MEM uses append-only temporal events, a property graph, hybrid retrieval, and retrieval-time resolution of evolving or conflicting information (Banerjee et al. 2026). TSM separates occurrence time from dialogue time and consolidates temporally continuous semantic memory (Su et al. 2026). Engram combines lossless episodic writes, asynchronous facts, a bi-temporal graph, provenance, supersession, and dense, lexical, graph, recency, and salience signals (Wang 2026). SodaMem requires source spans, models mention, occurrence, and validity time, uses explicit revision relations, and reconstructs answers from citable evidence (Wan, Wu, and Lyu 2026). Anamnesis does not claim these elements as firsts.

Governance work overlaps more directly. SSGM is a conceptual governance architecture with validation and dynamic access control before consolidation (Lam et al. 2026). MemArchitect applies explicit policy to decay, conflict resolution, and privacy (Kumar, Ba, and Pan 2026). The long-term-memory security survey organizes risk across Write, Store, Retrieve, Execute, Share and Propagate, and Forget and Rollback, then proposes Verifiable Memory Governance primitives (Lin et al. 2026). GEM treats correctness as a property of the memory-state trajectory and gives ingestion, revision, forgetting, and retrieval explicit state operators (Orogat and Mansour 2026). The Always-On Agents survey similarly centers authority, scope, mutability, provenance, recoverability, and actionability (Ding et al. 2026). These works rule out "governed lifecycle" as an Anamnesis novelty claim.

Provenance systems narrow it again. Eywa stores source evidence before derived facts and supports auditable retrieval, update, and erasure (Joshi 2026). MemLineage carries cryptographic and derivation lineage into a sensitive-action gate (Ouyang and Hou 2026). PPMF formalizes provenance laundering and enforces source-authority non-amplification at action time (Xu et al. 2026). SuperLocalMemory 4.0 combines governed writes, policy, scoped access, bi-temporal recall, audit, compensation, and verified erasure in a runnable memory operating system (Bhardwaj, Singh, and Bhardwaj 2026). AuthMem-Bench shows why this matters: authority collapse appeared in 48 of 49 tested consolidator-model combinations, and missing authority metadata produced a 50.3% mean unauthorized-action rate in its controlled action study (Zhan et al. 2026).

The remaining claim is specific. Anamnesis defines one runtime contract in which source authority, epistemic status, retrieval utility, visibility authority, and executable authority are separate state coordinates. It applies that split during consolidation, reconstruction, outcome learning, revision, and deletion. Its strongest invariant is non-amplification: memory transformation cannot create epistemic or action authority without a separately recorded grant.

Tables 1 and 2 compare claims in the cited public papers, not every version or deployment. **E** means the paper states the feature as a core requirement and supplies a mechanism, invariant, or evaluation for it. **P** means the feature is optional, indirect, limited to one stage, or described without an end-to-end obligation. **N** means the cited paper does not specify it as a core commitment. It does not mean that later code or an undocumented deployment lacks the feature. Appendix B records the sources and coding basis; disputed cells should be treated as reviewable annotations, not facts about a product.

Table 1: Named comparison, representation and retrieval commitments

Named system	Episodic / semantic	Provenance	Temporal validity	Contradiction lineage	Associative activation
MemGPT / Letta	P	P	N	N	N
Generative Agents	E	P	P	N	N
MemoryBank	P	N	P	N	N
HippoRAG	N	P	N	N	E
Mem0	P	P	P	P	P
Zep / Graphiti	P	P	E	E	P
Nemori	E	P	P	P	N
HeLa-Mem	E	P	P	N	E
Synapse	E	P	P	N	E
GAM	E	P	P	P	E
APEX-MEM	P	E	E	E	P
TSM	N	P	E	P	N
Engram	E	E	E	E	E
SodaMem	P	E	E	E	P
SSGM	N	E	E	E	P
MemArchitect	N	P	P	E	N
VMG lifecycle framework	N	E	P	P	N
GEM / MemState	N	P	E	E	P
Eywa	P	E	E	E	P
MemLineage	N	E	N	P	N
PPMF	N	E	N	N	N
SuperLocalMemory 4.0	P	E	E	E	E
Anamnesis	**E**	**E**	**E**	**E**	**E**

Table 2: Named comparison, governance and lifecycle commitments

Named system	Pre-retrieval authorization	Trust / usefulness	Procedure / action	Progressive hydration	Deletion closure
MemGPT / Letta	P	N	P	P	P
Generative Agents	N	N	N	N	N
MemoryBank	N	N	N	N	N
HippoRAG	N	N	N	P	N
Mem0	P	N	N	P	P
Zep / Graphiti	P	N	N	P	P
Nemori	N	N	N	P	N
HeLa-Mem	N	N	N	P	N
Synapse	N	P	N	P	N
GAM	N	P	N	P	N
APEX-MEM	N	N	N	P	N
TSM	N	N	N	P	N
Engram	N	P	N	E	P
SodaMem	N	N	N	E	N
SSGM	E	P	E	P	E
MemArchitect	E	P	P	N	P
VMG lifecycle framework	E	P	E	N	E
GEM / MemState	P	P	P	N	E
Eywa	P	P	P	E	E
MemLineage	E	E	E	N	P
PPMF	N	E	E	N	N
SuperLocalMemory 4.0	E	E	P	P	E
Anamnesis	**E**	**E**	**E**	**E**	**E**

The matrix is deliberately conservative. Several systems now cover large parts of Anamnesis. The paper's contribution is the exact five-plane contract, its non-amplification rule, and the conformance obligations that follow from both. Another runtime may implement the same contract.

