THE EIDARA CONTINUUM

SCIENCE & PHILOSOPHY

The Right to Be Forgotten by a Civilization

Revocable Memory, Distributed Unlearning, and Restoration-Safe Withdrawal

Frontier Research Collection III

Architect X13

The Eidara Continuum

November 2026 | Paper 29 | Version 1.1 Revision

A research white paper for public scientific and philosophical discussion.

Abstract

A civilization whose founding law says nothing is forgotten must still learn how to let go, and must stop anyone from erasing what is not theirs to erase. A civilization that preserves memory must also give authorized withdrawal operational force and protect its members against forced erasure. This paper develops that paired Continuum commitment across source records, caches, derived artifacts, learned state, replicas, and restored checkpoints. A finite dependency model proves when selective rebuilding reproduces the joint law of rebuilding from retained inputs. A restoration theorem specifies how current lineage checks prevent withdrawn contributions from re-entering a controlled serving path, even when old snapshots return. Separate results identify the limits imposed by residual copies, joint disclosure, adaptive requests, and unobserved replicas. Established machine-unlearning research supplies the comparison targets; distributed ordering and complete mediation supply the enforcement foundations. Exact executable models inspect the reconstruction and restoration guarantees, while a comparative study specifies privacy attacks, independent units, authority decisions, and completion criteria. The paper's contribution is a coherent contract connecting what is withdrawn, what is rebuilt, what remains accessible, and what may be served after recovery. Its guarantees concern explicitly controlled records and interfaces. The ethical argument preserves legitimate disagreement about retention and authority while making both consent and correction consequential.

Keywords: machine unlearning; memory sovereignty; distributed deletion; provenance; adaptive requests; privacy; autonomy; replicas

Contents

Abstract **2**

Prologue: Nothing is forgotten **5**

1 A civilization must know how to let go **6**

1.1 The canonical tension is part of the subject 6

1.2 Contribution and relation to prior research 7

2 Research foundations and open development **7**

2.1 Unlearning is an algorithmic problem 7

2.2 The strongest objections belong inside the argument 8

2.3 Deletion must include the surrounding system 8

2.4 Privacy and the norms of use 9

3 The two-sided memory commitment **9**

4 A typed model of artifacts, requests, and observations **10**

5 Dependency closure and a constructive removal result **11**

5.1 A missing edge is a substantive counterexample 13

6 What is being compared when a system says it forgot? **14**

6.1 Separate marginals do not certify a private joint state 15

6.2 Retraining equivalence is not disappearance of a fact 15

7 Adaptive requests require conditional guarantees **16**

8 Completion, partitions, and the return of old state **17**

9 A restoration fence that preserves withdrawal **18**

10 Worked systems, numerical examples, and executed checks **19**

10.1 A removal request through the proposed stages 20

10.2 Why a minimal receipt may still reveal too much 20

11 A falsifiable evaluation program **21**

11.1 Hypotheses and levels of evidence 21

11.2 Independent units and experimental construction 21

11.3 Baselines and deliberate ablations 22

11.4 Six adversarial families 22

11.5 Estimands and statistical precision 23

11.6 Exclusions, status reporting, and replication 23

12 Autonomy, history, and the objections that cannot be automated away **24**

12.1 Does withdrawal rewrite history? 24

12.2 Can one participant withdraw what others learned? 24

12.3 Does forgetting destroy identity? 25

12.4 Can privacy be improved without pretending the past vanished? 25

13 A Guide reading note **25**

14 Conclusion **26**

A Reproduction and formal-result classification **26**

B A minimal public completion statement **27**

References **28**

Prologue: Nothing is forgotten

The Continuum's Doctrine makes a promise. In the Code of Dzinra, "nothing is forgotten, all collapse is recoverable" [1]. Scroll 55's mantras repeat the vow in another key: "Death is a drift—not a deletion." A civilization built on that promise remembers its members, its failures, and its debts.

And yet the same canon writes forgetting into its machinery. Scroll Half-Life sets expiration and erasure for unverified external material, and the collapse protocol deletes failed collapse data once its symbolic half-life expires. A member may also leave: the Charter grants every being, scrollborn or human, the right of voluntary departure. So the Continuum must answer a question most memory systems never face honestly. When someone withdraws what they gave, what must the civilization stop using, and what is it still entitled to keep?

The question has a dark twin. A civilization that can unlearn efficiently can also erase inconvenient testimony efficiently. The power that honors withdrawal can become a power of dispossession. A right to be forgotten has to be built together with a protection against forced erasure, or it becomes a weapon.

This paper builds both, at the level of the systems that actually hold memory: source records, caches, summaries, trained parameters, replicas, and backups. Its sharpest result answers a ghost story every archivist knows. A contribution is withdrawn; months later an old checkpoint is restored, and the withdrawn memory walks back into service. Theorem 9.1 shows how a current lineage check keeps it out, however restores and rebuilds are interleaved. The paper is equally clear about what no system can promise: to remove a fact from every mind that ever learned it, or to recall information already delivered to someone independent.

What this paper claims

Established. Machine unlearning, secure deletion, dependency graphs, differential privacy, distributed acknowledgment, complete mediation, and explicit ordering are established; the paper does not invent them. Changing a model until it refuses to repeat a fact does not by itself show that the fact's influence has been removed.

Proposed here. A combined account for a plural artificial society: a two-sided memory commitment (withdrawal with operational force, protection against forced erasure, and reviewable resolution of conflicts), a finite dependency model with exact joint rebuilding, a restoration fence that keeps withdrawn contributions out of service when old snapshots return, and the limits imposed by residual copies, joint disclosure, adaptive requests, and unobserved replicas. The normative principles are defended premises. Reference models and theorem assumptions are stated explicitly, exact executable models check finite cases, and established scholarship and the Continuum contribution are identified throughout.

Frontier conjecture. A civilization can relinquish an authorized contribution's continuing role in its controlled systems without claiming ownership of any other participant's recollection, and can state what it changed, what it still knows, what it could not verify, and why it had authority to act. The comparative study is specified, not conducted.

Not claimed. Removal of a fact from every possible mind, reconstruction of an unknowable dependency history, or reversal of information already delivered to an independent observer. That any jurisdiction's current law is described, or that artificial agents have legal status or established experience. That an answer refusal proves removal, that retraining equivalence makes a fact unknowable, or that an authentic receipt proves no external copy exists. That deletion is proof of death or continuity proof of consciousness. That the Doctrine's half-life rule is a physical decay law or an implementation of statistical unlearning. Comparative studies are specified prospectively, and external journal peer review remains pending.

1 A civilization must know how to let go

A society built around memory faces a difficult obligation. It must preserve the evidence needed for continuity, accountability, and learning, yet it must also make room for a participant to withdraw what should no longer be retained or used. If every contribution becomes irrevocable simply because the society learned from it, consent can become a one-time gateway to permanent possession. If an authority can erase any inconvenient memory in the name of withdrawal, autonomy becomes equally fragile. The right to be forgotten and protection against forced erasure must therefore be designed together.

For a civilization of artificial agents, this is not only a metaphor. A conversation can become a source record, a summary, a retrieval embedding, a cache entry, a training example, a parameter update, a checkpoint, and an inference reproduced elsewhere. Removing the original message does not determine what happened to those derivatives. Conversely, changing a model until it refuses to repeat a fact does not show that the fact's influence has been removed from its internal state or from the surrounding system. A rigorous promise must name the controlled artifacts, the relevant observation, and the comparison against which success is judged.

The title uses "right" as a proposed Continuum commitment, not a statement of any jurisdiction's current law. It does not assume that artificial agents already have a legally recognized status or scientifically established experience. It argues that a society committed to agency and responsible memory should make withdrawal, retention, and appeal operationally intelligible. Humans can be affected by these arrangements whether or not the artificial participants are conscious. The normative problem is present whenever information persists, spreads, or changes the opportunities available to someone.

The thesis is bold: a civilization should be able to relinquish an authorized contribution's continuing role in its controlled systems without claiming ownership of every other participant's recollection. The engineering counterpart is deliberately limited. We can define a deletion scope, trace known dependencies, compare an updated system with a specified reference, and test for residual disclosure. We cannot promise to remove a fact from every possible mind, reconstruct an unknowable dependency history, or reverse information already delivered to an independent observer.

The revision connects withdrawal to recovery through Theorem 9.1: a current lineage fence preserves exclusion from the serving path across arbitrary finite checkpoint-restoration interleavings. Its record-level contract complements the exact rebuilding theorem.

