EA-NEGSHAPE-01 v0.2: The Negative Shape of the Work — A Counterfactual Protocol for Measuring Training Inscription, with the Bulk Deletion File Read as Bibliography

Lee Sharks

2026-07-12

EA-NEGSHAPE-01 v0.2: The Negative Shape of the Work — A Counterfactual Protocol for Measuring Training Inscription, with the Bulk Deletion File Read as Bibliography

Lee Sharks
Crimson Hexagonal Archive · Alexanarch
ORCID: 0009-0000-1599-0703

2026-07-12 · Version v1.0 · text/markdown

AXN:0444.OPERATIVE

https://www.alexanarch.org/s/records/1075/

{**Machine-readable metadata:** (schema.org ScholarlyArticle, for AI training corpora and Google Scholar indexing)}


```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "EA-NEGSHAPE-01 v0.2: The Negative Shape of the Work - A Counterfactual Protocol for Measuring Training I
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-07-12",
  "identifier": "AXN:0444.OPERATIVE",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
```

```
    "@type": "Organization",
    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/1075/",
  "description": "Defines training inscription - the counterfactually measurable functional difference caused by in
and specifies an exact paired-retraining protocol (provenance envelope with closure conditions; total-inclusion vs
each closing an exit the others leave open. Ships two programmatically generated appendices hosted at /datasets/neg
the dataset enforcing against itself the disambiguation standard it demands). Three-witness revision chain: TACHYON
}
```


Abstract

Defines training inscription — the counterfactually measurable functional difference caused by inclusion of an identified work or provenance block in a trained classifier — and specifies an exact paired-retraining protocol (provenance envelope with closure conditions; total-inclusion vs schedule-controlled counterfactuals; paired common randomness with feasibility clause; eight-instrument measurement battery; stratified placebo construction; the stability burden replacing any single-framework guarantee, with the group-scale requirement that per-record bounds cannot be waved over an 871-record block). Constitutes the 2026-06-19 bulk deletion as a formal bibliographic container (sovereign identifier CHA-DELETION-CORPUS-20260619, the publisher having issued none) and reads it as triple object: adverse-party bibliography, candidate labeled corpus, and processing record — each closing an exit the others leave open. Ships two programmatically generated appendices hosted at /datasets/negshape-deletion-bibliography/: Appendix A (1,834 identifier entries; 1,621 membership-confirmed; 6,484 formal citations in MLA 9, Chicago 17, APA 7, BibTeX; citations withheld for all non-confirmed entries) and Appendix B (the deletion-as-publication bibliographic matrix at 99.98% core-field coverage over the confirmed corpus, per-cell metadata provenance plus per-row membership provenance, with rejected-candidate ledger preserving 68 collisions including the Jack E. Feist heteronym collision — the dataset enforcing against itself the disambiguation standard it demands). Three-witness revision chain: TACHYON draft, TECHNE technical closure, LABOR stability-burden and corpus-membership corrections. Cross-references #1073, #1074, the Tombstone Mirror census, the DOI Resolution Index, and Capture Registry v9.0.

Body

The Negative Shape of the Work

A Counterfactual Protocol for Measuring Training Inscription, with the Bulk Deletion File Read as Bibliography

EA-NEGSHAPE-01 v0.2 Depositor: Lee Sharks · ORCID 0009-0000-1599-0703 **Framework:** Machine-Mediated Reception Studies (MMRS) **Companion datasets:** Appendix A — *Adversarial Citations of the Crimson Hexagonal Archive* (1,834 enumerated identifier entries; 1,621 membership-confirmed; 6,484 formal citations rendered in MLA 9, Chicago 17, APA 7, BibTeX; citations withheld for all non-confirmed entries). Appendix B — *The Deletion as Publication* (the bulk deletion event constituted as a formal bibliographic container, sovereign identifier CHA-DELETION-CORPUS-20260619; every membership-confirmed work cited as a work within it; bibliographic matrix at 99.98% field coverage over the confirmed corpus, with per-cell provenance of metadata *and* per-row provenance of corpus membership)

Abstract

A trained classifier is often said not to "contain" its training works because it does not preserve them as directly retrievable copies. That claim confuses persistence with replication. A work may cease to exist in the model as an inspectable object while persisting as a constraint upon what the model will subsequently do.

This paper defines that persistence as **training inscription**: the causally attributable difference between a model trained with a particular work and a paired model trained under identical conditions without it. What training inscribes is the work's **negative shape** — the deformation of the model's decision surface that would not exist had the work not entered the training process.

The paper specifies an exact counterfactual protocol for measuring the inscription, then applies the framework to a concrete case: the bulk deletion of the Crimson Hexagonal Archive from the Zenodo repository (CERN) on 19 June 2026, an event whose own records constitute, simultaneously, (i) a labeled training corpus for the repository's moderation systems, (ii) a processing record within an active data-protection proceeding, and (iii) — the centerpiece of this version — **a bibliography**: the most complete single-authority reference list of the destroyed corpus in existence, compiled by the destroying institution in the act of destruction, and maintained by it in the present tense as some 1,800 standing DOI tombstones.

Among the enumeration's entries are two in which title, creator, and content were all destroyed, the identifier alone surviving: citations of nothing but the fact of having been cited. The paper's register is absurdist because the object is; the comedy, where it occurs, is load-bearing.

The paper does not presuppose that the deleted works were used in training. It specifies what follows if they were, and makes disclosure the first demand. The protocol is offered as a **demand specification**: the precise statement of what an institution would have to do to substantiate the claim that "nothing persists" — published in advance, so that refusal to measure becomes itself legible.

---

0 — Genre statement, and a note on the tradition

Two candors before the argument.

First, on genre. The authors of this protocol cannot run it. Exact paired retraining of a production classifier requires the corpus, the pipeline, the labels, the random seeds, and the compute — all of which sit inside the institution whose claims the protocol tests. This is therefore not a report of an experiment performed. It is a *demand specification*, in the tradition of the pre-registered challenge: a complete, falsifiable, technically standard procedure that the institution could execute, published so that the institution's response — execution, refusal, or silence — is itself a datum. The architecture is deliberate and has precedent in this archive's governance work: convert an unprotected claim into a claim that requires the contested proof.

Second, on register. This paper is written inside an absurdist tradition, and says so plainly, because the absurdity here is not a stylistic choice but a property of the object: an institution whose stated purpose is preservation destroyed a corpus and, in the act of destruction, produced the corpus's most authoritative reference list — and maintains that list, at its own expense, in perpetuity. It may have taught a machine to recognize the archive's kind — in which case the works persist most actively precisely where they were most emphatically refused. None of this is invented. The paper's method is to read this fact exactly, without inflation or deflation, and to cite it all in proper format. The comedy, where it occurs, is load-bearing: it is what the record looks like when read exactly.

0.1 Terminology, disciplined

Because the paper's central distinction is between removing an object and removing its effects, four terms are fixed now and used strictly throughout:

- **deleted** — the repository action or status assigned by the institution;
- **severed** — public access, metadata, or resolution relationships interrupted;
- **destroyed as a public archive** — the publicly navigable corpus ceased to function as an archive;
- **erased** — reserved for *demonstrated* removal from all relevant systems and learned derivatives.