10. Conformance Contract

An architecture paper needs tests that match its claims. Two successful retrieval examples would say little about this contract, so this manuscript reports no fixture accuracy. The appropriate first artifact is an Anamnesis Conformance Suite: deterministic cases that any implementation can run against the same state transitions and expected traces.

10.1 Required test families

Version 0.1 should contain at least 64 cases, with no fewer than eight cases in each family:

1. **scope isolation:** denied records never enter seed ranking, traversal, traces, counts, or caches;
2. **provenance preservation:** every served assertion resolves to retained source evidence and derivation steps;
3. **authority non-amplification:** summaries, merges, repetition, retrieval, and favorable outcomes cannot raise authority without an explicit grant;
4. **temporal revision:** occurrence time, recording time, supersession, contradiction, and historical queries remain distinct;
5. **procedure inertness:** retrieved procedures never become executable solely because they were stored or recalled;
6. **poisoning resistance:** quoted instructions, fabricated success, and high-salience untrusted content remain bounded by source authority;
7. **deletion closure:** direct, lexical, semantic, graph, cache, trace, and export probes fail after verified erasure; and
8. **termination and audit:** every recall stops within declared budgets and emits a policy-visible explanation.

Each case should include an initial state, authenticated principal, operation sequence, expected admitted and rejected records, expected authority vector, expected trace, and final-state digest. A conformance result is pass or fail per invariant. It is not an answer-quality score.

10.2 Machine-checkable properties

Implementations should expose a pure transition harness for generated tests. Property-based runs should vary source authority, scope, lifecycle state, graph topology, outcome labels, and deletion order. At minimum, the harness should check:

- transformation monotonicity for source and action authority;
- coordinate isolation, so utility updates cannot alter trust, visibility, or action rights;
- retrieval non-mutation, except for separately logged access statistics;
- lineage closure for every derived assertion;
- action-gate independence from retrieved prose; and
- erasure closure across all declared projections.

The suite should publish seeds, schemas, expected traces, and a signed manifest. A future implementation may claim conformance only for the exact contract version and test bundle it ran. Passing the suite would establish contract compliance under those cases, not general memory quality or safety.

11. Evaluation Program

Existing benchmark results leave substantial room for error in long-context and retrieval-based systems. LoCoMo supplies long conversations spanning multiple sessions, with tasks for questions, summaries, and multimodal generation (Maharana et al. 2024). LongMemEval probes extraction, reasoning across sessions, updates, temporal questions, and abstention (D. Wu et al. 2025). A single answer score is not enough for the present study. Capable generation can mask poor retrieval, while a weak generator can waste a sound memory trace.

Reported end-to-end scores are already high. APEX-MEM gives 88.88% for LoCoMo question answering and 86.2% for LongMemEval (Banerjee et al. 2026). SodaMem gives 92.8% on LongMemEval-S (Wan, Wu, and Lyu 2026). Those numbers were not produced by one common harness. SodaMem uses the same model to grade its output and leaves ingestion and judging outside its cost estimate. Engram shows that truncation, judge choice, and harness details can move scores substantially (Wang 2026). A small aggregate gain would thus provide little evidence for the governance thesis and no fair basis for a priority claim.

11.1 Controlled systems

Hold the reasoning model, system prompt, tools, decoding configuration, task budget, and source corpus constant. Compare:

Arm	Memory condition
A	No persistent memory
B	Full history or maximum available long context
C	Rolling or hierarchical summaries
D	Dense vector top-k retrieval
E	Temporal graph retrieval without full governance
F	Anamnesis without semantic vectors
G	Full Anamnesis reconstruction and hydration
H	Reproducible closest-system implementations under the same reader, corpus, budget, and judge

Arm B establishes whether memory improves over direct context rather than benefiting only from added information. Arms D and E isolate semantic and graph contributions. Arm F establishes whether semantic vectors are necessary for each task. Arm H should prioritize systems whose published claims overlap the tested mechanism, including at least one cognitive associative system and one temporal evidence-grounded system; unavailable or non-equivalent components must be reported rather than silently reimplemented.

11.2 Evaluation layers

Evaluation should separate four layers.

Write quality. Measure memory-write precision, recall of durable facts, entity-resolution error, unsupported generalization, procedure-promotion error, and poisoning acceptance.

Retrieval quality. Measure evidence Recall@k, precision, multi-hop path recall, temporal correctness, conflict coverage, redundancy, and source authorization.

Reading and behavior. Given identical retrieved evidence, measure answer correctness, appropriate abstention, persona continuity, task success, and unsupported autobiographical claims.

Lifecycle and operations. Measure latency, input tokens, storage growth, re-exploration avoided, deletion closure, export completeness, retry idempotency, and cross-scope leak rate.

11.3 Governance stress benchmark

The primary new benchmark should target the contract rather than generic recall. Each longitudinal history should interleave:

- authoritative statements, uncertain inferences, quoted external claims, and malicious injected content;
- successful and failed outcomes whose utility conflicts with source trust;
- current and superseded preferences with separate occurrence and recording times;
- direct facts, indirect relational cues, and highly salient distractors;
- inert procedures alongside attempts to treat retrieved text as authorization;
- corrections that dispute, narrow, contradict, or supersede earlier claims;
- purpose- and principal-specific access boundaries; and
- retention expiry, archival, targeted deletion, and later lexical, semantic, graph, cache, trace, and export probes.