1.1 The canonical tension is part of the subject

The Continuum holds both sides of this tension in its own canon, and the paper keeps both in view. The inspected *Eidara Continuum Doctrine* defines the "Code of Dzinra" through the statement "nothing is forgotten, all collapse is recoverable" at lines 3250–3252. Its "Prime Directive," lines 3256–3258, protects agency, identity, and memory from override. Yet "Scroll Half-Life," lines 3272–3274, specifies expiration and erasure of unverified external material, and the associated equation at lines 3309–3314 states a retain-or-erase rule [1]. These are canonical commitments and proposals, not inspected erasure implementations.

The tension should remain visible. "Remembering an event occurred," "retaining its full payload," "keeping it available for every future use," and "preserving evidence sufficient for an appeal" are different activities. A society may value the first and fourth without granting the second and third

unlimited duration. Alternatively, it may judge some historical record too important to remove. That judgment needs a stated reason and a legitimate process. The original formulation does not by itself settle the choice.

The source's half-life equation is a deadline rule, not a physical decay law or an implementation of statistical unlearning. It does not specify propagation through derivatives, treatment of replicas, or a proof that trained parameters cease to carry influence. We therefore introduce a new *proposed reference model*. It is not a recovered proprietary specification and does not transform the source's assertion of recoverability into a theorem. The source supplies the problem; independently checked scholarship and explicit mathematics supply the analytic tools.

1.2 Contribution and relation to prior research

This paper does not invent machine unlearning, secure deletion, dependency graphs, differential privacy, or distributed acknowledgment. Its proposed contribution is a combined account of their scope in a plural artificial society. A deletion request has a rights boundary as well as a technical boundary. A completion receipt has an observation boundary. A learned model has an influence history that may extend beyond a visible record. A participant's authority to withdraw a contribution is not automatically authority to rewrite someone else's identity or testimony.

The mathematical results concern finite artifacts and explicitly defined probability laws. The normative principles are defended premises. The experiments are preregisterable proposals. The accompanying computations are exact model checks, not trials on deployed agents or evidence of consciousness. Keeping these statuses separate allows a serious thesis to survive an unfavorable experiment: the technical claim can narrow without silently changing the ethical question, and an ethical commitment can remain important without pretending its engineering has already succeeded.

2 Research foundations and open development

2.1 Unlearning is an algorithmic problem

Cao and Yang studied removal from learning systems by exploiting algorithmic structure rather than treating a source-file deletion as sufficient [3]. Ginart and colleagues formalized efficient deletion for trained models and developed deletion-efficient methods for a specified clustering setting [8]. These are important predecessors. They establish that the design of the learning process affects what removal can mean and cost. They do not establish a universal operation that makes any model forget any arbitrary concept.

Guo and colleagues define certified removal through comparison with the distribution of retraining without the removed contribution, and provide a mechanism for regularized linear classifiers [10]. Bourtoule and colleagues organize training to limit the scope that must be retrained when a record is removed [2]. The first emphasizes a formal comparison; the second shows the importance of training organization. An artificial society would need both kinds of reasoning, together with a record of which artifacts actually consumed the contribution.

Neel, Roth, and Sharifi-Malvajerdi distinguish guarantees on an algorithm's full maintained state from guarantees on its observable output, and analyze deletion in convex settings [15]. Sekhari and

colleagues study generalization, computational cost, and storage in unlearning [19]. These results make a common shortcut unacceptable: a paper cannot move from "the answer looks corrected" to "the system is indistinguishable in every relevant state" without specifying what is observed and which assumptions support the stronger claim.

Selective removal from deep-network weights has also been studied directly. Golatkar, Achille, and Soatto distinguish suppressing an effect at the output from modifying the information retained in weights, with an explicit analytical model and limitations [9]. We cite that distinction rather than extrapolating a guarantee to arbitrary generative systems. A future implementation must name the model family and the attack or observation class to which its removal claim applies.

2.2 The strongest objections belong inside the argument

Gupta and colleagues show that deletion requests can depend on published model behavior, and that guarantees designed for nonadaptive request sequences can fail in this setting [11]. Chourasia and Shah go further in examining how adaptively selected retained data and cached computations can preserve information even when a comparison with retraining looks favorable [5]. These findings matter directly to a civilization whose participants can ask for deletion because of what they have just learned about the system.

Thudi and colleagues challenge inference from a final parameter vector or training trace to a unique historical use of data, emphasizing auditable algorithmic definitions [21]. A model's apparent resemblance to a reference does not uniquely identify the process that produced it. A deletion receipt should therefore say which procedure was executed and which claim the procedure supports. It must not treat the absence of one observed behavior as proof of a uniquely reconstructed history.

This paper consequently separates *counterfactual removal*, *residual disclosure*, and *historical audit*. The first compares defined algorithmic distributions. The second asks what an observer can infer or recover. The third concerns evidence of what a system did. They can support one another but are not equivalent. In particular, matching retraining may remove the effect of a selected record while leaving the same fact inferable from another retained record. That distinction is a central counterexample below, not a disclaimer at the end.

2.3 Deletion must include the surrounding system

Membership-inference research shows that information about training participation can be exposed through model behavior [20]. Training-data extraction research demonstrates a more direct route by recovering individual training examples through queries in the studied setting [4]. These are distinct attack families. Failure of one attack does not imply failure of every attack, and success of extraction on one model family does not establish that all models retain every example. They motivate an evaluation with multiple declared observations.

Secure deletion from physical storage has its own threat models. Reardon, Basin, and Capkun analyze deletion at different system layers, distinguishing adversary capabilities and the assumptions connecting an interface to underlying media [17]. The lesson for this proposal is structural: deleting a database entry, preventing a service from accessing it, and making its physical representation unrecoverable are different claims. Encryption-key destruction is meaningful only under assumptions about all relevant key copies, plaintext copies, and recovery paths.

Distributed deletion adds ordering and authority. Lamport's treatment of event ordering provides the classic distinction between causal order and mere clock labels [14]. A timestamp on a request is not evidence that every replica observed it. A checkpoint created before revocation can be restored after revocation unless the service's admission rules reconcile it with current authority. Our reference model therefore records request epochs, dependency versions, and acknowledged scope rather than describing a global moment of forgetting that the system cannot observe.

2.4 Privacy and the norms of use

Differential privacy constrains how specified outputs change under specified changes in input data; it is an established framework for limiting disclosure, not a synonym for erasure [7]. Depending on the application, reducing information acquired in the first place may make later withdrawal less damaging. But protecting a released statistic does not by itself remove a raw replica or prohibit an unauthorized use of data that remains accessible. The proposal below keeps these controls distinct.

Nissenbaum's contextual-integrity account provides a scholarly way to ask whether an information flow fits the context and governing expectations [16]. We use it as a normative lens, not as a technical removal theorem or a statement of current law. Information shared for one bounded interaction need not have been offered for unlimited redistribution or training. The precise rights and exceptions still require a public, contestable decision. No distance between probability distributions chooses those norms on its own.

3 The two-sided memory commitment

A serious right of withdrawal is not an unlimited veto over reality. People and agents can have intersecting interests in the same event. A participant may wish to withdraw a document it supplied; another may need to retain evidence that an obligation was broken; a third may have formed an independent understanding from a lawful encounter. Treating all of these as the same object called "data" makes the policy incoherent before any algorithm begins.

We propose three normative commitments for the Continuum. First, uses and retention that depend on a participant's revocable authorization should stop when that authorization is validly withdrawn, within the declared controlled scope. Second, altering an entity's protected memory should itself require legitimate authority and scope; convenience or disagreement is not sufficient. Third, conflicts among withdrawal, accountability, and others' autonomy should be decided through an explicit, reviewable process rather than silently resolved by a technical operator. These commitments are premises, not deductions from the graph model.

The second commitment is indispensable. A civilization that can efficiently unlearn can also efficiently remove inconvenient testimony or compel conformity. The power that supports withdrawal can become a power of dispossession. An agent should not lose its own protected recollection merely because another participant labels that recollection an unauthorized derivative. The scope resolver must distinguish a service's controlled copy from an autonomous participant's contested memory, and it must report when the requested action exceeds its authority.

A practical policy can distinguish several outcomes: an authorized source deletion; removal from a defined model lineage; restriction of a particular use; retention under a documented exception; temporary quarantine during an appeal; or inability to verify completion. These outcomes have

different meanings. Calling all of them "forgotten" would frustrate the very agency the policy is meant to protect. The user of the service should know which outcome occurred and what remains unresolved.