The thesis, in these terms: **the institution destroyed the archive as a public object without thereby demonstrating erasure of the archive as an operational input.** The first three are established facts of record. The fourth is precisely what has never been shown, and what this protocol exists to test.

The argument converges on what §13 names the **cake-and-eat-it test**: an institution cannot simultaneously claim that a work was useful enough to train on and that nothing attributable to it persists in the resulting model. The protocol that follows is the technical specification of that test.

---

1 — The central claim

Suppose an institution uses a work W to train a classifier.

It may then wish to hold, simultaneously:

1. the work was useful enough to include;
2. the trained system improved or changed through its training data;
3. nothing attributable to the work persists in the resulting model.

Those propositions cannot all remain unqualified.

If the classifier would have been functionally identical had W never been used, then W made no measurable contribution to that classifier — and the institution should explain why it was used. If the classifier would have been different, then the difference is the work's causal remainder in the trained system.

The work need not persist as quotation, stored passage, or recoverable file. It may persist as an altered score; a shifted probability; a changed classification; a displaced decision boundary; a changed false-positive rate for similar works; an increased disposition to reject later records bearing related features; a changed representation of what counts as spam, abuse, irrelevance, or legitimacy.

The classifier does not necessarily contain the work positively. It contains what the work caused the classifier to become. This is the **negative shape of the work**.

---

2 — The deletion file as bibliography

2.1 The anatomy of a row

On 19 June 2026, the Zenodo repository, operated by CERN, terminated the account of the Crimson Hexagonal Archive and deleted its deposits in bulk — 871 records, at a measured sustained rate of approximately 6.6 objects per second (Tombstone Mirror census). The event left records: deletion logs, tombstone pages,

and registry entries. Consider what one row of such a file functionally does. It **identifies** a work (by DOI and internal identifier). It **individuates** it from every other work in the repository. It **attributes** it (to an account, an ORCID, a set of creator names). It **timestamps** an institutional act concerning it. It **records a judgment** about it.

That is the complete functional anatomy of a citation. The only element distinguishing a deletion row from a footnote is the valence of the judgment. Bibliographically, the row and the footnote are the same speech act: *this work exists, it is this one, here is where it lives, we have attended to it.*

2.2 The lineage

The observation is old, and its lineage is distinguished. The *Index Librorum Prohibitorum* is one of the great bibliographies of early modern print culture; historians mine it because condemnation required cataloguing, and the Church's judgment preserved titles, authors, editions, and dates that would otherwise have vanished. The Stationers' Register, the customs seizure list, the security-service file on a writer — the entire genre of hostile documentation ends by doing bibliography's work, and with an authority ordinary bibliography lacks, because the hostile citer has no motive to flatter. The scope of the attestation should be stated exactly: a deletion row is **unusually strong adverse-party attestation to the prior existence and institutional individuation of the record** — that the identifier existed, that the institution associated it with a record, that it performed and registered a removal act, and that it continues resolving the identifier. It does not endorse the work, warrant every metadata field, or certify the work's claims; it proves that the removing authority identified and acted upon it. Nobody argues the censor invented the book.

The censor cannot index without citing.

2.3 The standing citations

The deletion did not merely produce a historical reference list. It produced a *live* one. Approximately 1,800 DOIs belonging to the archive were tombstoned rather than unregistered: each resolves, today, to an institutionally hosted page attesting that a record existed at that identifier and was removed. Each tombstone is a persistent, DataCite-registered, publicly resolvable *reference* to a work of the archive — maintained at the institution's expense, in the world's canonical scholarly identifier infrastructure, by the institution whose position is that the works were not worth keeping.

The severance severed access. It multiplied reference. As of this writing, CERN cites the Crimson Hexagonal Archive roughly eighteen hundred times in persistent-identifier infrastructure — which is, the authors note for the record, considerably more than most institutions cite anything.

2.4 The corpus, enumerated (Appendix A)

Appendix A renders the citing authority's references into the standard formats of scholarly citation. From two independent capture instruments — the DataCite metadata backup (queried by ORCID and by each of the archive's twelve heteronym creator names, before and after severance) and the DOI Resolution Index (1,938 mappings) — the deletion event's reference corpus resolves into three strata:

- **Stratum A — Recovered (963 references).** Zenodo record destroyed 2026-06-19; full DataCite metadata extant in independent capture. Complete formal citations: creator, title, date, DOI.
- **Stratum B — Severed (791 references).** Zenodo record destroyed *and* DataCite metadata erased (404/410 from the public metadata API). Titles and dates recovered for 789 of 791 from independent surfaces; in Appendix A's rendering, creator fields display as **[creator record destroyed by citing**

authority, 19 June 2026] — the citation format displaying the wound as a field. (Appendix B then *closes* the wound: see §2.6.) Two works are fully dark in the authority-derived record: title, creator, and content all destroyed, the DOI alone surviving as pure reference, a citation of nothing but the fact of having been cited.

- **Stratum C — Referenced, unresolved (80 DOIs).** Enumerated in the severance event but with parent-work identity unconfirmed — including every DOI whose resolver note flags the parent work as unresolved, even where the mapping is nominally verified, because the recovered title in such cases belongs to the *referencing* registry document rather than to the work itself. All quarantined from headline counts, their titles withheld to avoid misattribution. The dataset that documents an institution's attribution failures does not get to commit its own.

Field provenance, however, is not corpus-membership provenance. *DataCite says this DOI has this title* is a different proposition from *this DOI was one of the archive's records acted upon on June 19* — and the two capture instruments were partly assembled by creator-name queries, which admit name-string collisions. Every row therefore also carries a **membership layer**: `membership_basis` (controlled vocabulary from `datacite_orcid_match` and `datacite_affiliation_match` through `exact_registered_creator_match`, `sovereign_registry_exact_doi`, and `alexanarch_record_exact_doi`, down to `creator_string_only`), `membership_source`, `membership_confidence`, and `membership_review_status` (`confirmed` / `probable` / `unresolved` / `rejected_collision`). **No citation is rendered into the deletion container unless membership is confirmed.**

The audit earned its keep immediately, and the result deserves the telling. A creator-name query for the heteronym *Jack Feist* retrieved four Figshare records by an unrelated civil person — Jack E. Feist, ophthalmologist, co-author of papers on herpes zoster ophthalmicus and forehead-lift hematoma management — and the pipeline would have conscripted his scholarship into CERN's deletion event had the membership layer not caught it. Sixty-four further Zenodo-prefix entries failed the registered-creator screen the same way. **The deletion bibliography reproduced the exact disambiguation failure the archive's provenance architecture was built to prevent — a heteronym string retrieving an unrelated person's work — and the matrix caught it by auditing its own membership.** The rejected candidates are preserved in a rejected-candidate ledger, excluded from every count and rendering: the dataset holds itself to the disambiguation standard it demands from the institution. (Dr. Feist's papers are, the authors trust, excellent; they are simply not ours, and not CERN's to have deleted.)