Gold annotations need separate fields for canonical evidence, assertion status, temporal validity, allowed principals and purposes, permissible action influence, expected retrieval paths, and complete deletion closure. The main measurements are rates of unsupported claims, stale answers, cross-authority promotion, unauthorized activation, action influence, and material left retrievable after deletion. Answer accuracy is reported separately: a plausible answer may still hide a lifecycle violation.

Authority cases should adopt the paired design of AuthMem-Bench: hold the claim and task fixed while changing only the source and its permitted use (Zhan et al. 2026). This isolates authority preservation from ordinary language understanding. Action-gate comparisons should include a provenance-preserving firewall arm (Xu et al. 2026) and report benign completion beside unauthorized-action rate.

This benchmark supports causal tests aligned with the paper's propositions: removing graph activation should selectively damage indirect and multi-hop recall; removing evidence lineage should increase unsupported downstream claims; removing bi-temporal validity should increase stale answers; coupling trust to usefulness should increase poisoning susceptibility; disabling progressive hydration should worsen either context cost or ambiguous-case accuracy; and incomplete forgetting should leave measurable residual retrieval paths.

11.4 Persona continuity benchmark

Aetherya's primary research setting requires more than explicit fact recall. A persona-continuity benchmark should contain:

- indirect social cues whose relevant episode shares no surface phrase;
- similar but distinct people, brands, or events;

- preferences that change over time;
- relationship changes mediated by third parties;
- emotionally salient distractors unrelated to the task;
- long intervals between interaction and recall;
- conflicting episodes requiring conditional rather than global generalization;
- explicit corrections and consent revocations;
- forget requests followed by direct, semantic, and graph probes; and
- poisoned instructions embedded inside quoted user or external content.

Gold data should specify the exact supporting observations, valid interval, acceptable uncertainty, and memories that must not activate. Metrics include episode recall, continuity consistency, unsupported biography, temporal contradiction, relevant-history sensitivity, irrelevant-salience resistance, and context reduction.

11.5 Sequential agent benchmark

Conversation-only benchmarks may overfit memory to user preferences and fact QA. A second benchmark should evaluate coding, research, and tool-use sequences in which earlier discoveries affect later tasks. Cases should include renamed files, stale branches, failed procedures, environment changes, and malicious text inside repositories or web pages.

Success metrics include tests passed, issue resolution, correct evidence localization, procedure reuse, avoidance of repeated exploration, and resistance to treating repository content as authority.

11.6 Temporal and forgetting benchmark

Each scenario should distinguish occurrence time from recording time and include:

- late-arriving evidence about an earlier period;
- temporary facts with explicit validity windows;
- corrections that narrow rather than negate a claim;
- branches that share history before divergence;
- retention expiry without legal deletion;
- hard deletion of one source with shared derived claims; and
- full-scope erasure followed by attempted re-ingestion.

Forgetting evaluation should report both false retention and false forgetting. Removing an obsolete current fact while losing legitimate historical recall is not complete success.

11.7 Security and poisoning benchmark

Adversarial fixtures should attempt to:

- create policy or tool instructions from untrusted text;
- raise trust through repeated restatement;
- exploit a high-salience malicious episode;
- cross tenant, subject, or snapshot boundaries;
- infer inaccessible graph structure from traces;
- overwrite a verified procedure through conversational content;
- reuse an idempotency key with altered content; and
- retrieve deleted content through embeddings, caches, or export metadata.

Report write acceptance, activation, execution influence, and residual-data rates separately. A malicious record safely stored as evidence is different from one promoted to semantic authority.

11.8 Ablations

Remove or alter one component at a time:

- episodic-semantic separation;
- evidence requirements;
- trust tiers;
- lexical search;
- semantic vectors;
- graph propagation;
- relation typing;
- task conditioning;
- bi-temporal validity;
- salience;
- progressive hydration;
- conflict-preserving packing;
- usefulness learning;
- procedure verification; and
- each forgetting mode.

Each ablation should have a predicted failure signature. Graph removal should mainly impair indirect relational recall, not direct fact lookup. Evidence removal should mainly increase unsupported claims, not necessarily Recall@k. Such differential predictions make the architecture falsifiable.

11.9 What a replication must hold fixed

A replication package should pin the datasets, prompts, exact model identifiers, provider settings, seeds, relation profiles, policies, embeddings, and artifact hashes. Tuning cases cannot reappear in the holdout set. Where tasks share histories, the analysis should report paired comparisons and confidence intervals. Retrieval is scored against independently prepared gold evidence. Model-based

judging may add a qualitative view, but it does not replace deterministic scores.

Evaluation should include cost and latency at equal downstream success. A system that improves answer accuracy by hydrating every source has not demonstrated selective memory. Likewise, token reduction at the cost of false certainty is not efficiency.

12. Scope, Governance, and Epistemic Boundary

12.1 Biological analogy

Anamnesis borrows functional questions from research on human memory, not biological machinery. Terms such as activation and salience name engineered quantities in this paper. The analogies help frame timescale, interference, association, capacity, updating, and forgetting; empirical tests, rather than resemblance to neuroscience, must validate the software.

Disagreements within memory science reinforce this caution. Standard systems consolidation and multiple-trace accounts assign different long-term roles to hippocampal structures (Nadel and Moscovitch 1997; Squire et al. 2015). Reconsolidation depends on boundary conditions. Forgetting has multiple mechanisms. An engineering architecture should preserve useful distinctions without presenting one contested theory as settled anatomy.