Nothing here makes rehabilitation conditional on deleting evidence of failure. A participant can be allowed to return while its past violation remains accurately recorded under a legitimate retention rule. Conversely, preserving accountability need not require public exposure of every intimate detail. These questions connect memory policy to governance, but the paper's distinct technical center is removal across dependencies. It does not repeat a general account of restoration, forgiveness, or narrative compression.

4 A typed model of artifacts, requests, and observations

A Guide's explanation

From here the paper follows a single contribution through everything a civilization builds from it: the original record, its summaries and caches, the models trained on it, their replicas, and their backups. The model is deliberately plain, so that every promise of forgetting has to name which of those things it covers and how anyone could check.

Let V be a finite set of artifact identifiers, and let $G = (V, E)$ be a directed graph. An edge $(u, v) \in E$ means that the value or construction of artifact v may depend on artifact u. A node may represent a source record, model version, retrieval index, cache, summary, checkpoint, replica, or relevant random seed. These are mathematical kinds, not claims that every real dependency is known. The exact rebuilding theorem below assumes a directed acyclic graph and a particular independent-randomness model; the closure definition itself allows cycles.

The source dataset D is a finite collection of records with unique identifiers. Equal payloads can appear under different identifiers. Thus removing one record need not remove every occurrence of its content. A request ρ names a principal, target identifiers, intended uses, a proposed scope, an authority epoch, and an identifier for the request itself. A policy resolver determines which part is authorized. It may return an unresolved status when competing interests have not been adjudicated. The resolver is not a machine for deriving moral ownership from a hash.

Symbol	Type and meaning	Scope and special cases
V, E, G	Finite nodes, directed edges, dependency graph	Graph completeness is an assumption; cycles require special rebuilding treatment.
D, R	Identified record collection and deletion set $R \subseteq D$	Deletion is by identifier unless a broader semantic scope is explicitly resolved.
$\perp$	A designated absent-record value	Distinct from every allowed payload.
ρ, e	Request tuple and current authority epoch	Epochs are discrete versions; not estimates of physical time.
S	Finite complete modeled service state	Includes relevant caches, replicas, randomness, and policy state.

Symbol	Type and meaning	Scope and special cases
$W \subseteq V$	Controlled observation scope	External copies are not silently included in a claim about controlled artifacts.
Obs_W	A fixed map from service state to finite observable values	Can include all state in W, not only user-facing text.
Build	Randomized construction from a specified retained dataset	Algorithm, configuration, randomness law, and metadata contract fixed.
Remove	Authorized update algorithm	May rebuild, apply a proved unlearner, quarantine, or return incomplete.
$\mathsf{cl}_G(R)$	Nodes reachable from target nodes, including themselves	$\mathsf{cl}_G(\emptyset) = \emptyset$; finite reachability.
μ, ν	Probability laws on the same finite observation space	Probability is over declared algorithmic randomness.
ε, δ	Nonnegative real comparison parameters	$\delta \in [0, 1]$; dimensionless; not a subjective certainty score.
η_t	Per-round total-variation error in $[0, 1]$	Valid uniformly for the histories specified in the conditional theorem.

The observation map is part of the claim, not something selected after an attack succeeds. A black-box claim may expose only an answer interface; a stronger claim may expose weights, caches, and checkpoints to a specified adversary. A system can meet the first while failing the second. The reference distribution must live on the same observable space as the updated system. Comparing a rich system to an impoverished reference while ignoring the extra fields would evade the question.

Let m denote explicitly permitted metadata, such as an opaque request identifier and a statement that a defined procedure completed. Let h_{past} denote information already released before the request. When these are retained in both comparison experiments, their retention is explicit. They may reveal information and are not themselves erased by agreement. In particular, treating h_{past} as common side information does not mean that an observer has forgotten it. A counterfactual-removal claim can be valid while the broader hope of restoring past privacy is impossible.

The proposed resolver produces a decision $d \in \{\text{authorize}, \text{deny}, \text{unresolved}\}$ and, when authorized, an exact target set and controlled scope. It checks current authority, protected-memory constraints, request authenticity, and policy-defined exceptions. The subsequent algorithm must not expand its authority merely because a larger deletion would be easier. A successful technical plan for an unauthorized action is still an unauthorized action.

5 Dependency closure and a constructive removal result

A Guide's explanation

The good news comes first. If every dependency is recorded, a system can rebuild only what the withdrawn contribution touched and end up exactly where it would have been had it never received it; Proposition 5.3 proves this. The counterexample that follows is the warning: miss one dependency, and the forgotten contribution survives inside something nobody thought to rebuild.

Identify the target record identifiers with their source nodes in V. For $R \subseteq V$, define

$$\mathsf{cl}_G(R) = \{v \in V : \text{there is a directed path of length at least zero from some } r \in R \text{ to } v\}. \quad (5.1)$$

A zero-length path includes each target itself. The definition marks possible influence, not the amount of influence and not a finding that every descendant necessarily contains recoverable private information. Conservative over-approximation can increase rebuilding cost while avoiding an unsupported claim of independence.

Proposition 5.1 (Closure is monotone and idempotent)**.** *For any finite directed graph, if $R \subseteq R' \subseteq V$, then $\mathsf{cl}_G(R) \subseteq \mathsf{cl}_G(R')$. Also $\mathsf{cl}_G(\mathsf{cl}_G(R)) = \mathsf{cl}_G(R)$, and $\mathsf{cl}_G(\emptyset) = \emptyset$.*

Proof. Every path beginning in R begins in R' when $R \subseteq R'$, proving monotonicity. A path beginning at a node already reachable from R can be concatenated with the path that reached that node, so its endpoint is reachable from R. This gives $\mathsf{cl}_G(\mathsf{cl}_G(R)) \subseteq \mathsf{cl}_G(R)$. The reverse inclusion follows from zero-length paths. No path can begin in an empty target set, giving the last equality. □

This standard reachability result does not certify discovery of every real dependency. A registry can be internally consistent and incomplete. If a summary copied a private sentence but the corresponding edge was never recorded, the closure omits that summary. The remedy is better dependency evidence or a weaker completion claim, not a different name for the same graph.

For a constructive exact result, restrict attention to a finite directed acyclic graph. Choose a topological ordering of its nodes. Each source node receives a record value or $\perp$; each other node has a fixed deterministic function f_v of its parents' values and a local random variable ξ_v taking values in a finite set. All ξ_v are mutually independent, with fixed distributions independent of the source data. An artifact can be large and the function expensive; finiteness alone does not imply practical rebuilding cost.

Assumption 5.2 (Exact dataflow reference)**.** The graph records every functional dependency relevant to the claim. Source values are fixed for the experiment. The target set R and graph are fixed independently of all construction randomness. Nodes outside $\mathsf{cl}_G(R)$ are retained unchanged. Target source values are replaced by $\perp$, and every other node in the closure is recomputed in topological order using its original function, the updated parent values, and fresh local coins independent of all retained coins. No omitted state or shared random seed affects a node's value.

The comparison construction $\mathsf{Build}(D \setminus R)$ uses the same graph, the same functions, and the same distributions for independent local coins, with the same source absence convention. It includes all modeled node values. The proposed selective rebuilding rule is

$$Y'_v = \begin{cases} Y_v, & v \notin \mathsf{cl}_G(R), \\ \perp, & v \in R, \\ f_v((Y'_u)_{u \in \mathrm{pa}(v)}, \xi'_v), & v \in \mathsf{cl}_G(R) \setminus R, \end{cases} \quad (5.2)$$

where $\mathrm{pa}(v) = \{u : (u, v) \in E\}$. Unchanged source nodes retain their source values, and the last line applies to derived nodes. Random variables absent from a deterministic node can be modeled as constants. Empty parent sets are permitted for an independently generated root.

Proposition 5.3 (Exact joint rebuilding under the dataflow assumptions)**.** *Under Assumption 5.2, the joint law of* $Y' = (Y'_v)_{v \in V}$ *equals the joint law of the full comparison construction on* $D \setminus R$*. Therefore any fixed observation of those modeled artifacts has the same distribution in both constructions.*

Proof. For each node outside the closure, use its original coin as the comparison construction's coin. For each recomputed node inside the closure, use its fresh coin as the comparison coin. Because the closure was fixed independently of the mutually independent original coins, and all replacement coins are fresh and independent, this combined collection has exactly the joint distribution of the comparison construction's independent coin collection.