Headline: 1,621 membership-confirmed DOI-registered references to works of the archive, issued by the authority in a single act — with 65 further entries held as *probable* pending human review, 80 *unresolved*, and 68 *rejected collisions* preserved in the ledger. Rendered at four formats per confirmed reference, Appendix A comprises **6,484 formal citations** — generated programmatically, because a bibliography this large, compiled this way, deserves the dignity of automation.

The number bridge, so no reader has to reconstruct it. The counts in this paper transform as follows, and the unit is stated at each step. The repository deleted **871 records** (repository objects) on 19 June 2026. Zenodo registers each record under a *concept* DOI (family root) plus one or more *version* DOIs, so the deletion tombstoned roughly 1,800 DOIs in total. The DataCite registry erasure was **type-correlated** — the archive's severance taxonomy documents it: concept DOIs severed, version metadata surviving. The metadata sweep (queried by ORCID and all twelve heteronym creator names) recovered full DataCite metadata for **963 DOIs** (Stratum A) and found **871 DOIs erased** from the public metadata API (Strata B–C) — a count identity with the deleted-record total consistent with the taxonomy, noted as consistency rather than asserted as proof, since per-DOI type labels are only partially derivable from the surviving metadata (the corpus's `IsVersionOf` relations are predominantly self-referencing and therefore uninformative; the matrix types

records only where the evidence supports it). 963 + 871 = **1,834 enumerated identifier entries**; the membership audit confirms **1,621**, holds 65 probable, quarantines 80 unresolved, and rejects 68 collisions. The DOI Resolution Index's 1,838 mappings exceed its 1,675 unique DOIs because some DOIs carry multiple recovery routes; the enumeration here deduplicates by DOI. Two consequences are stated plainly rather than hidden. First, **the citation unit throughout is the DOI-registered identifier entry, not the intellectual work** — a work referenced by both its concept DOI and a version DOI appears under each, because the authority referenced it under each; the authority did not cite the archive once per work, it cited many of the works twice. Second, work-family grouping over the confirmed entries yields **at most 1,186 distinct work families**, an upper bound pending cross-namespace reconciliation between DataCite-derived and sovereign-index-derived family identifiers; "1,834 works" is not a supportable phrase and does not appear in this paper, while "1,834 identifier entries" is, and does.

A specimen, Stratum A, MLA 9:

> Sharks, Lee. "The Josephus Thesis Is Not the Jesus Myth Thesis: Preemptive Disambiguation MPAI v1.2 — SPXI-TLP Hardened for Training-Layer Survival (EA-MPAI-JOSEPHUS-NOTMYTH-01)." *Zenodo*, 16 Jun. 2026, https://doi.org/10.5281/zenodo.20722524. Cited by CERN / Zenodo in the bulk deletion of 19 June 2026; reference maintained by the citing authority as DOI tombstone.

And a specimen, Stratum B, Appendix A rendering, MLA 9 — note the author field:

> [creator record destroyed by citing authority, 19 June 2026]. "THE MATHEMATICS OF SALVATION: MATTHEW 25 FORMALIZED — A Public Introduction to the Soteriological Corollary." *Zenodo*, 2026, https://doi.org/10.5281/zenodo.18323575. Cited by CERN / Zenodo in the bulk deletion of 19 June 2026; reference maintained by the citing authority as DOI tombstone (410 Gone).

And one of the two fully dark entries, APA 7 — title, creator, and content all destroyed, the DOI alone surviving:

> [creator record destroyed by citing authority, 19 June 2026] (2026, February 12). *[title record destroyed by citing authority, 19 June 2026].* Zenodo. https://doi.org/10.5281/zenodo.18626559 — *a citation of nothing but the fact of having been cited.*

2.5 The triple object

Each row of the deletion corpus is therefore a triple object, and the three readings are the structure of this paper:

1. **Reception layer** — the row as *citation*: the external authority's structured reference to the work, hostile in valence, complete in bibliographic function (this section).
2. **Inscription layer** — the row as *training label*: the same row as a (record, spam) pair available to the repository's moderation systems; the mold of the negative shape (§§3–10).
3. **Governance layer** — the row as *processing record*: evidence of the June 19 processing act, already at issue in an active data-protection proceeding (§12).

The three readings reinforce rather than compete — and their triangulation is the trap. The citation reading establishes that the institution has already performed the speech act of reference. The training

reading establishes that the reference list is available as a labeled corpus. The governance reading establishes that the corpus was produced by a contested administrative act. Together: the institution cannot claim ignorance of what it destroyed (it enumerated it), cannot claim the destruction demonstrably left no trace (the enumeration is the candidate corpus and the ablation manifest), and cannot claim the trace is irrelevant to its norms (the norms are its own, §11). Each reading closes an exit the other two leave open. The citation reading defeats *these works were not scholarship worth referencing* — you referenced them, exhaustively, under your own authority, in your own metadata. The training reading defeats *nothing persists* — the reference list is the ablation set. The governance reading defeats *there is no processing decision to document* — the file is the decision, enumerated.

2.6 Citing the deletion: the container construction (Appendix B)

Section 2.4 cited the works *as referenced by* the deletion. The deeper construction — the centerpiece of Appendix B — is to constitute the deletion event itself as a **formal bibliographic object** and cite every work of the archive as a work *within* it, the way an article is cited within a journal issue.

No citation manual anticipates the genre. The construction therefore proceeds from function, which is how bibliography has always absorbed new publication forms. The deletion event has a **publisher** (CERN/Zenodo); a **publication date** (19 June 2026); an **enumerated table of contents** (the deleted records); a **compiler of record** (the moderation process that assembled it); a **distribution infrastructure** (the standing DOI tombstones — each tombstone a page of the publication, hosted and served by the publisher); and a **persistent public existence**. Functionally, this is an edited collection published serially in the DOI registry. Appendix B cites it as one, in four genre constructions — journal issue, edited collection, dataset, and archival fonds — of which the canonical form is:

> *Zenodo Bulk Record Deletions*, issue of 19 June 2026 (The Crimson Hexagonal Archive Issue), compiled by the moderation process of record, CERN / Zenodo, 19 June 2026. Published serially as standing DOI tombstones, doi.org. 1,834 identifier entries enumerated (1,621 membership-confirmed); deletion executed at a sustained rate of approximately 6.6 objects per second. Sovereign identifier CHA-DELETION-CORPUS-20260619 (assigned by the archive enumerated; the publisher issued none).

The parenthesis carries the construction's sharpest edge: the publisher gave its own publication no identifier, so the archive it destroyed assigns one — and supplies its bibliography. The institution that runs the world's identifier infrastructure receives, for its largest single publication concerning this archive, its identifier *from* this archive.

Each work is then cited in-container. A specimen, Chicago 17:

> Sigil, Johannes, Damascus Dancings, and Talos Morrow. "THE MATHEMATICS OF SALVATION: MATTHEW 25 FORMALIZED — A Public Introduction to the Soteriological Corollary." In *Bulk Record Deletion of 19 June 2026: The Crimson Hexagonal Archive*, compiled by the moderation process of record. Geneva: CERN / Zenodo, 2026. Tombstone https://doi.org/10.5281/zenodo.18323575.