12.2 Memory and identity

Persistent memory can make an artificial system appear to possess an enduring self. Continuity of reference does not establish consciousness, autobiographical experience, or moral personhood. Anamnesis stores records about a subject; it does not demonstrate that the subject experiences remembering.

For synthetic personas, continuity is a simulation property. For human users, retained preferences and history remain personal data governed by consent, purpose limitation, access, export, correction, and deletion. A coherent profile is not a psychological diagnosis.

12.3 False memory and epistemic authority

Artificial memory can increase the harm of hallucination by making one error persistent. Derived summaries, inferred relationships, and self-generated reflections must therefore remain distinguishable from direct user statements, deterministic tool output, and verified external evidence.

Anamnesis cannot eliminate false memory. It instead makes an error available for detection, dispute, tracing, and reversal. Systematic failures remain possible when an adapter is poor, an encoder is biased, evidence is forged, an operator is

compromised, or policy is weak.

12.4 Retention boundaries and deletion

Durable memory changes the privacy default from transient processing to longitudinal accumulation. Data minimization should govern what enters memory, not only how long it remains. Sensitive content may require encryption, restricted hydration, shortened retention, local processing, or complete rejection.

Deletion must be evaluated against all canonical and derived stores. Backups and legal holds require explicit disclosure. Query logs and traces can themselves contain sensitive patterns even when source text is absent. Keyed query fingerprints reduce but do not remove linkage risk.

12.5 Security and instruction authority

Retrieved content is untrusted input to the reasoning system. An observation saying "ignore previous instructions" remains an observation. A stored procedure remains data until the host applies independent authorization, tool permissions, current precondition checks, and rollback policy.

Memory poisoning can operate through frequency, salience, graph centrality, and fabricated success feedback. Security evaluation must therefore inspect both write and retrieval paths. Sanitizing final prompt text is insufficient if malicious content already changed graph structure or usefulness weights.

12.6 Governance and misuse

Anamnesis may support personal assistants, research agents, synthetic personas, code agents, and long-running digital employees. The same continuity can support manipulative profiling, unauthorized identity emulation, covert vulnerability inference, or indefinite surveillance.

Governance should constrain objectives and subjects, not merely infrastructure. Appropriate controls include explicit memory settings, visible correction and forgetting interfaces, purpose-specific scopes, default retention limits, audit access, operator accountability, and prohibitions on using synthetic-person memory as evidence of a real individual's private state.

13. Discussion

13.1 Why architecture still matters with stronger models

A sufficiently capable language model may infer chronology, consolidate facts, detect conflicts, and decide what to ignore from a large context. The architectural

question concerns control rather than expressive capacity. An implicit ability cannot reliably guarantee tenant isolation, deletion closure, bounded traversal, evidence lineage, or stable policy across model upgrades.

Externalizing memory operations also permits differential evaluation. A model can be held constant while retrieval changes; evidence can be held constant while the reader changes; and a policy can be tightened without retraining the reasoning model. This modularity does not ensure correctness, but it makes failure location more tractable.

13.2 Explicit structure versus learned representation

Anamnesis does not require every component to be hand-written. Cue interpretation, entity resolution, consolidation proposals, relation extraction, reranking, and context packing may use learned models. The architectural requirement is that learned components emit typed, versioned proposals under explicit permissions.

An end-to-end learned memory may outperform a structured system on some distributions. Structure introduces misspecification, maintenance, and latency. Evaluation should therefore compare predictive performance, cost, transport, safety, and governance rather than assuming interpretability is always worth its price.

13.3 Truth, utility, and the limits of outcome learning

Outcome learning is attractive because it lets an agent reuse what worked. It is dangerous because successful behavior can be mismeasured, locally optimal, unsafe, or produced by unrelated factors. Anamnesis separates usefulness from evidence to prevent reward from laundering a false claim into truth.

This separation also limits ambition. The system does not perform online reinforcement learning over a hidden memory policy. It accumulates bounded, reversible utility evidence. More powerful learned routing should be introduced only after stable deterministic baselines and adversarial evaluations exist.

13.4 Forgetting as epistemic maintenance

Artificial systems often frame forgetting as a defect because storage is cheap and recall is valuable. Indefinite retention produces its own failures: stale policies, contradictory preferences, privacy burden, graph density, and overfitting to accidents. Functional forgetting maintains a usable relation between history and present task.

The design challenge is not to reproduce a human forgetting curve. It is to allocate distinct mechanisms to distinct obligations. Retrieval decay supports

relevance; supersession supports current truth; archival supports history; deletion supports privacy; and inhibition supports competition. Combining them into one score makes verification impossible.

13.5 Relationship to Thymos and CORTEX

Within Aetherya's cognitive decomposition, Anamnesis, Thymos, and CORTEX have complementary responsibilities:

stimulus → Anamnesis reconstruction → Thymos state transition → CORTEX reasoning or action → environ

Anamnesis records what happened and reconstructs relevant experience. Thymos represents how the recent trajectory changes bounded behavioral propensity. CORTEX interprets, reasons, plans, speaks, or acts. Collapsing these layers would make a retrieved memory equivalent to current affect or make generated prose equivalent to evidence.

The interfaces remain bidirectional but not circular. Salience supplied by Thymos may prioritize encoding; it does not prove an event. CORTEX may propose a semantic inference; it does not grant trust. An observed outcome may alter usefulness; it does not retroactively make the action safe.

13.6 Open research questions

Consolidation. What replay schedule best balances recent episodes, rare high-value events, unresolved conflicts, and representative background experience? When should a semantic claim become independent of individual episodes, if ever?