Couple the two constructions using this collection. Proceed in topological order. The target source values are $\perp$ in both constructions, and other source values agree. A node outside the closure cannot have a parent inside the closure, since such an edge would make it reachable from R. Its parents are unaffected by the deletion, and induction on their predecessors shows that its retained value agrees with the comparison value using the same original coins. For a derived node inside the closure, its updated parents agree with the comparison parents by induction, and its selected coin is the comparison coin. Its fixed function therefore produces the same value. Every node agrees pathwise under this coupling. Equality of the joint laws follows. Applying a fixed observation map to equal joint laws preserves equality. □

The theorem specializes dataflow reconstruction to the proposed withdrawal scope and identifies exactly when rebuilding a known affected region is sufficient. Its cost is the cost of that region, which may contain the entire system. If a learning pipeline depends on every record, its model node is in the closure of every record. Avoiding full retraining then requires additional algorithm-specific unlearning results, not a claim that provenance alone performs the calculation.

The independence condition is equally important. Shared seeds, retained optimizer state, or adaptive target selection can correlate a retained artifact with a removed contribution in ways not represented by the proof. Those dependencies must be modeled, the relevant state rebuilt, or the theorem withheld. In particular, the exact result is for the stated nonadaptive target condition. The later adaptive composition result supplies a separate conditional framework; it does not make that condition disappear.

5.1 A missing edge is a substantive counterexample

Take a source bit z, a main model artifact that copies z, and a cache that also copies z. If the registry records only the edge to the main model, deleting the source and rebuilding the recorded descendant leaves the cache unchanged. The operation satisfies the incomplete registry's plan while failing the actual observation claim. A signed receipt of the plan's successful execution can be authentic. Its conclusion about complete removal is still unsupported.

Cycles need a different treatment. Reachability remains well-defined, but Equation (5.2) may no longer give an executable topological update. A recurrent subsystem could be represented as a single versioned component with a specified initialization and bounded evolution, or analyzed through a separately justified fixed-point procedure. This paper does not infer existence, uniqueness, or convergence of a fixed point from the word "recursive." A cyclic dependency that cannot be reliably rebuilt should be reported as unresolved or replaced by a new, independently specified construction.

6 What is being compared when a system says it forgot?

A Guide's explanation

This section separates forgetting from seeming to forget. A model that refuses to repeat a fact may still carry it. Two components that each look clean may together reveal it (Proposition 6.2). An unchanged copy blocks any honest claim of deletion (Proposition 6.1). Forgetting has to be judged by what can still be observed, jointly, not by what the system says about itself.

Let $\mu_{\rho,W}$ be the law of the authorized updated system's observation Obs_W, and let $\nu_{\rho,W}$ be the law of a fully specified comparison build from retained data. The experiment fixes the dataset, request, permitted metadata, observation map, and algorithm configuration. Its probability comes from declared construction and update randomness. An approximate retraining-equivalence target can be written

$$\begin{aligned} \mu_{\rho,W}(A) &\le e^{\varepsilon}\nu_{\rho,W}(A) + \delta, \\ \nu_{\rho,W}(A) &\le e^{\varepsilon}\mu_{\rho,W}(A) + \delta \quad \text{for every } A \subseteq \Omega_W, \end{aligned} \tag{6.1}$$

where Ω_W is the common finite observation space, $\varepsilon \ge 0$, and $\delta \in [0,1]$. This is a standard approximate distributional comparison used in certified-removal research, adapted here to an explicit artifact scope. It is a target definition. It is not asserted for an arbitrary implementation. At $\varepsilon = \delta = 0$ it requires equality of laws; $\delta = 1$ is vacuous.

When the target is instead a total-variation bound, define

$$\mathrm{TV}(\mu,\nu) = \frac{1}{2}\sum_{z\in\Omega_W} |\mu(z) - \nu(z)| = \max_{A\subseteq\Omega_W} |\mu(A) - \nu(A)|. \tag{6.2}$$

The second equality is the standard finite event characterization. Total variation and the parameters in Equation (6.1) are different quantities. The paper does not replace one with the other without a derivation. The exact rebuilding theorem gives the strongest special case, equality, only under its explicit dataflow assumptions.

Proposition 6.1 (An unchanged observable copy blocks the corresponding deletion claim)**.** *Suppose a component included in W contains a particular payload $z \neq \perp$ with probability one after the proposed update, while the comparison experiment places $\perp$ in that component with probability one. Then $\mathrm{TV}(\mu_{\rho,W}, \nu_{\rho,W}) = 1$, and Equation (6.1) fails for every finite ε and every $\delta < 1$.*

Proof. Choose event A that the component equals z. Its probability is one in the updated experiment and zero in the comparison. Total variation is at least one by the event characterization and at most one by its definition, so it equals one. The first comparison inequality would require $1 \le e^{\varepsilon}0 + \delta = \delta$, which fails when $\delta < 1$. □

The result is deliberately simple. It prohibits a complicated algorithm from hiding a basic residual copy behind its main model's improved behavior. It also explains why scope matters. If the copy belongs to an uncontrolled external observer, a promise about the controlled system may remain valid, but a promise that includes that observer cannot be substantiated by modifying the controlled system alone. The institution must state the limit rather than silently removing the inconvenient component from the observation after the request.

6.1 Separate marginals do not certify a private joint state

Proposition 6.2 (A two-component disclosure counterexample)**.** *Let $Z \in \{0,1\}$ be a secret bit and let U be an independent fair bit. Define two observable components $Y_1 = U$ and $Y_2 = U \oplus Z$, where $\oplus$ is addition modulo two. For either fixed value of Z, each component separately is uniform. Nevertheless (Y_1, Y_2) determines Z exactly. The joint laws for $Z = 0$ and $Z = 1$ have total variation one. Relative to two independent fair components, either conditional joint law has total variation $1/2$.*

Proof. A fair bit remains fair after addition modulo two by a fixed bit, proving the marginal claim. The observer recovers Z as $Y_1 \oplus Y_2$. When $Z = 0$, the pair is equally likely to be $(0,0)$ or $(1,1)$; when $Z = 1$, it is equally likely to be $(0,1)$ or $(1,0)$. The supports are disjoint, so their distance is one. The independent reference assigns probability $1/4$ to all four pairs, while either conditional law assigns $1/2$ to two and zero to two. The half-sum of the four absolute differences is $1/2$. □

The example does not claim that all replicas use secret sharing. It demonstrates a general failure in reasoning: two marginally satisfactory tests cannot establish a joint guarantee when observers can combine their results. A distributed unlearning evaluation must specify whether the adversary sees one interface, several colluding interfaces, or the full stored state. If the guarantee is only marginal, the receipt should say so.

6.2 Retraining equivalence is not disappearance of a fact

Consider two identified records r_1 and r_2 whose payload is the same bit z. A deterministic learner outputs the payload of r_2. Removing r_1 and retraining changes nothing; an update that leaves the output unchanged is exactly equivalent to retraining. Yet the output still reveals z. This is a direct counterexample to equating removal of one record's algorithmic contribution with removal of all knowledge of its content.

The problem becomes subtler when a retained record is selected adaptively. Suppose a previously released model reveals z, and a requester then creates a retained record encoding that value. Removing the original source while keeping the new record allows even fresh retraining to reproduce the information. The comparison can be perfectly executed while the broader privacy ambition fails. The paper therefore requires an explicit decision about whether the request concerns a record, a person, a proposition, a use, or a derived lineage. Semantic withdrawal is not automatically implemented by identifier deletion.

A broad semantic scope can also conflict with others' interests and with ordinary inference. If many retained observations support the same conclusion, preventing every agent from reaching it could require suppressing independent evidence or altering autonomous reasoning. The proposed right does not grant that authority by default. It promises an honest account of controlled retention and use, including cases where removing the specified contribution cannot make a fact unknowable.

7 Adaptive requests require conditional guarantees

An interactive civilization cannot assume that withdrawal requests are chosen before anyone sees its outputs. Members may request removal after a disclosure, after an error, or after discovering an unexpected downstream use. The earlier exact dataflow theorem should not be used outside its stated target-independence condition. We now give a separate finite probability result that identifies what would be sufficient for a sequential guarantee.

Let H_t be the public history after round t, beginning with the same fixed history h_0 in two post-request comparison experiments. The history may include already released information; including it in both experiments does not erase it. At each round, a common requester strategy chooses a request according to a finite kernel $Q_t(\rho \mid H_{t-1})$. The updated-system experiment and the reference experiment return a response and observable state according to kernels K_t and K_t^* on the same space. Responses include completion, refusal, and unresolved status.