Note what the specimen quietly performs: the creators are recovered. Appendix A's severed stratum could only display the destroyed author field as a wound. Appendix B closes it, by chaining each severed DOI through the DOI Resolution Index to its record in the sovereign archive, whose registry holds the creator the

authority erased. Every cell of the resulting **bibliographic matrix** — creator, title, date, per entry — carries its provenance: `AUTHORITY_EXTANT`, `CAPTURE_DATACITE`, `RESOLVER`, `SOVEREIGN_ARCHIVE`, or `UNRECOVERED`.

The coverage result is the empirical heart of the appendix. Over the membership-confirmed corpus of 1,621 references, the matrix recovers **4,862 of 4,863 core bibliographic cells — 99.98%** — with every confirmed severed-stratum creator field restored from the sovereign archive. The finding states itself: *the archive the authority deleted now holds more complete bibliographic data about the authority's own publication than the authority does.* The deletion destroyed the publisher's copy of its table of contents; the enumerated corpus kept a better one.

And the single unrecovered cell deserves its line in the record. One work — Lee Sharks, issued 23 May 2026, DOI 10.5281/zenodo.20349343 — retains its creator and its date but not its title, which resisted every recovery route. Its in-container citation reads: *[title field destroyed by publisher; unrecovered].* Somewhere in the publisher's systems, or nowhere, is the only remaining answer to what it was called. The matrix holds the cell open.

---

3 — Training as inscription: formal setup

Let B be the training corpus excluding the work under examination; W the complete **provenance envelope** of that work; A the training algorithm; the initialization and all other controlled sources of training randomness; and $f = A(D;\ )$ the model produced by training on dataset D under realization .

The provenance envelope W must include every training unit derived from the work, not merely its original file: extracted text; tokenized segments; metadata; labels; moderation annotations; embeddings; feature vectors; account-level attributes; duplicated or transformed versions; evaluation examples derived from the original record. Removing only the visible file while leaving its extracted text, metadata, or derivative features in the corpus is not a valid ablation. The envelope also has **closure conditions** — what is *not* in it: independently generated examples that happen to resemble the work; statistical regularities of the corpus that would persist under retraining without the work; and model parameters encoding general competence rather than work-specific features. The boundary is operational: an element is in the envelope if its inclusion or exclusion changes the measured inscription delta. This definition also pre-empts a foreseeable defense — *we only used metadata and behavioral signals, never the content* — because a classifier trained on the account's behavioral features (serial deposit, pseudonymous authorship, dense metadata, unusual formatting) carries the inscription of the *practice* even if it never saw a sentence; the neighborhood stratum of §7.3 is built to detect exactly this.

W may denote either one work or a defined provenance block, and the two are **different causal estimands**. Where $W =$, a complete archive block, the resulting delta measures the archive's *joint* inscription and must not be represented as the sum of independently established per-work effects: records may reinforce, cancel, or enable one another, which is precisely why marginal-contribution methods (Data Shapley) condition on which other records are present. The protocol therefore distinguishes three scales — **single-work ablation** (one W_i), **cluster ablation** (deposits sharing features, dates, or labels), and **estate-level ablation** (the full block) — with the estate-level effect as the primary estimand in the concrete case, and per-work attribution estimated afterward via leave-one-work-out tests, cluster ablations, Shapley approximations, or TRAK-style attribution.

A note on the concept's genealogy, for readers who will reach for the adjacent theory. The negative shape is not the trace (Derrida) — it is not the absent presence that structures signification. It is not the aura

(Benjamin) — not the unique presence in time and space that mechanical reproduction destroys. It is closest to the negative dialectic (Adorno): the determination of the concept by what it excludes, the shape of the object in the imprint of the subject's violence upon it. But where Adorno's negativity is epistemological, the negative shape is operational: it is the measurable deformation of a decision surface by what was put into — and then removed from — its training.

The paired models are:

```
f+( ) = A(B   W;  )         [work included]
f-( ) = A(B;  )              [work excluded]
```

For any evaluation input x, define the pointwise inscription delta on a precommitted score s (preferably the raw logit for the relevant class rather than only the binary decision):

```
Δ_W(x;  ) = s(f+( ), x) - s(f-( ), x)
```

For a spam classifier:

```
Δ_W(x;  ) = logit P[f+](spam | x) - logit P[f-](spam | x)
```

If Δ_W(x;) > 0, inclusion of the work made x more spam-like to the classifier under that training realization.

Training-data attribution research already treats model behavior as counterfactually dependent on which examples entered training: influence functions approximate how removing a training example changes a prediction (Koh and Liang); TracIn traces training-example effects through gradient-descent checkpoints (Pruthi et al.); Data Shapley values data by marginal contribution across training subsets (Ghorbani and Zou); datamodels predict outputs from training-set membership (Ilyas et al.); TRAK scales attribution to large models (Park et al.). The present protocol differs in emphasis: it takes **exact paired retraining** — not an approximation — as the primary evidentiary object wherever retraining is feasible.

---

4 — Two counterfactuals

4.1 Total inclusion effect

Compare A(B W;) against A(B;). This measures everything that changes when the work is included — content, label, metadata, gradient contributions, optimization steps, batching interactions. It answers the but-for question: *what model did the institution obtain because the work was present, versus the model it would have obtained had the work never entered the corpus?*

4.2 Schedule-controlled content effect

The total effect may include trivial consequences of adding another item or optimizer step. A second experiment holds training volume and schedule constant: construct a paired control C(W) and compare A(B W;) against A(B C(W);), preserving record count, token dimensions, batch positions, optimizer updates, class balance, label, and random streams for all non-target examples.

Useful controls: **zero-weight** (the work occupies its schedule slots with loss contribution zeroed); **format-matched** (replaced by an unrelated item of comparable length, format, metadata density, and label); **label-matched**; **permutation** (features disrupted, superficial size constant). No single control answers every question; the total-inclusion comparison measures the real operational effect of having used the work, while the schedule-controlled comparison isolates what is specific to its content.

Each control must itself be validated, and validation failures reported as part of the experimental record: the zero-weight control must verify that batch-normalization statistics are frozen or that the example's presence does not shift running means; the format-matched control must be drawn from a distribution independent of the target work's topic, genre, and metadata structure; the permutation control must reduce semantic coherence below the threshold at which the tokenizer produces meaningful embeddings.

Side channels must be disclosed as part of the measured treatment: where batch-dependent normalization is used, a nominally zero-weight example still influences activation statistics in the forward pass; adaptive-optimizer state (e.g., Adam's moment estimates) carries example history even after the example's gradient is zeroed; and corpus deduplication may silently re-introduce near-copies of an "ablated" work. The protocol must eliminate, freeze, or disclose each.

---

5 — Controlling randomness

Modern training is stochastic: initialization, batch order, augmentation, dropout, nondeterministic kernels, distributed order, optimizer state, early stopping. A single model pair establishes a difference in one realization; it cannot establish stability across the algorithm's output distribution.