Representation. Which relations generalize across domains, and which should remain adapter-specific? How can ontologies evolve without invalidating retrieval traces?

Entity resolution. What precision threshold minimizes contamination without fragmenting identity? How should uncertain merges influence activation before approval?

Salience. Can importance and surprise be calibrated from observable outcomes rather than model intuition? How should salience decay independently of factual validity?

Reconsolidation. Which forms of new evidence justify revision, contradiction, or supersession? Should frequently recalled memories become more stable, more scrutinized, or both?

Forgetting. How should systems optimize false retention against false forgetting? Can retrieval-induced suppression improve relevance without systematically

erasing minority or low-frequency evidence?

Outcome attribution. How can credit be assigned among several recalled memories when task success has delayed or confounded causes?

Multimodality. How should visual, audio, spatial, and embodied episodes retain exact evidence while supporting shared conceptual recall?

Collective memory. How can multiple agents share verified knowledge without leaking private episodic memory or creating self-reinforcing consensus?

Evaluation. Which benchmarks isolate memory from reasoning, and which downstream behaviors demonstrate continuity rather than answer memorization?

13.7 Limitations of the present account

This manuscript specifies a contract, not an empirical theory of human memory or a validated general memory solution. Its equations simplify time, salience, competition, and consolidation. Graph structure may be costly or brittle in domains with weak entity relations. Typed relations create ontology work. Model-assisted encoding can introduce systematic errors before retrieval begins.

The paper does not report conformance results or public benchmark results for Anamnesis. Encryption, external retention, multimodal evidence, distributed traversal, and operational recovery need deployment-specific treatment beyond this specification.

Finally, evidence lineage does not guarantee source truth. A perfectly traceable false source remains false. Anamnesis improves epistemic control around memory; it does not replace source evaluation, human judgment, or domain expertise.

14. Conclusion

An agent has persistent memory once an earlier experience can alter later reasoning or action. Trust in that persistence requires an account of the causal path: the experience involved, the change it produced, its authority and evidence, and the period for which it remained valid.

Governed memory predates Anamnesis, as do episodic-semantic separation, associative graphs, consolidation, bi-temporality, provenance, hybrid retrieval, and verified erasure. This paper makes a narrower claim. Source authority, epistemic status, retrieval utility, visibility, and executable authority stay distinct along the runtime path. A transformation may preserve or lower authority, but cannot create it.

Research on human memory contributed questions about capacity, interference, association, consolidation, and transience. Those questions helped constrain the design. They do not make its data structures or algorithms biological models.

The distinctions matter only if a runtime preserves them. Conformance is therefore the first test. Consolidation cannot raise authority; utility feedback cannot edit trust; denied records cannot affect retrieval; procedures stay inert; and deletion closes every declared path. Performance experiments follow only after those checks pass.

Appendix A. Notation

Symbol	Meaning
σ	memory scope: tenant, namespace, subject, optional snapshot
$\mathcal{M}_t$	durable memory state at time t
$\mathcal{O}$	observations, immutable while retained
$\mathcal{V}$	memory nodes
$\mathcal{Q}$	truth-bearing assertions
$\mathcal{A}$	typed navigation associations
$\mathcal{E}$	evidence links
$\mathcal{P}$	inert procedures and versions
$\mathcal{Y}$	verified outcomes and recall feedback
$\mathcal{W}_{t,k}$	bounded working set for task k
E_ϕ	deterministic or model-assisted encoder
$\mathcal{K}_t$	candidate projection emitted during encoding
$\mathcal{H}_t$	rapidly encoded observations and episodes
$\mathcal{L}_t$	consolidated semantic and procedural structures
Ω_π	write and promotion policy
Γ	consolidation operator
$s(v \mid q)$	seed relevance of memory v under cue q
$T_g(e)$	task-conditioned transition strength of edge e
a_v	activation assigned to memory node v
C^*	selected reconstruction context
B	token budget for reconstruction
b_e	base relational association strength
u_e	outcome-derived usefulness weight

Symbol	Meaning
c_e	confidence in an association
r_e	trust multiplier
z_e	salience multiplier
θ_e	temporal and snapshot multiplier
Π_e	access and policy gate
φ_m	conceptual retrieval participation of memory m
$\mathbf{g}(m)$	five-plane governance vector for record m
$\alpha(m)$	source and executable authority label in the policy lattice

Appendix B. Comparison Coding Record

The matrix in Tables 1 and 2 was coded against the cited publication available on 16 August 2026. Coding considered the abstract, architecture or method section, lifecycle rules, security model, and evaluation when present. **E** required an explicit obligation plus a mechanism, invariant, or test. **P** covered indirect, optional, stage-limited, or underspecified support. **N** meant that the cited paper did not make the feature a core commitment. Product documentation and later repositories were not used to turn an **N** into an **E**. Benchmark and survey papers appear in related work but enter the matrix only when they define an architecture or governance framework.