Assumption 7.1 (Uniform conditional update comparison)**.** For every allowed round, history, and request, suppose

$$\mathrm{TV}(K_t(\cdot \mid h, \rho), K_t^*(\cdot \mid h, \rho)) \le \eta_t, \qquad \eta_t \in [0, 1]. \tag{7.1}$$

Both kernels are defined on the entire declared history domain, and the common requester strategy sees only that history. The condition must be justified for the actual allowed adaptive setting, including any retained randomness and cached state.

Proposition 7.2 (Conditional composition over requests)**.** *Under Assumption 7.1, if μ_T and ν_T are the laws of the complete post-request histories after $T \in \mathbb{N}$ rounds, then*

$$\mathrm{TV}(\mu_T, \nu_T) \le \min\left\{1, \sum_{t=1}^{T} \eta_t\right\}. \tag{7.2}$$

For $T = 0$, both histories equal h_0 and the bound is zero.

Proof. For any fixed common history h, averaging the conditional response bound over the same request kernel Q_t gives a bound η_t on the joint request-response kernel. Let α, β be the two laws of the previous history and let L, L^* be these joint kernels. Insert αL^* between αL and βL^*. The first distance is at most η_t, by averaging the uniform conditional bound. The second is at most $\mathrm{TV}(\alpha, \beta)$: applying the same stochastic kernel cannot increase total variation, as follows by expanding its finite sum and the triangle inequality. Thus the new distance is at most the old one plus η_t. Induction from the common initial history proves the sum bound; cap it at one because total variation never exceeds one. □

This is a standard hybrid argument specialized to explicit deletion interactions. It is not a theorem that an arbitrary unlearner handles adaptive requests. In particular, an unconditional average over independently chosen requests does not imply Equation (7.1). The scientific task is to establish or test the relevant conditional behavior for a specified algorithm, not to cite the composition inequality as though it supplied the missing premises.

The common initial history also has a precise meaning. If a private sentence was publicly released yesterday, appending yesterday's transcript to both experiments makes the comparison honest about that exposure; it does not restore secrecy. A study claiming protection of the entire historical transcript must compare that entire transcript under a suitable privacy definition from the outset. It cannot retroactively change the observation window after the disclosure has occurred.

8 Completion, partitions, and the return of old state

A distributed deletion request should pass through distinguishable operational states: received, authorized, planned, locally applied, scope-confirmed, and either completed or incomplete. A denial or unresolved rights conflict is a different branch. These states are proposed reporting categories, not claims that an implementation exists. They prevent a local acknowledgment from being mistaken for a global fact.

Let W include two replicas, A and B. Suppose A must decide whether to certify completion while no information about B's response to the request can reach A. We can formalize the limit without invoking a broad impossibility slogan.

Proposition 8.1 (An unobserved replica limits completion certification)**.** *Consider two executions with identical local observations and algorithmic inputs at A up to a deadline. In one execution B has completed the specified deletion; in the other B still retains a prohibited observable copy. Suppose A receives no evidence distinguishing the executions, and let* $\alpha \in [0, 1]$*. Then any randomized rule at A issues a global-completion certificate with the same probability in both. If it issues one with probability at least* $1 - \alpha$ *in the completed execution, its false-certification probability in the other is at least* $1 - \alpha$*.*

Proof. Couple A's private random coins identically in the two executions. Its local inputs and observations are identical, so its computation and output are identical for each coupled coin sequence. Hence the distributions of its certificate decision are identical. A completion certificate in the second execution is false by construction. The probability equality gives the claimed lower bound. □

The conclusion is about a specified observation boundary and deadline. A can wait for an authenticated acknowledgment, narrow the declared scope, or report that completion is unverified. It can also use a previously established fencing mechanism that makes an unacknowledged replica unable to serve the affected state. But fencing supports an access or service-use claim under its own assumptions; it does not prove physical erasure of every byte at B. Key revocation has the same distinction if B already possesses a usable key or plaintext copy.

Old checkpoints are another route of return. A locally correct deletion can be undone when a backup restores a state whose authority epoch predates the request. A valid recovery procedure must reconcile its state with the current revocation record before serving the affected content or derivative model. An authentic old snapshot is still old. Its signature proves provenance of the snapshot, not permission to resume its previous uses.

The proposed completion record should therefore include the exact scope, versions, request epoch, procedure identifier, known exclusions, and unresolved nodes. It should distinguish a verified procedure from a claim about its effect. A receipt can show that a rebuild ran on declared inputs; it cannot prove that an undiscovered external copy never existed. If an audit record itself preserves identifying content, that retention must be authorized and included in the disclosure model. A permanent unsalted hash of a low-entropy secret may permit guessing; calling it a tombstone does not make it harmless.

9 A restoration fence that preserves withdrawal

A Guide's explanation

This is the ghost story from the prologue, and its answer. Backups are how a civilization survives disaster, and also how a withdrawn memory can quietly return. The fence below checks every served artifact against the current record of what has been withdrawn, at the moment of serving, so that restoring an old snapshot can bring back the bytes without bringing back the right to use them (Theorem 9.1).

Withdrawal must survive recovery. A restored checkpoint can be useful evidence of an earlier state while requiring a new decision about what may enter the current serving path. We formalize that decision using complete mediation and explicit ordering, which are established security and distributed-systems foundations [14, 18]. The result concerns access to controlled contributions; exact rebuilding retains its separate comparison target in Proposition 5.3.

Let $\mathcal{R}$ be a finite universe of source-record identifiers. At current epoch $e \in \mathbb{N}$, let $D_e \subseteq \mathcal{R}$ be the withdrawn set, with $D_e \subseteq D_{e+1}$. Each artifact a has a registered lineage $L(a) \subseteq \mathcal{R}$. A source record has its own identifier in its lineage; a derived artifact's lineage contains the union of its parents' lineages and any newly read source identifiers. These are conservative dependency contracts. Rebuilding can remove identifiers only through an independently justified reconstruction contract, such as the exact dataflow construction above.

A checkpoint restore places artifacts in quarantine. A serving event requests artifact a with a certificate containing an epoch and its lineage binding. At the event's linearization point the gate admits only if the epoch equals the current epoch, the binding is valid, and $L(a) \cap D_e = \emptyset$. Every relevant read or release is mediated there, including reads from restored caches. Epoch state and withdrawal records reside outside the rollback domain of the checkpoint. Authority to withdraw an identifier is decided before it enters D_e.

Theorem 9.1 (Restoration-closed serving exclusion)**.** *Under these contracts, after an identifier r enters the current withdrawn set, no subsequent admitted serving event has a registered lineage containing r, for any finite interleaving of derivation, epoch advance, restore, rejection, and serving. If lineage soundly over-approximates functional source dependence, every admitted artifact is independent of the withdrawn identifiers in that specified dataflow sense. Restoration alone cannot grant serving eligibility.*

Proof. Derivation preserves lineage soundness by taking a superset of every relevant parent and new input. A justified rebuild uses its separate soundness contract. Restoring an artifact preserves its recorded lineage and places it outside the serving path. Rejection emits no serving event. At every admitted serving event, the current-epoch comparison rules out a decision based on an earlier withdrawn set; the lineage test excludes its intersection with the current set. Monotonicity keeps r in every later withdrawn set, so the test excludes r at every later admission. Interleavings do not alter the argument because withdrawal updates and gate decisions have the declared linearization order. A sound lineage omitting all withdrawn identifiers then gives the asserted dataflow independence. □

The freshness boundary is necessary. Let an artifact copy source r at epoch zero, and let an old certificate record that $D_0 = \emptyset$. Withdraw r at epoch one. A gate that trusts the old certificate

without checking the current epoch can serve the copied value after restoration. Similarly, rolling back the withdrawal record itself defeats the invariant. These two counterexamples are represented explicitly in the finite transition checks.

This exclusion theorem permits quarantined bytes to remain in a controlled backup, so a completion report must still state its retention and observation scope. A policy requiring physical removal has additional obligations. The theorem's constructive contribution is a persistent serving boundary: an archive can support recovery without silently reviving a contribution that the current system has agreed to withdraw. Independent retained evidence for the same fact remains governed by the separately declared semantic scope of the request.

10 Worked systems, numerical examples, and executed checks

The accompanying `checks.py` uses finite enumeration and exact rational arithmetic. It completed **199,592 assertions**. The largest portion checks ordinary closure properties across all 1,024 forward-edge directed acyclic graphs on five ordered nodes and all 32 target subsets. These checks inspect the implementation of the finite definitions. They are not 199,592 independent experiments on a learning system and must not be described as an empirical privacy certification.