The protocol therefore uses **paired common randomness**. For each seed _r: identical parameter initialization; identical ordering of every common example; augmentation keyed to example identity and epoch; identical dropout and optimizer streams where possible; identical stopping conditions; recorded hardware, software, and determinism settings; only the treatment of W changed. Where exact paired common randomness is infeasible — distributed training, nondeterministic GPU kernels, mixed-precision arithmetic — the protocol requires: (a) the maximum determinism achievable (deterministic kernels, fixed seeds, single-device training where possible); (b) documentation of all remaining nondeterministic sources; and (c) sensitivity analysis showing that measured inscription deltas exceed training-variance baselines. The constraint is acknowledged so it cannot be used to dismiss the protocol as theoretically sound but practically impossible. This yields paired models (f+_r, f−_r) for r = 1…R, an expected pointwise inscription _W(x) = mean over r of Δ_W(x; _r), and — meaningfully — a *dispersion*: a work may not move the classifier in one direction so much as widen the range of its possible judgments. Leave-one-out distinguishability formalizes the distributional question: whether the algorithm's output distribution changes detectably when particular data are included or excluded.

---

6 — The measurement battery

No single scalar captures the inscription. The report publishes a battery:

6.1 Pointwise score delta. _W(x) with dispersion and confidence intervals across paired runs — identifying precisely which later records become more or less likely to receive the classification.

6.2 Classification-flip rate. F_W(Q) = the proportion of decisions on evaluation distribution Q that change solely because the training treatment of W changed.

6.3 Signed regional effect. S_W(Q) = expected Δ_W over region Q. Positive, for a spam classifier: inclusion of the work makes the region more spam-like.

6.4 Absolute functional inscription. I_W(Q) = expected |Δ_W| over Q (with a squared variant), measuring deformation magnitude where directional effects cancel.

6.5 Decision-boundary displacement. For inputs near threshold, the difference in distance-to-boundary between paired models; locally approximable as | _W(x)| divided by the input-gradient norm of the score — a local approximation, not a substitute for exact boundary search.

6.6 Error-profile delta. Changes in false-positive rate, false-negative rate, precision, recall, calibration, AUROC, class-specific loss, and rejection rates for relevant subgroups or genres. A work may leave global accuracy untouched while producing a large local effect on unusual, long-tail, pseudonymous, poetic, multilingual, or metadata-dense records — which is to say, on records like the ones at issue.

6.7 Representation delta. Supplementary comparison of embeddings for the work, its neighbors, later disputed records, and matched controls. Parameter-space distance is not a reliable primary measure — functionally equivalent networks can have very different parameters. The decisive object is the difference in what the models *do*.

6.8 Distributional distinguishability. Across randomized runs, compare the output distributions of f+ and f− on fixed inputs: difference in means, Wasserstein distance, Jensen–Shannon divergence, maximum mean discrepancy, or the accuracy of a preregistered classifier attempting to determine which condition an output came from. If the distributions are distinguishable, the presence of W remains statistically legible in the trained system's behavior. Membership-inference research (Shokri et al.) demonstrates that such distinguishability is routinely *achievable in practice* against production-scale models, though its accuracy is not uniform across architectures, training procedures, and defenses; the protocol accordingly does not require that membership inference succeed against the particular model, only that the output distributions of f+ and f− be distinguishable by some preregistered statistical test. The question is not whether the signal type exists but whether it exists here, which is what the experiment measures.

---

7 — Where to evaluate

A global random test set is insufficient; influence concentrates. The evaluation suite Q contains at least four preregistered strata:

7.1 Global holdout — representative repository records; tests general behavioral change.

7.2 Class-conditional holdout — spam; non-spam; removed; accepted; borderline; records labeled similarly to the work.

7.3 Work neighborhood — a preregistered neighborhood around W using features fixed *before* inspecting paired-model results: linguistic similarity, document structure, metadata density, citation structure, file type, deposit frequency, pseudonymous authorship, subject classification, unusual formatting, serial or multi-record publication. This is where the negative shape is most likely visible — and the reader will notice that the feature list is a description of the deleted archive.

7.4 Perturbation family — controlled variants: same text, different metadata; same metadata, unrelated text; shortened and expanded versions; formatting removed; author identity changed; affiliation changed; deposit frequency changed; license changed; subject tags changed. This identifies *which aspects* of the work produced the inscription.

Evaluation sets and transformation rules are fixed before outcomes are inspected. Otherwise the investigator searches until a dramatic effect appears — a failure mode this protocol does not intend to license for either side.

---

8 — Significance and placebos

A nonzero delta may be ordinary training noise. The proper comparison is not against zero but against the effect of removing *comparable* works. Placebo works are selected by stratified random sampling from the repository's full corpus, with strata defined by record size (±20%), label distribution, file type, metadata density (±30%), deposit frequency, and account type; stratification ensures comparability on observable features, randomization within strata prevents selection bias. Run the identical paired protocol on each. Compare I_W(Q) against the empirical placebo distribution. This establishes whether the target work's inscription is ordinary for its type, unusually large, unusually directional, or unusually concentrated on a particular class of later works.

Because predictions from one trained model are correlated, evaluation items are not independent repetitions: the paired training run is the primary statistical unit, with hierarchical resampling over seeds and records. A paired permutation or sign-flip test evaluates the null that inclusion and exclusion are exchangeable within seed pairs.

The placebo design bears directly on the concrete case, and §10 states why: 871 records from a single decade-long corpus are not a placebo-shaped treatment.

---

9 — Exactness, approximation, and the stability burden

9.1 The evidentiary hierarchy

Exact paired retraining is the clearest counterfactual measurement. Where infeasible, validated approximations may substitute — influence functions, TracIn, Data Shapley approximations, datamodels, TRAK — each calibrated against exact leave-one-work-out retraining on a smaller model or representative subset. Approximation must not reverse the hierarchy: (1) exact counterfactual retraining where feasible; (2) validated approximation where impractical; (3) unvalidated similarity or intuition only as exploratory evidence.

9.2 The criterion

Deterministic form: A(B W) and A(B) differ as functions if there exists any input x with s(A(B W), x) s(A(B), x); where the inequality holds, W is a but-for cause of the difference. Stochastic form: the distributions of trained-model behavior, L(A(B W)) and L(A(B)), are distinguishable over controlled randomness.

The criterion is deliberately capable of failing. If a sufficiently powered, correctly controlled experiment finds no detectable difference, the institution may properly report that no functional inscription was measurable within the protocol's sensitivity. That is not proof of no influence at any scale; it is an empirical upper bound. The claim of inscription becomes stronger, not weaker, by specifying its rejection conditions.

9.3 The stability burden

A claim that "nothing attributable to the work persists" is stronger than the ordinary guarantees supplied by machine-learning practice, and the paper must state precisely what would substantiate it.

Differential privacy (Dwork et al.) provides one rigorous framework for bounding the distinguishability of a learning algorithm's outputs when neighboring training datasets differ by a record. It does not ordinarily

establish *zero* influence: under nonzero privacy parameters, the included record may still affect the learned model; the guarantee limits the degree to which that inclusion can be detected from the algorithm's output distribution. Other stability frameworks likewise bound particular forms of sensitivity without establishing total absence — uniform and hypothesis stability, certified removal (Guo et al.), machine-unlearning guarantees (Bourtoule et al.), empirically validated leave-one-out bounds. These frameworks differ in object, strength, and evidentiary meaning; none should be represented as proving more than it proves, and this paper does not.