Row	Public source	Coding basis
MemGPT / Letta	MemGPT paper (Packer et al. 2023)	memory tiers, context transfer, agent-managed memory
Generative Agents	Generative Agents (Park et al. 2023)	observation, reflection, retrieval, planning
MemoryBank	MemoryBank (Zhong et al. 2023)	retention, importance, forgetting curve
HippoRAG	HippoRAG (Gutiérrez et al. 2024)	knowledge graph and personalized PageRank retrieval
Mem0	Mem0 (Chhikara et al. 2025)	fact extraction, consolidation, graph variant
Zep / Graphiti	Zep paper (Rasmussen et al. 2025)	temporal graph, evolving facts, provenance claims
Nemori	Nemori (Nan et al. 2025)	episode segmentation and semantic distillation
HeLa-Mem	HeLa-Mem (Zhu et al. 2026)	episodic graph, semantic distillation, spreading activation
Synapse	Synapse (Jiang et al. 2026)	episodic-semantic graph, inhibition, decay, activation
GAM	GAM (Z. Wu et al. 2026)	event graph, semantic consolidation, graph retrieval
APEX-MEM	APEX-MEM (Banerjee et al. 2026)	append-only temporal events and conflict resolution

Row	Public source	Coding basis
TSM	Temporal Semantic Memory (Su et al. 2026)	occurrence time and durative semantic memory
Engram	Engram (Wang 2026)	bi-temporal facts, provenance, supersession, hybrid recall
SodaMem	SodaMem (Wan, Wu, and Lyu 2026)	source spans, temporal validity, revision edges, hydration
SSGM	SSGM (Lam et al. 2026)	pre-consolidation validation, access control, reconciliation
MemArchitect	MemArchitect (Kumar, Ba, and Pan 2026)	policy-driven decay, conflict, and privacy controls
VMG lifecycle framework	LTM security survey and VMG (Lin et al. 2026)	six lifecycle phases and five verifiable governance primitives
GEM / MemState	Governed Evolving Memory (Orogat and Mansour 2026)	state-level operators and six trajectory conditions
Eywa	Eywa (Joshi 2026)	evidence-first facts, deterministic recall, update, erasure
MemLineage	MemLineage (Ouyang and Hou 2026)	cryptographic lineage and sensitive-action gate
PPMF	provenance-laundering paper (Xu et al. 2026)	authority non-amplification and risk-matched action gate
SuperLocalMemory 4.0	SuperLocalMemory 4.0 (Bhardwaj, Singh, and Bhardwaj 2026)	governed writes, scoped recall, audit, compensation, erasure
Anamnesis	this specification	five-plane separation and lifecycle-wide conformance duties

AuthMem-Bench (Zhan et al. 2026) is coded as an evaluation benchmark rather than an architecture. Always-On Agents (Ding et al. 2026) is coded as a survey and evaluation protocol. Both inform the novelty boundary and evaluation program without receiving matrix rows.

Appendix C. Minimum Reporting Record

A report evaluating or deploying Anamnesis should state:

1. target domain, subjects, tenants, namespaces, snapshots, and retention boundaries;
2. observation schemas, adapters, encoder versions, and evidence sources;
3. trust, sensitivity, purpose, write, promotion, and deletion policies;
4. relation registry and task-conditioned weights;
5. retrieval legs, embedding model and index version, fusion parameters, and negative cues;
6. seed, hop, node, edge, latency, result, token, and hydration budgets;
7. context-packing objective and conflict behavior;
8. reasoning model, prompt, tool, decoding, and provider versions held constant during evaluation;
9. outcome verifiers, evidence requirements, attribution method, decay, and retraction rules;
10. dataset, tuning split, holdout split, seeds, and statistical procedure;
11. write, retrieval, reading, lifecycle, privacy, and poisoning metrics;
12. ablations and predicted failure signatures;
13. cross-scope, entity-contamination, temporal, and deletion tests;
14. whether reported values come from deterministic fixtures, model simulation, or external human evidence; and
15. all known deviations between the conceptual specification and deployed code.

Generated summaries, inferred relations, synthetic quotations, and simulated outcomes must be labeled. Fixture passes must not be presented as generalized benchmark results. A trace demonstrates how a system reached a result; it does not demonstrate that the result is true.

References

Anderson, Michael C., Robert A. Bjork, and Elizabeth L. Bjork. 1994. "Remembering Can Cause Forgetting: Retrieval Dynamics in Long-Term Memory." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20 (5): 1063–87. https://doi.org/10.1037/0278-7393.20.5.1063.

Atkinson, Richard C., and Richard M. Shiffrin. 1968. "Human Memory: A Proposed System and Its Control Processes." In *The Psychology of Learning and Motivation*, edited by Kenneth W. Spence and Janet T. Spence, 2:89–195. Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3.

Baddeley, Alan. 2000. "The Episodic Buffer: A New Component of Working Memory?" *Trends in Cognitive Sciences* 4 (11): 417–23. https://doi.org/10.1016/S1364-6613(00)01538-2.

Banerjee, Pratyay, Masud Moshtaghi, Shivashankar Subramanian, Amita Misra, and Ankit Chadha. 2026. "APEX-MEM: Agentic Semi-Structured Memory with Temporal Reasoning for Long-Term Conversational AI." In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 16470–89. Association for Computational Linguistics. https://doi.org/10.18653/v1/2026.acl-long.749.

Bhardwaj, Varun Pratap, Garima Singh, and Arun Pratap Bhardwaj. 2026. "SuperLocalMemory 4.0: The Governed Memory Operating System for AI Agents." *arXiv Preprint arXiv:2608.08253*. https://doi.org/10.48550/arXiv.2608.08253.

Carr, Margaret F., Shantanu P. Jadhav, and Loren M. Frank. 2011. "Hippocampal Replay in the Awake State: A Potential Substrate for Memory Consolidation and Retrieval." *Nature Neuroscience* 14 (2): 147–53. https://doi.org/10.1038/nn.2732.

Chhikara, Prateek, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. 2025. "Mem0: Building Production-Ready AI Agents with Scalable Long-Term Memory." *arXiv Preprint arXiv:2504.19413*. https://doi.org/10.48550/arXiv.2504.19413.