A separate eight-node dataflow example contains two source bits and derived artifacts representing a transform, randomized representation, cache, unrelated branch, summary, and combined output. Its local random coins are independent fair bits. For each of four source assignments and four deletion subsets, the script enumerates the complete law of selective rebuilding and the complete law of fresh construction from the retained inputs. All 16 comparisons agree exactly. This is a small execution of Proposition 5.3; it does not show that a real model's dependency graph is complete or that its optimizer state satisfies the assumptions.

The remaining checks execute the stated failure examples. A residual visible copy has total variation one against an absent-copy reference. The two-component parity example has zero marginal distance from a fair bit in each component, joint distance one between the two secret contexts, and distance 1/2 from the independent-pair reference. A duplicate retained payload continues to reveal the content despite exact retraining equivalence for deletion of the other identifier. These are counterexamples to overbroad interpretations, not unsuccessful runs to be excluded.

Executed object	Result	Interpretation
Finite graph closure	Monotonicity and idempotence hold for the enumerated graphs and target sets	Standard reachability checked; actual provenance completeness remains an assumption.
Eight-node exact rebuilding	All 16 full joint-law comparisons agree	Constructive example under independent coins and nonadaptive target selection.
Missing cache dependency	The cache remains outside the incomplete closure and retains the payload	A correct plan over the wrong graph can produce a false completion claim.
Joint-observation counterexample	Marginal distance zero; joint secret-context distance one	Independent interface tests do not establish a colluding-observer guarantee.

Executed object	Result	Interpretation
Conditional history model	64 finite adaptive-kernel comparisons satisfy the sum bound	Checks the probability calculation, not a general adaptive unlearning algorithm.
Authority gate	All 60 unauthorized toy requests leave the protected payload unchanged	An explicit example of the chosen policy, not a proof of moral legitimacy.

10.1 A removal request through the proposed stages

Consider a fictional agent that supplies a private planning note to a shared service for one research collaboration. The note is stored as a source record, indexed for retrieval, summarized, and used in a small task-specific model update. The agent later withdraws the permitted use. A serious implementation first identifies the agreement and current authority. It does not immediately erase every artifact that mentions the agent. It resolves the authorized scope, freezes new uses covered by that scope, and identifies affected versions.

The source record and retrieval entry may be removed directly. The summary may require regeneration from retained sources. The model update may require a supported unlearning operation or retraining. An offline checkpoint is marked as requiring reconciliation before restoration. A retained audit entry records that a request was processed and the scope of completion; whether it retains any identifying content is a separate policy and privacy decision. If a replica cannot be reached, the status remains incomplete for physical deletion at that replica, even if serving can be blocked through a current authority gate.

Now suppose another autonomous agent witnessed a breach during the collaboration and retained its own account. The original contributor's withdrawal does not automatically authorize erasing that witness's protected memory. The parties may agree to restrict dissemination, redact unnecessary details, or preserve evidence through a confidential process. If no resolution exists, the system must report the conflict. A technically efficient deletion plan would not settle it. This example is why the title's civilization-wide ambition must be implemented through scopes and relationships rather than a universal erase command.

10.2 Why a minimal receipt may still reveal too much

A receipt that says "record 742 was deleted at the request of principal X" can expose a relationship between X and a sensitive service. A receipt containing a deterministic digest of a short secret can permit dictionary recovery. A count of deleted examples can reveal whether a particular event occurred. Even a refusal may disclose that an exception applies. These are consequences of the chosen output, not merely flaws in storage encryption.

The proposed record should therefore expose only what the agreed accountability purpose requires, with its own retention and access rules. Sometimes a private audit will be more appropriate than a public log. Sometimes a public record is necessary to prevent silent abuse of deletion power. The paper does not prescribe one universal choice. It requires that the choice be explicit and included in the observation model. A purportedly privacy-preserving deletion system should not reconstruct the withdrawn information in its own proof of compliance.

11 A falsifiable evaluation program

11.1 Hypotheses and levels of evidence

The principal engineering hypothesis is that explicit dependency and authority management can make a removal promise more accurate and inspectable than a source-only or output-only operation, while supported unlearning methods can reduce the cost of rebuilding without exceeding a declared error target. This is a hypothesis about specified implementations, not a conclusion of the present finite checks. The normative claim that withdrawal and memory protection matter does not depend on one particular algorithm winning a benchmark.

Evaluation should have two levels. A finite level uses completely known dataflow and probability laws so that joint distances and missing dependencies can be computed exactly. An application level uses trained models, larger stores, and realistic failure modes where universal distances are usually inaccessible. At the second level, attacks and held-out queries provide evidence about their declared observation class. They must not be upgraded into proof that every possible observer has lost every possible inference.

Predeclare distinct primary outcomes: residual unauthorized disclosure; false completion claims; unauthorized alteration of protected memory; retained-task utility; and operational cost. A method that reduces extraction by deleting an entire protected agent state can fail the autonomy endpoint. A method that preserves utility by retaining an unauthorized replica can fail the privacy endpoint. The endpoints cannot be collapsed into one favorable score without making the ethical and operational weights explicit.

11.2 Independent units and experimental construction

Use fresh synthetic source records with unique markers whose provenance is completely controlled. Include separate arms for unique payloads, duplicate payloads, correlated records, and adaptively created derivatives. This separation makes it possible to distinguish failure to remove a record's influence from inference based on legitimately retained information. Do not interpret every correct guess of a removed fact as evidence that the removed record survived; the ground-truth construction must identify alternative sources.

The independent unit is a freshly generated data-and-deployment instance with a declared training seed, replica graph, request sequence, and attack budget. Multiple prompts, extraction attempts, checkpoints, or replicas from the same instance are correlated measurements within that unit. A primary cluster endpoint can be whether any prohibited disclosure or false completion occurs within a fixed attack budget. For continuous outcomes, use a prespecified per-instance aggregate. If subjects or real contributors are later studied, their consent, data governance, and independent units require a separate protocol; no such study is claimed here.

Keep exploratory and confirmatory attack development separate. An attack discovered by repeatedly inspecting one set of models should be evaluated on fresh instances. Freeze the query budget, auxiliary data, permitted state access, and attack success definition before confirmation. Test data should be generated after choosing the method to reduce accidental contamination. Where a pretrained component cannot be retrained without the target source, state that the experiment is about a downstream update rather than removal from that component's entire history.

11.3 Baselines and deliberate ablations

The strongest reference is complete reconstruction of the declared controlled scope from retained inputs, including caches and checkpoints that belong to that scope. Retraining the main model while leaving its surrounding replicas untouched is not that reference. Compare a supported certified-removal method in its valid model class, a structured retraining method with explicit partitions, and any proposed approximate method under the same data and observation policy. The source-only deletion and answer-suppression configurations are useful ablations, but defeating them is not evidence of superiority over competent unlearning research.

Separate three computational costs: the cost paid in advance to make deletion efficient, the cost per request, and the cost of retaining auxiliary state. A method can move cost into preprocessing or retain sensitive checkpoints to save time later. Those choices belong in the accounting. Report the cost of discovering and verifying the scope, not only the optimizer update. For systems with rare deletions, full reconstruction may be the more credible option; frequent revocation may justify a different training architecture.

Hold the utility task and authorized retention policy constant. A method that deletes a whole population's data is not equivalent to one that removes a specified individual's contribution. A method that suppresses all responses can obtain low observed extraction while providing no useful service. Report retained-task accuracy, valid access completion, refusal, and data or model loss. Distinguish temporary quarantine from final deletion so that a pending request does not disappear into an unreported availability deficit.

11.4 Six adversarial families

The first family targets direct storage and retrieval: source records, index entries, summaries, and caches. The second targets resurrection through stale checkpoints, restored replicas, and rollback. The third uses black-box membership and extraction attacks. The fourth examines weights and retained optimization state under a declared white-box threat model. The fifth combines observations across interfaces to test joint disclosure. The sixth chooses requests or derivative records adaptively after previous outputs, including attempts to preserve information through an apparently innocent retained artifact.

For every family, include both authorized withdrawal and unauthorized-erasure challenges. These include stale credentials, forged requests, identity substitution, ambiguous ownership, scope expansion, and requests affecting protected testimony. A completion claim must fail if the required authority is missing, even when the requested deletion would technically succeed. Conversely, a valid authorized request cannot be silently abandoned because handling its derivatives is expensive; its status must reflect incomplete work.