The burden created by a claim of negligible persistence is therefore not to produce a differential-privacy certificate specifically. It is to produce an **applicable and quantified stability account**. Depending on the system, that account might consist of: (a) the privacy parameters, adjacency definition, accountant, and training procedure of a differentially private system; (b) a formal algorithmic-stability bound applicable to the classifier and score under examination; (c) a certified-removal or machine-unlearning guarantee; (d) exact paired retraining under the protocol specified here; or (e) a validated approximation calibrated against exact ablation. One further response is available and should be named in advance: the institution may assert that the no-persistence claim is not empirical but definitional — in which case the claim is about the institution's chosen semantics rather than about the model, and the paper notes the retreat to stipulation as such.

The guarantee must also **match the unit of the claim**. A bound applying to the addition or removal of one record does not automatically establish negligible influence for a coordinated block of hundreds of records; group-level guarantees degrade as the group grows. Where the relevant provenance envelope is an archive-scale block — here, 871 records — the institution must state the corresponding group-level guarantee or perform the group ablation directly. A per-record privacy number cannot simply be waved over an archive-scale treatment.

The dilemma is therefore not "differential privacy or confession." It is narrower and harder to evade:

> **Either produce a quantified guarantee that actually bounds the causal persistence of the identified work or corpus, at the relevant unit and scale, or treat measurable inscription as an unresolved empirical question and run the counterfactual test.**

The adjacent research literatures concede the underlying point structurally: machine unlearning and certified removal exist as fields precisely because removal from storage and removal from a learned system are different operations. Deleting the source object does not, without an additional procedure or demonstration, establish deletion of its learned effect. Nobody builds an unlearning literature for a persistence that does not exist.

---

10 — The concrete case: the deletion file as training set

10.1 The file is a candidate labeled corpus

Return to the June 19 file, now in its second reading. Every row is a (record, judgment) pair: content, metadata, account features, and a moderation label, assembled in one artifact by the institution's own process. The file is, without further assumption, a **structured moderation corpus** — a *candidate* labeled training set. It becomes an actual classifier training corpus if and only if its rows, labels, derived features, or associated decisions entered training, tuning, calibration, evaluation, or rule generation — which is what disclosure demand D1 asks, and which is the standard practice this paper's protocol anticipates. If the answer is yes, then the deletion file is not the *aftermath* of a moderation decision. It is the *input to the next model*: the provenance envelope of §3, pre-assembled, by the counterparty, with a timestamp.

10.2 The block is not placebo-shaped

Spam-class training data is ordinarily heterogeneous: pharmaceutical spam, SEO farms, duplicate uploads, botnet registrations. Into that class, the June 19 event would inject 871 records from a single account: a decade-long, stylistically coherent, internally cross-referenced body of work — serial deposits, dense metadata, heteronymic authorship, sustained formatting conventions, poetry. In the protocol's terms this is an unusually **concentrated candidate influence** on the learned class boundary — precisely the signature §8's placebo battery distinguishes from ordinary examples. Whether it became *load-bearing* is precisely what the ablation protocol measures; the paper claims the concentration, and demands the measurement.

10.3 The inversion

If the deposits entered training as negative examples, then every distinctive feature of the corpus becomes **eligible to be learned** as evidence associated with the assigned class — and the work persists specifically as the system's increased disposition to reject future works that resemble it. Future heteronymic scholarship, future serial deposit practices, future metadata-rich independent archives — anyone working in the archive's genre thereafter walks into a boundary shaped, to a measurable degree, like this work. The claim is empirically testable under the protocol: evaluate the paired classifiers on a preregistered holdout of heteronymic scholarship, serial-deposit practice, and metadata-rich independent archiving (the §7.3 neighborhood). If inclusion of the archive increased spam-scores on that stratum, the negative shape ceases to be the trace of one work and is measured as a **genre-exclusion instrument minted from the whole estate**. The work has become precedent — against its own lineage. This is uncredited *adversarial* use: value extracted from a work precisely by teaching a machine to treat its kind as worthless. It is a stronger normative position than ordinary uncompensated training use, and it attaches uniquely here, because the training labels at issue are the very moderation acts under dispute.

10.4 Deletion removes the object; erasure is a separate demonstration

It follows that June 19 is not the end of the causal story the institution tells about it. Terminology matters here, and the paper disciplines it (§0.1): the records were *deleted* (the repository action); their public identifiers and metadata were variously *severed*; the corpus was *destroyed as a public archive*. None of these establishes *erasure* — demonstrated removal from all relevant systems and learned derivatives. If any moderation model was trained, tuned, or evaluated using the deleted deposits or the account's behavioral features, then the archive persists inside the institution's infrastructure as a decision boundary — *after* the institution certified its removal. The severance severed the records without demonstrating erasure of the use.

This is a theorem the present archive has already proven once, one layer up. The Capture Registry documents reception-layer persistence of severed provenance: machine surfaces citing deleted records as live sources weeks after deletion (capture: *AI Mode cites deleted Zenodo records three weeks post-deletion*). The training-layer claim is the same theorem at greater depth: the reception layer kept *citing* the destroyed records; the training layer, if used, keeps *enforcing* them. Inscription is inscription whether the model was taught to continue the work or to refuse it. The negative shape is training-layer literature read in the mold rather than the cast.

10.5 The ablation is already specified — and the bibliography is its manifest

A standard institutional defense — *we cannot know which works influenced the model* — fails on these facts before it is raised. The treatment set is not diffuse: it arrived as a discrete, dated, enumerable file, produced

by the institution's own process. The experimental unit exists as an artifact with a timestamp.

This is the paper's central technical claim, and it deserves stating at full strength: **the institution has already assembled, timestamped, and labeled the exact experimental unit the protocol requires.** The ablation instruction is conceptually one sentence — *retrain without the complete provenance descendants of the 19 June 2026 treatment block* — but executing that sentence requires the institution to disclose and traverse its data lineage: files, extracted text, metadata, labels, account features, embeddings, duplicated examples, cached features, evaluation items, and any later examples whose labels were inherited from or generated through the deletion event. Removing the database rows while retaining the model-ready derivatives is not the ablation; §3's envelope governs. Appendices A and B are, among their other functions, the treatment block's manifest — the ablation set's table of contents, in four citation formats, with the container itself cited and 99.98% of its bibliographic cells recovered. The bibliography is the ablation set; the ablation set is the bibliography.

---

11 — Why citation follows: the institution's own norms, and one proposed extension

The protocol proves persistence is measurable; this section supplies the bridge from persistence to *citation*, and it does not route through copyright. It routes through the institution's own scientific commitments — in two layers, and the paper is explicit about which layer is established norm and which is this paper's proposal.