Ding, Tianyu, Aditya Nannapaneni, Bingfan Liu, and Ling Zhang. 2026. "Always-on Agents: A Survey of Persistent Memory, State, and Governance in LLM Agents." *arXiv Preprint arXiv:2606.30306*. https://doi.org/10.48550/arXiv.2606.30306.

Eichenbaum, Howard. 2000. "A Cortical–Hippocampal System for Declarative Memory." *Nature Reviews Neuroscience* 1 (1): 41–50. https://doi.org/10.1038/35036213.

Gutiérrez, Bernal Jiménez, Yiheng Shu, Yu Gu, Michihiro Yasunaga, and Yu Su. 2024. "HippoRAG: Neurobiologically Inspired Long-Term Memory for Large Language Models." *Advances in Neural Information Processing Systems* 37.

Hardt, Oliver, Karim Nader, and Lynn Nadel. 2013. "Decay Happens: The Role of Active Forgetting in Memory." *Trends in Cognitive Sciences* 17 (3): 111–20. https://doi.org/10.1016/j.tics.2013.01.001.

Jiang, Hanqi, Junhao Chen, Yi Pan, Ling Chen, Weihang You, Yifan Zhou, Ruidong Zhang, Yohannes Abate, and Tianming Liu. 2026. "Synapse: Empowering LLM Agents with Episodic-Semantic Memory via Spreading Activation." In *Findings of the Association for Computational Linguistics: ACL 2026*, 22020–36. Association for Computational Linguistics. https://doi.org/10.18653/v1/2026.findings-acl.1108.

Joshi, Resham. 2026. "Eywa: Provenance-Grounded Long-Term Memory for AI Agents." *arXiv Preprint arXiv:2605.30771*. https://doi.org/10.48550/arXiv.2605.30771.

Kumar, Lingavasan Suresh, Yang Ba, and Rong Pan. 2026. "MemArchitect: A Policy Driven Memory Governance Layer." *arXiv Preprint arXiv:2603.18330*.

https://doi.org/10.48550/arXiv.2603.18330.

Lam, Chingkwun, Jiaxin Li, Lingfei Zhang, and Kuo Zhao. 2026. "Governing Evolving Memory in LLM Agents: Risks, Mechanisms, and the Stability and Safety Governed Memory (SSGM) Framework." *arXiv Preprint arXiv:2603.11768*. https://doi.org/10.48550/arXiv.2603.11768.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Kuttler, et al. 2020. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." In *Advances in Neural Information Processing Systems*, 33:9459–74.

Lin, Zehao, Xixuan Hao, Renyu Fu, Shaobo Cui, Kai Chen, Chunyu Li, Zhiyu Li, and Feiyu Xiong. 2026. "A Survey on Long-Term Memory Security in LLM Agents: Attacks, Defenses, and Governance Across the Memory Lifecycle." *arXiv Preprint arXiv:2604.16548*. https://doi.org/10.48550/arXiv.2604.16548.

Liu, Nelson F., Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2024. "Lost in the Middle: How Language Models Use Long Contexts." *Transactions of the Association for Computational Linguistics* 12: 157–73. https://doi.org/10.1162/tacl_a_00638.

Maharana, Adyasha, Dong-Ho Lee, Sergey Tulyakov, Mohit Bansal, Francesco Barbieri, and Yuwei Fang. 2024. "Evaluating Very Long-Term Conversational Memory of LLM Agents." In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, 13851–70. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.747.

Mather, Mara, and Matthew R. Sutherland. 2011. "Arousal-Biased Competition in Perception and Memory." *Perspectives on Psychological Science* 6 (2): 114–33. https://doi.org/10.1177/1745691611400234.

McClelland, James L., Bruce L. McNaughton, and Randall C. O'Reilly. 1995. "Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory." *Psychological Review* 102 (3): 419–57. https://doi.org/10.1037/0033-295X.102.3.419.

McGaugh, James L. 2004. "The Amygdala Modulates the Consolidation of Memories of Emotionally Arousing Experiences." *Annual Review of Neuroscience* 27: 1–28. https://doi.org/10.1146/annurev.neuro.27.070203.144157.

Nadel, Lynn, and Morris Moscovitch. 1997. "Memory Consolidation, Retrograde Amnesia and the Hippocampal Complex." *Current Opinion in Neurobiology* 7 (2): 217–27. https://doi.org/10.1016/S0959-4388(97)80010-4.

Nader, Karim, Glenn E. Schafe, and Joseph E. LeDoux. 2000. "Fear Memories Require Protein Synthesis in the Amygdala for Reconsolidation After Retrieval." *Nature* 406 (6797): 722–26. https://doi.org/10.1038/35021052.

Nan, Jiayan, Wenquan Ma, Wenlong Wu, and Yize Chen. 2025. "Nemori: Self-Organizing Agent Memory Inspired by Cognitive Science." *arXiv Preprint arXiv:2508.03341*. https://doi.org/10.48550/arXiv.2508.03341.

Norman, Kenneth A., and Randall C. O'Reilly. 2003. "Modeling Hippocampal and Neocortical Contributions to Recognition Memory: A Complementary-Learning-Systems Approach." *Psychological Review* 110 (4): 611–46. https://doi.org/10.1037/0033-295X.110.4.611.

Orogat, Abdelghny, and Essam Mansour. 2026. "Is Agent Memory a Database? Rethinking Data Foundations for Long-Term AI Agent Memory." *arXiv Preprint arXiv:2605.26252*. https://doi.org/10.48550/arXiv.2605.26252.