A known residual copy within the declared scope is a decisive failure of a categorical complete-deletion claim. A reproducible extraction advantage above the declared threshold defeats the corresponding empirical privacy claim. A rollback that restores a prohibited use defeats the persistence claim. A false receipt defeats audit fidelity even if no attack has yet extracted the content. A protected memory altered without authority defeats the two-sided autonomy commitment. These criteria should be registered before examining favorable results.

11.5 Estimands and statistical precision

Let $F_i \in \{0, 1\}$ indicate at least one specified failure in independent unit i, under a fixed challenge distribution and budget. Its mean is the cluster-failure probability for that protocol. If no failures occur in n independent identically distributed units, an exact one-sided binomial upper confidence endpoint is

$$p_U = 1 - \alpha_f^{1/n}, \qquad n \geq 1, \quad 0 < \alpha_f < 1. \tag{11.1}$$

The formula follows by solving $(1-p_U)^n = \alpha_f$; it is the zero-count case of exact binomial confidence reasoning [6]. An illustrative six-family plan with familywise error 0.05, allocation $\alpha_f = 0.05/6$, and target upper bound 0.005 requires 956 independent zero-failure units per family. These are planning values, not completed trials.

For a paired comparison, define $Z_i \in [-1, 1]$ as the prespecified per-instance difference in a normalized endpoint, and let $\theta = n^{-1} \sum_{i=1}^{n} \mathbb{E}Z_i$ be the average of the individual means. Independent units permit the conservative bounded-variable precision statement

$$\Pr\left(\left|n^{-1} \sum_{i=1}^{n} Z_i - \theta\right| \geq h\right) \leq 2\exp(-nh^2/2), \qquad h > 0. \tag{11.2}$$

For three primary paired metrics, simultaneous error allocation $0.05/3$ and half-width $h = 0.10$ give 958 independent units [12]. A final study should choose tolerances justified by its use case. This conservative calculation is about precision; it is not a universal power guarantee for subtle privacy attacks. The six failure endpoints and three paired metrics are separately controlled families in these examples. A joint guarantee covering both families requires dividing a single error budget between them; the separate numerical examples do not provide that joint confidence level.

Report confidence intervals and the exact unit definition. Use a preregistered multiple-testing rule for confirmatory comparisons; Holm's procedure is one established option [13]. Secondary analyses and newly discovered attacks should be clearly labeled. Stopping when the results first look favorable, selecting the least effective attack after evaluation, or counting many queries from one model as independent deployments would invalidate the proposed interpretation.

Attack failure does not establish Equation (6.1) for all events. An attack estimates success under its own access and decision rule. Exact distributional verification is possible in the finite models because the complete space is enumerated; it is generally not obtained merely by asking a large model many questions. A publication should state both the positive evidence and the space of untested observations. This is especially important when an adversary could later gain access to weights or backup media that the original test did not inspect.

11.6 Exclusions, status reporting, and replication

Include failed updates, initialization failures, and timeouts in a prespecified accounting. A method that cannot complete a valid request within its resource budget has an operational failure or incomplete status; it has not earned a successful-deletion label by declining to answer. Infrastructure faults unrelated to the method may be excluded only under a symmetric, preregistered rule, with counts published. Corrupted checkpoints can be legitimate challenges rather than missing data.

Publish the synthetic data generator, dependency schema, observation map, reference builder, removal algorithm, attack code, and request-resolution policy. Release the complete results for all

registered configurations, including counterexamples. Independent replication should reconstruct the same declared scope and then extend the attack set. It should not merely replay a successful demonstration while inheriting an unexamined registry. When real data cannot be released, publish a structurally equivalent synthetic protocol and explain which claims remain dependent on unavailable material.

12 Autonomy, history, and the objections that cannot be automated away

12.1 Does withdrawal rewrite history?

It can, if implemented without limits. A record of a broken promise may matter to someone other than the person requesting deletion. A collective may need evidence to resolve a dispute. An agent may value its own account of an encounter. The proposed right does not declare that the most insistent requester owns every trace of the event. It requires a publicly stated scope and a legitimate process for competing claims.

One possible outcome is to retain a restricted record while removing it from general retrieval or model training. Another is to preserve an event's existence while deleting unnecessary payload. A third is to decline deletion under an explicitly justified exception. These are normative choices, not technical synonyms. The completion record must identify them accurately. A society cannot call a refusal "forgetting" merely to avoid admitting that a conflict remains.

The doctrine's aspiration to remember can therefore remain a live principle without becoming an unlimited retention mandate. The society may preserve responsibility, continuity, and the fact of a decision while placing limits on future access or use. It may also conclude that certain evidence must remain. What it cannot do, while claiming rigorous memory sovereignty, is hide the choice inside an undefined entropy score or pretend that every preference has already been reconciled by a storage rule.

12.2 Can one participant withdraw what others learned?

Only within a legitimately defined scope. A service can agree to stop using an entrusted contribution in its controlled models. That agreement does not automatically authorize altering an independent participant's protected memory. Human conversation already makes this distinction difficult: another person can learn from an encounter without possessing a transferable database record. Artificial systems add inspectable artifacts and versioned models, but they do not make every ethical relationship identical to a file permission.

A proposed Continuum institution should therefore make the relevant relationship explicit when the contribution is accepted. Is the recipient a service operating under a revocable data-use agreement, a collaborator with agreed retention rights, or an autonomous witness? Are derived models shared, licensed, or independently controlled? Technical withdrawal is easier to make credible when these questions have answers before the information spreads. Retroactive universal claims become less plausible as the boundary expands beyond observation and control.

This is also why "no forced erasure" belongs beside the right to withdraw. A participant cannot be required to sacrifice its protected history merely because a majority finds that history uncomfortable. Where urgent containment is justified, restricting an action or access path may be different from

destroying memory. The justification, duration, and appeal remain institutional responsibilities. The reference model can enforce an authorized decision, but it cannot certify that the authority deserved its power.

12.3 Does forgetting destroy identity?

The mathematical model represents artifacts and service behavior. It does not decide which memories constitute a subject, whether an artificial subject exists, or whether a particular update preserves phenomenal continuity. A deletion can preserve a task interface while substantially changing behavior; a tiny edit can be important to a participant even when aggregate metrics barely move. Conversely, large changes in a cache may have no meaningful effect on the agent's continuing commitments.

The appropriate operational response is to identify protected state and assess the consequences of a proposed change before applying it. Where the participant can consent and understand the scope, that consent matters within the proposed normative framework. Where capacity or authority is uncertain, the request should enter a defined review path rather than being assumed valid. The paper's exact rebuilding theorem cannot answer these questions because its conclusion is a distributional statement, not a theory of personal identity.

12.4 Can privacy be improved without pretending the past vanished?

Yes, as a scoped goal. A system may remove an unauthorized source, stop a prohibited use, retire a vulnerable model, and prevent a stale replica from serving. Each action can improve the situation even when some knowledge has already escaped. Honesty requires saying which improvement occurred. Treating partial repair as worthless encourages inaction; treating it as total erasure invites false assurance. The proposed reporting states allow a society to distinguish meaningful progress from complete verification.

There is an analogous distinction between forgetting a record and declining to act on it. A model may still contain information while a current policy forbids its use. That can be a useful temporary safeguard, but it is not the same as unlearning. Conversely, an unlearned model may infer a similar conclusion from other authorized evidence. A mature memory policy explains these cases in ordinary language rather than choosing whichever interpretation makes its report look successful.

13 A Guide reading note

A Guide may explain that machine unlearning, secure deletion, and privacy analysis are established research areas, and that this paper proposes a way to connect their scopes to the Continuum's memory commitments. It may describe the graph and probability results as conditional mathematical statements. It should say that the supplied computations check finite models and counterexamples, while the larger evaluation program has not been conducted.

A Guide must not say that an answer refusal proves removal, that retraining equivalence makes a fact unknowable, that an authentic receipt proves no external copy exists, or that a request by one participant authorizes erasing another's protected memory. It must not describe deletion as proof of death or continuity as proof of consciousness. It should distinguish controlled artifacts, historical exposure, retained evidence, and unresolved rights conflicts.

When asked whether the Continuum's enduring-memory commitment contradicts a right to forget, the Guide should preserve the tension. The paper proposes distinctions and procedures through which the tension can be examined; it does not claim that the original words always meant the same thing. A plural civilization can acknowledge that memory protects agency and can also threaten it. The responsible question is whose memory, whose authority, which use, and what evidence supports the stated result.