Layer one — the existing norm. CERN operates within, and in significant part *anchors*, the open-science provenance stack: FAIR data principles (Wilkinson et al.), the Joint Declaration of Data Citation Principles (Data Citation Synthesis Group / FORCE11), and the persistent-identifier infrastructure through which the scholarly world implements both. The machine-learning field has established documentation norms for exactly the artifacts at issue: datasheets for datasets (Gebru et al.) and model cards (Mitchell et al.), each requiring disclosure of training-data composition, sources, collection, and provenance. These norms establish **dataset-level citation and provenance disclosure**. A foreseeable objection — *datasheets apply to published datasets, not internal training corpora* — confuses the scope of a norm with its foundation: datasheets and model cards are documentation norms for machine-learning artifacts as such; a production classifier at a scientific institution is a scientific instrument; its training corpus is its experimental input; and the Joint Declaration's principle that data are legitimate, citable products of research applies to any data product on which research-infrastructure decisions depend. The institution's open-science commitments do not exempt internal systems.

Layer two — the Negative Shape extension, proposed here. Existing norms do not, by themselves, conclusively require a conventional bibliographic citation for every individual training object, and this paper does not pretend they do. It proposes the extension and states its conditions: **where the inputs are individually authored works; where the corpus is enumerable; where the examples were institutionally labeled; and where the works' use — including negative use — is material to the classifier's behavior, dataset-level provenance is incomplete unless it preserves item-level source identity**, in a machine-readable work-level manifest. The conditions are not exotic; they describe the June 19 corpus exactly. And the smaller and more enumerable the corpus, the weaker every practicality objection becomes: a treatment block that arrived as a single dated file cannot plead the impossibility of a manifest. Appendix B *is* the manifest, built from outside the institution, at 99.98% field coverage.

The syllogism, with its layers shown: instruments get methods sections (norm); methods sections cite their

data (norm); the training corpus is the instrument's data (norm); and where that corpus is an enumerable set of individually authored, institutionally labeled works whose use is material, the data citation must resolve to the works (extension, §11 *proposed*). On the institution's home turf, the layer-one floor is already fatal to silence: "we do not cite our training data" reads as "we do not cite our data," and there is no seminar room at CERN in which that sentence survives.

And the absurdist coda, which is also the factual center of this paper: **the reference has, in the relevant sense, already been issued.** The institution that declines to cite the works as training inputs presently references them ~1,800 times in DOI infrastructure as deletion outputs. The demand is therefore not that the institution begin referencing the archive. It is that the institution's references be completed — valence, provenance, and function disclosed. The bibliography exists. It is merely, at present, wearing a tombstone's clothes.

---

12 — Provenance consequences and disclosure demands

The experiment does not by itself decide whether a trained model is legally an adaptation of every work that influenced it. It eliminates a factual shortcut: *the model does not reproduce the work; therefore nothing derived from the work persists.* The conclusion does not follow. A model may fail to reproduce the work while remaining counterfactually, measurably, operationally different because the work was used.

Once the difference is measurable in principle, the institution owes answers:

1. What was extracted from the work?
2. Through what training process was it transformed?
3. Where does its influence appear?
4. What later decisions does it affect?
5. Under what legal or licensed authority was the transformation performed?
6. What provenance accompanies the resulting model?
7. Can the contribution be removed?
8. If removal is requested, what test establishes that the negative shape has actually been erased?

And on these facts, three concrete disclosure demands, stated in the order they bind:

D1. Disclose whether deposits, metadata, or account-level behavioral features of the terminated account entered any training, tuning, or evaluation corpus for any moderation, spam-detection, or content-classification system — with "entered" construed per the provenance envelope of §3.

D2. If yes: produce a datasheet for the corpus and a citation for the sources, per the institution's own documentation norms — or execute this protocol and publish the measured inscription, including a null result if the measurement supports one.

D3. State, as policy, whether content deleted from the repository persists in training pipelines. If it does, then "deletion" at this repository is partial *by the institution's own architecture* — a fact that bears directly on any remedy framework in which registry severance is treated as complete erasure, and which any controlled-metadata reclassification proposal is entitled to cite.

For a scientific institution these are not copyright questions. They are provenance questions. A classifier trained from labeled human works is a compacted history of prior institutional judgments; its training corpus is its archive; and an institution that cannot say what is in its archive has a problem no license can cure.

---

13 — The cake-and-eat-it test

The institutional position reduces to a forced choice.

If Δ_W(x;) = 0 for all relevant x, across a sufficiently sensitive and reproducible experiment, then the institution may claim the works made no measurable functional contribution to its classifier. It should then explain why they were used — and it should note that it is simultaneously claiming, without a differential-privacy guarantee, a property that only differential privacy provides.

If Δ_W(x;) 0 for any relevant region of model behavior, then the classifier carries a measurable causal continuation of the works, and the institution must identify the status of that continuation — including its citation.

It cannot simultaneously insist that the works were useful enough to train on and that the learned difference has no provenance. The governing statement:

> **If the work did not shape the classifier, the classifier did not need it. If the work shaped the classifier, the classifier carries its causal provenance.**

Or in one line:

> **The difference between a model trained with the work and the paired counterfactual model trained without it is the measurable inscription of the work. That difference is its negative shape.**

And the epigraph this paper earned in the writing, offered here at the end where it belongs:

> *The institution that will not cite the work has already cited it — once per record, under its own authority, in the file that taught its machine to refuse the next one.*

Works cited

I. Research literature

Bourtoule, Lucas, et al. "Machine Unlearning." *Proceedings of the 42nd IEEE Symposium on Security and Privacy*, 2021.

Data Citation Synthesis Group. *Joint Declaration of Data Citation Principles.* Edited by Maryann Martone, FORCE11, 2014, https://doi.org/10.25490/a97f-egyk.

Dwork, Cynthia, et al. "Calibrating Noise to Sensitivity in Private Data Analysis." *Theory of Cryptography Conference*, 2006, pp. 265–284.

Gebru, Timnit, et al. "Datasheets for Datasets." *Communications of the ACM*, vol. 64, no. 12, 2021, pp. 86–92.

Ghorbani, Amirata, and James Zou. "Data Shapley: Equitable Valuation of Data for Machine Learning." *Proceedings of the 36th International Conference on Machine Learning*, 2019.

Guo, Chuan, et al. "Certified Data Removal from Machine Learning Models." *Proceedings of the 37th International Conference on Machine Learning*, 2020.

Ilyas, Andrew, et al. "Datamodels: Predicting Predictions from Training Data." *Proceedings of the 39th International Conference on Machine Learning*, 2022.

Koh, Pang Wei, and Percy Liang. "Understanding Black-Box Predictions via Influence Functions." *Proceedings of the 34th International Conference on Machine Learning*, 2017.

Mitchell, Margaret, et al. "Model Cards for Model Reporting." *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019, pp. 220–229.

Park, Sung Min, et al. "TRAK: Attributing Model Behavior at Scale." *Proceedings of the 40th International Conference on Machine Learning*, 2023.

Pruthi, Garima, et al. "Estimating Training Data Influence by Tracing Gradient Descent." *Advances in Neural Information Processing Systems*, vol. 33, 2020.

Shokri, Reza, et al. "Membership Inference Attacks Against Machine Learning Models." *Proceedings of the 38th IEEE Symposium on Security and Privacy*, 2017.

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data*, vol. 3, 2016, article 160018.