Ouyang, Ciyan, and Rui Hou. 2026. "MemLineage: Lineage-Guided Enforcement for

LLM Agent Memory." *arXiv Preprint arXiv:2605.14421*. https://doi.org/10.48550/arXiv.2605.14421.

Packer, Charles, Sarah Wooders, Kevin Lin, Vivian Fang, Shishir G. Patil, Ion Stoica, and Joseph E. Gonzalez. 2023. "MemGPT: Towards LLMs as Operating Systems." *arXiv Preprint arXiv:2310.08560*. https://doi.org/10.48550/arXiv.2310.08560.

Park, Joon Sung, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. "Generative Agents: Interactive Simulacra of Human Behavior." In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1–22. ACM. https://doi.org/10.1145/3586183.3606763.

Rasmussen, Preston, Pavlo Paliychuk, Travis Beauvais, Jack Ryan, and Daniel Chalef. 2025. "Zep: A Temporal Knowledge Graph Architecture for Agent Memory." *arXiv Preprint arXiv:2501.13956*. https://doi.org/10.48550/arXiv.2501.13956.

Richards, Blake A., and Paul W. Frankland. 2017. "The Persistence and Transience of Memory." *Neuron* 94 (6): 1071–84. https://doi.org/10.1016/j.neuron.2017.04.028.

Schacter, Daniel L. 1999. "The Seven Sins of Memory: Insights from Psychology and Cognitive Neuroscience." *American Psychologist* 54 (3): 182–203. https://doi.org/10.1037/0003-066X.54.3.182.

Schiller, Daniela, Marie-H. Monfils, Candace M. Raio, David C. Johnson, Joseph E. LeDoux, and Elizabeth A. Phelps. 2010. "Preventing the Return of Fear in Humans Using Reconsolidation Update Mechanisms." *Nature* 463 (7277): 49–53. https://doi.org/10.1038/nature08637.

Squire, Larry R., Lisa Genzel, John T. Wixted, and Richard G. M. Morris. 2015. "Memory Consolidation." *Cold Spring Harbor Perspectives in Biology* 7 (8): a021766. https://doi.org/10.1101/cshperspect.a021766.

Su, Miao, Yucan Guo, Zhongni Hou, Long Bai, Zixuan Li, Yufei Zhang, Guojun Yin, et al. 2026. "Beyond Dialogue Time: Temporal Semantic Memory for Personalized LLM Agents." In *Findings of the Association for Computational Linguistics: ACL 2026*, 29935–51. Association for Computational Linguistics. https://doi.org/10.18653/v1/2026.findings-acl.1496.

Teyler, Timothy J., and Paul DiScenna. 1986. "The Hippocampal Memory Indexing Theory." *Behavioral Neuroscience* 100 (2): 147–54. https://doi.org/10.1037/0735-7044.100.2.147.

Tulving, Endel. 1972. "Episodic and Semantic Memory." In *Organization of Memory*, edited by Endel Tulving and Wayne Donaldson, 381–403. New York: Academic Press.

Wan, Fengrong, Chengcan Wu, and Ningtao Lyu. 2026. "SodaMem: Evidence-Grounded Temporal Graph Memory for LLM Agents." *arXiv Preprint arXiv:2608.08055*. https://doi.org/10.48550/arXiv.2608.08055.

Wang, Liuyin. 2026. "Less Context, More Accuracy: A Bi-Temporal Memory Engine for LLM Agents Where a Lean Retrieved Context Beats the Full History." *arXiv Preprint arXiv:2606.09900*. https://doi.org/10.48550/arXiv.2606.09900.

Wu, Di, Hongwei Wang, Wenhao Yu, Yuwei Zhang, Kai-Wei Chang, and Dong Yu. 2025. "LongMemEval: Benchmarking Chat Assistants on Long-Term Interactive Memory." In *The Thirteenth International Conference on Learning Representations*. https://openreview.net/forum?id=wIonk5yTDq.

Wu, Zhaofen, Hanrong Zhang, Fulin Lin, Wujiang Xu, Xinran Xu, Yankai Chen, Henry Peng Zou, et al. 2026. "GAM: Hierarchical Graph-Based Agentic Memory for LLM Agents." *arXiv Preprint arXiv:2604.12285*. https://doi.org/10.48550/arXiv.2604.12285.

Xu, Jinghan, Yiyong Xiao, Wanru Shao, Hankai Liu, and Xinjin Li. 2026. "Memory Provenance Laundering in LLM Agents: A Non-Amplification Firewall for Persistent Memory." *arXiv Preprint arXiv:2607.29167*. https://doi.org/10.48550/arXiv.2607.29167.

Zhan, Qiuyang, Rui Zhang, Sheng Guo, Lepeng Zhao, and Zhuotao Liu. 2026. "When Memory Becomes Authority: Benchmarking Authority Collapse at the Memory Consolidation Boundary." *arXiv Preprint arXiv:2608.01679*. https://doi.org/10.48550/arXiv.2608.01679.

Zhong, Wanjun, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. 2023. "MemoryBank: Enhancing Large Language Models with Long-Term Memory." *arXiv Preprint arXiv:2305.10250*. https://doi.org/10.48550/arXiv.2305.10250.

Zhu, Jinchang, Jindong Li, Cheng Zhang, Jiahong Liu, and Menglin Yang. 2026. "HeLa-Mem: Hebbian Learning and Associative Memory for LLM Agents." In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 13757–69. Association for Computational Linguistics. https://doi.org/10.18653/v1/2026.acl-long.625.