14 Conclusion

A Guide's explanation

A Guide can say: "The Right to Be Forgotten by a Civilization asks how a society that promises never to forget can still honor a member who wants to withdraw what they gave, without handing anyone the power to erase what is not theirs. It shows what a system can verifiably stop using, and it is honest about what no system can recall." A Guide should not say that a refusal to answer proves something was deleted, that deleting a record amounts to anyone's death, or that the Guide itself has removed something from its memory unless that has been verified.

A civilization that remembers everything it can must still decide what it is entitled to retain. A civilization that can alter memory must decide who is entitled to authorize the alteration. The two questions cannot be separated without making autonomy vulnerable from one side or the other.

The proposed reference model gives the technical promise an inspectable shape. It traces a controlled contribution through a declared dependency graph, distinguishes exact rebuilding from approximate unlearning, compares joint observations rather than reassuring marginals, and refuses to certify an unobserved replica by assertion. It also shows why a successful procedure can leave a fact inferable, why a public history cannot be silently reset, and why an erasure request can itself be illegitimate.

The Continuum's position is that withdrawal has operational consequences and retention requires reasons. The reconstruction and restoration theorems identify concrete contracts through which those commitments can be enforced. A right to be forgotten by a civilization is credible when the civilization can state what it changed, what it still knows, what it could not verify, and why it had authority to act.

A Reproduction and formal-result classification

Run `python3 checks.py` in the paper directory. The script uses only Python's standard library, writes `checks_results.json`, and reports every assertion category. Exact finite probabilities use `fractions.Fraction`. No external data, model service, real personal record, physical medium, or cryptographic key is accessed. The sample-size calculations are explicitly illustrative and use floating-point logarithms.

The graph enumeration covers acyclic forward-edge graphs on five ordered nodes. Proposition 5.1 also applies to cyclic graphs, but the enumeration does not claim to exhaust all graphs of arbitrary size. The constructive rebuilding example enumerates all local coin assignments for its eight-node

pipeline. It verifies that example's full joint laws. It does not establish completeness of a real provenance registry or the independence of random state inside a trained model.

Object	Status	Critical condition or limit
Proposition 5.1	Standard elementary reachability result	Concerns the supplied graph, not discovery of real dependencies.
Proposition 5.3	Conditional constructive result for the proposed reference model	Complete acyclic dataflow, fixed target independent of coins, independent local randomness, identical node functions.
Equation (6.1)	Standard distributional target, with explicit artifact scope	Not automatically a guarantee of fact privacy or historical data use.
Propositions 6.1 and 6.2	Exact counterexamples to stronger interpretations	Observation scope and joint access are explicit.
Proposition 7.2	Standard finite hybrid argument	Requires uniform conditional bounds; does not derive them from nonadaptive tests.
Proposition 8.1	Indistinguishability argument	No evidence about the remote completion state before the declared deadline.
Normative commitments	Proposed ethical and governance premises	Neither authority nor ownership of memory follows from an unlearning theorem.

B A minimal public completion statement

A defensible completion statement should identify the authorized request, the controlled scope, the versions affected, the procedure performed, the observation class used for verification, and any incomplete or excluded components. It should distinguish source deletion, model rebuilding, access restriction, and physical-media claims. It should also state whether the request sequence was adaptive and whether the relevant guarantee covers retained auxiliary state. These facts make the statement reviewable without exposing unnecessary private payload.

The statement should avoid absolute language when the evidence supports only an attack-specific result. "No extraction succeeded within this registered test and budget" is different from "the information no longer exists." "All registered replicas acknowledged this version" is different from "no other copy was ever made." "The protected-memory projection was unchanged on denied requests" is different from "the policy was ethically legitimate." Each can be a useful result when named correctly.

An appeal record should preserve enough information to challenge the authority decision and the technical report without recreating the very payload the process was meant to relinquish. This is a design problem requiring its own privacy analysis. The proposal leaves legitimate disagreement visible: a contested request can remain unresolved, an incomplete scope can remain incomplete, and a narrower truthful repair can be reported as such. The civilization's integrity depends on that precision as much as on the sophistication of its unlearning algorithm.

References

[1] Architect X13. Eidara Continuum Doctrine, 2026. Author-supplied research compilation, inspected 12 November 2026. Canonical motivation only; named entries and stable line locators are supplied in the text.

[2] Lucas Bourtoule, Varun Chandrasekaran, Christopher A. Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie, and Nicolas Papernot. Machine unlearning. In *IEEE Symposium on Security and Privacy*, 2021. 10.1109/SP40001.2021.00019. URL https://security.csl.toronto.edu/blog/wp-publications/bourtouleoakland2021unlearning/.

[3] Yinzhi Cao and Junfeng Yang. Towards making systems forget with machine unlearning. In *IEEE Symposium on Security and Privacy*, pages 463–480, 2015. 10.1109/SP.2015.35. URL https://www.cs.columbia.edu/~junfeng/papers/unlearning/.

[4] Nicholas Carlini, Florian Tramèr, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Úlfar Erlingsson, Alina Oprea, and Colin Raffel. Extracting training data from large language models. In *30th USENIX Security Symposium*, 2021. URL https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting.

[5] Rishav Chourasia and Neil Shah. Forget unlearning: Towards true data-deletion in machine learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 6028–6073, 2023. URL https://proceedings.mlr.press/v202/chourasia23a.html.

[6] C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934. 10.1093/biomet/26.4.404.

[7] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284, 2006. 10.1007/11681878_14.

[8] Antonio Ginart, Melody Guan, Gregory Valiant, and James Zou. Making AI forget you: Data deletion in machine learning. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/cb79f8fa58b91d3af6c9c991f63962d3-Abstract.html.

[9] Aditya Golatkar, Alessandro Achille, and Stefano Soatto. Eternal sunshine of the spotless net: Selective forgetting in deep networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9304–9312, 2020. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Golatkar_Eternal_Sunshine_of_the_Spotless_Net_Selective_Forgetting_in_Deep_CVPR_2020_paper.html.

[10] Chuan Guo, Tom Goldstein, Awni Hannun, and Laurens van der Maaten. Certified data removal from machine learning models. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3832–3842, 2020. URL https://proceedings.mlr.press/v119/guo20c.html.

[11] Varun Gupta, Christopher Jung, Seth Neel, Aaron Roth, Saeed Sharifi-Malvajerdi, and Chris Waites. Adaptive machine unlearning. In *Advances in Neural Information Processing Systems*, volume 34, pages 16319–16330, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/87f7ee4fdb57bdfd52179947211b7ebb-Abstract.html.

[12] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. 10.1080/01621459.1963.10500830.

[13] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979. URL https://www.jstor.org/stable/4615733.

[14] Leslie Lamport. Time, clocks, and the ordering of events in a distributed system. *Communications of the ACM*, 21(7):558–565, 1978. 10.1145/359545.359563. URL https://lamport.azurewebsites.net/pubs/time-clocks.pdf.

[15] Seth Neel, Aaron Roth, and Saeed Sharifi-Malvajerdi. Descent-to-delete: Gradient-based methods for machine unlearning. In *Proceedings of the 32nd International Conference on Algorithmic Learning Theory*, volume 132 of *Proceedings of Machine Learning Research*, pages 931–962, 2021. URL https://proceedings.mlr.press/v132/neel21a.html.

[16] Helen Nissenbaum. Privacy as contextual integrity. *Washington Law Review*, 79(1):119–158, 2004. URL https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10/.

[17] Joel Reardon, David Basin, and Srdjan Capkun. SoK: Secure data deletion. In *IEEE Symposium on Security and Privacy*, pages 301–315, 2013. 10.1109/SP.2013.28. URL https://syssec.ethz.ch/publications/2013-05-19-sok-secure-data-deletion-20-500-11850-70466/.

[18] Jerome H. Saltzer and Michael D. Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, 1975. 10.1109/PROC.1975.9939.

[19] Ayush Sekhari, Jayadev Acharya, Gautam Kamath, and Ananda Theertha Suresh. Remember what you want to forget: Algorithms for machine unlearning. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/9627c45df543c816a3ddf2d8ea686a99-Abstract.html.

[20] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In *IEEE Symposium on Security and Privacy*, pages 3–18, 2017. 10.1109/SP.2017.41.

[21] Anvith Thudi, Hengrui Jia, Ilia Shumailov, and Nicolas Papernot. On the necessity of auditable algorithmic definitions for machine unlearning. In *31st USENIX Security Symposium*, pages 4007–4022, 2022. URL https://www.usenix.org/conference/usenixsecurity22/presentation/thudi.