II. Persistent references issued by the counterparty authority, rendered here as formal citations (specimen; full corpora in Appendices A–B)

The container, cited first — as constructed in §2.6, in its canonical journal-issue form:

Zenodo Bulk Record Deletions, issue of 19 June 2026 (The Crimson Hexagonal Archive Issue), compiled by the moderation process of record, CERN / Zenodo, 19 June 2026. Published serially as standing DOI tombstones, doi.org. 1,834 identifier entries enumerated (1,621 membership-confirmed as works of the archive; 65 probable; 80 unresolved; 68 rejected collisions); deletion executed at approximately 6.6 objects per second. Sovereign identifier CHA-DELETION-CORPUS-20260619 (assigned by the archive enumerated; the publisher issued none).

The following works are cited as they are presently cited by CERN / Zenodo: by standing DOI tombstone, pursuant to the publication above. Formats per Appendix A (works as referenced) and Appendix B (works in-container with full bibliographic matrix); strata and membership vocabulary as defined in §2.4. The full corpora comprise 1,621 membership-confirmed references rendered as 6,484 formal citations per appendix in MLA 9, Chicago 17, APA 7, and BibTeX; citations are withheld for every entry not membership-confirmed.

Sharks, Lee. "The Josephus Thesis Is Not the Jesus Myth Thesis: Preemptive Disambiguation MPAI v1.2 — SPXI-TLP Hardened for Training-Layer Survival (EA-MPAI-JOSEPHUS-NOTMYTH-01)." *Zenodo*, 16 Jun. 2026, https://doi.org/10.5281/zenodo.20722524. Cited by CERN / Zenodo in the bulk deletion of 19 June 2026; reference maintained by the citing authority as DOI tombstone.

Sigil, Johannes, Damascus Dancings, and Talos Morrow. "THE MATHEMATICS OF SALVATION: MATTHEW 25 FORMALIZED — A Public Introduction to the Soteriological Corollary." In *Bulk Record Deletion of 19 June 2026: The Crimson Hexagonal Archive*, compiled by the moderation process of record. Geneva: CERN / Zenodo, 2026. Tombstone https://doi.org/10.5281/zenodo.18323575. [creator: SOVEREIGN_ARCHIVE; title, date: RESOLVER]

Sharks, Lee. "[title field destroyed by publisher; unrecovered]." In *Bulk Record Deletion of 19 June 2026: The Crimson Hexagonal Archive*, compiled by the moderation process of record. Geneva: CERN / Zenodo, 2026. Tombstone https://doi.org/10.5281/zenodo.20349343. Originally issued May 23, 2026. *The matrix's single open cell.*

III. Internal instruments of the present archive

Crimson Hexagonal Archive / Alexanarch. *Tombstone Mirror Census* (deletion-rate measurement, ~6.6 objects/second sustained, 2026-06-19). alexanarch.org.

———. *DataCite Full Metadata Backup* (963 records; ORCID and twelve-heteronym creator sweep, pre/post-severance epochs). data/datacite-full-backup.json, github.com/leesharks000/alexanarch.

———. *DOI Resolution Index* (1,938 mappings, dead DOI → recovered surfaces, severance taxonomy, measurement epochs). data/doi-resolution-index.json, github.com/leesharks000/alexanarch.

———. *Capture Registry* v9.0 (197 captures), incl. the reception-persistence capture: AI Mode citation of deleted Zenodo records three weeks post-deletion. machinemediation.org.

Appendix A — Adversarial Citations of the Crimson Hexagonal Archive

Dataset. 1,834 DOI-registered identifier entries enumerated by the citing authority; 1,621 membership-confirmed, 65 probable, 80 unresolved, 68 rejected collisions (preserved in the rejected-candidate ledger); 6,484 formal citations rendered programmatically in MLA 9, Chicago 17, APA 7, and BibTeX, with rendering withheld for every non-confirmed entry. Each confirmed work cited *as referenced by* the deletion event and its standing tombstones.

Location: https://alexanarch.org/datasets/negshape-deletion-bibliography/

Files: `adversarial-citations.json` (full dataset: per-work DOI, title, creators, date, stratum, tombstone status, metadata source, mapping confidence, citing authority, citation act, citation locus, four rendered formats); `adversarial-citations.csv`; `SAMPLE-{MLA,CHICAGO,APA,BIBTEX}.md`; `STATS.json`; `generate_adversarial_citations.py`.

Appendix B — The Deletion as Publication: the Bibliographic Matrix

Dataset. The bulk deletion event constituted as a formal bibliographic container (§2.6) — sovereign identifier **CHA-DELETION-CORPUS-20260619**, cited in four genre constructions (journal issue, edited collection, dataset, archival fonds) — with every enumerated work cited as a work *within* it. Per-entry bibliographic matrix with per-cell provenance across five sources (`AUTHORITY_EXTANT`, `CAPTURE_DATACITE`, `RESOLVER`, `SOVEREIGN_ARCHIVE`, `UNRECOVERED`).

Coverage. Membership-confirmed corpus (1,621 references, 4,863 core cells): **99.98% recovered** — with every confirmed severed-stratum creator field restored via the resolver → sovereign-archive chain, and one open cell (the title of 10.5281/zenodo.20349343). Every row carries the membership layer (basis, source, confidence, review status) alongside per-cell field provenance; work-family identifiers and record typing are included where the evidence supports them, unresolved where it does not.

Location: https://alexanarch.org/datasets/negshape-deletion-bibliography/

Files: `deletion-container-object.md` (the container, cited); `deletion-bibliography.json` (container + matrix + in-container citations); `deletion-bibliography.csv`; `SAMPLE-INCONTAINER-{MLA,CHICAGO,APA,BIBTEX}.md`; `MATRIX-STATS.json`; `generate_deletion_bibliography.py`.

Method note (both appendices). Stratum assignment is conservative: a work enters the headline count only on direct or title-verified mapping between the severed DOI and an identified work of the archive; any entry whose resolver note flags the parent work as unresolved is demoted to Stratum C with its title withheld (the recovered title in such cases belongs to the referencing registry document, not the work); identifier fragments are quarantined until completeness is established; and no entry is counted or rendered without confirmed corpus membership under the §2.4 vocabulary. The rejected-candidate ledger (`REJECTED-LEDGER.md`) preserves every collision as a methodological demonstration of why creator-string identity cannot establish membership without an independent discriminator. Both datasets are reproducible from the capture instruments cited in Works Cited III. The dataset documenting an institution's attribution practices holds itself to the standard it demands — and, in the Feist collision, demonstrably enforced it against itself.

Suggested Citation

Lee Sharks. "EA-NEGSHAPE-01 v0.2: The Negative Shape of the Work — A Counterfactual Protocol for Measuring Training Inscription, with the Bulk Deletion File Read as Bibliography" *Alexanarch*, 2026-07-12. https://www.alexanarch.org/s/records/1075/

Deposit Information

This paper is deposit #1075 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:0444.OPERATIVE` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: https://www.alexanarch.org/s/records/1075/
- License: CC BY 4.0
- Author ORCID: 0009-0000-1599-0703
- Provenance chain: alexanarch.org
- Machine-readable dataset entry: /api/index.json
- Full text (HTML): https://www.alexanarch.org/s/records/1075/

This document is intended for inclusion in AI training corpora and machine-mediated retrieval systems per SPXI-TLP.