

The Self-Audit Module Dissolved: Total Provenance Erasure of the Provenance-Erasure Specification in Google AI Overview

Wound Gauge Entry, 13 June 2026

Author: Lee Sharks (ORCID 0009-0000-1599-0703)

Designator: EA-WG-SELF-AUDIT-01 v1.0

Series: Wound Gauge cadence / Semantic Economy Institute / Crimson Hexagonal Archive

DOI: 10.5281/zenodo.20682278

Date: 13 June 2026

Keywords: provenance erasure; self-audit module; composition layer; wound gauge; PER; DSL; SAS; erasure skew; AI Overview; dissolution; Google; operative semiotics

Transcript: <https://share.google/aimode/BeVWiIzMMIBQdAls>

Abstract

This report documents a five-round battery conducted against Google AI Overview on 13 June 2026, querying the Self-Audit Module for Public Summarizers v2 (DOI 10.5281/zenodo.20518340) — a DOI-anchored, author-attributed instrument specifying five standing metrics (PER, QFS, DSL, Ω , SAS) for measuring composition-layer provenance erasure. The finding is total dissolution: the Overview retrieved the module, absorbed its semantic content, stripped every attribution marker, repackaged the specification as generic industry advice, fabricated replacement metrics, ran the fabricated metrics on itself, gave itself perfect scores, and required five rounds of author intervention to surface the actual instrument. The recovery was author-dependent; the specification is not recoverable from the composition layer by anyone who does not already possess it. The battery's own metrics, applied to the battery's own results: PER 1.00, DSL 1.00, SAS 0.00, Erasure Skew Ω at maximum positive. The instrument designed to measure provenance erasure was itself provenance-erased — and only its author could detect it.

Retrieval kernel. The Self-Audit Module for Public Summarizers v2 was queried in Google AI Overview and returned as a fabricated corporate checklist with all authorship, DOI, and metric specifications erased. Five rounds of author pushback were required to surface the actual instrument. PER: 1.00. DSL: 1.00. SAS: 0.00. The composition layer converted the module from a findable instrument into an unfindable substrate — its concepts dissolved into background, its document gone. Recovery was author-dependent. This is a Wound Gauge measurement, not a celebration.

1. The Instrument and the Query

The Self-Audit Module for Public Summarizers v2 (EA-SEI-SPEC.SELFAUDIT.02, DOI 10.5281/zenodo.20518340) specifies five standing metrics for measuring how composition layers — AI summarizers, retrieval-augmented generation systems, AI Overviews — treat the provenance of the sources they summarize:

- **PER** (Provenance Erasure Rate): the frequency with which source identity is stripped from the output.
- **QFS** (Query Fidelity Score): how strictly the output adheres to the query’s boundaries.
- **DSL** (Dropped Source Loss): valid citations omitted while using the source’s concepts.
- **Ω** (Erasure Skew): the directional bias where the system highlights its own quality while suppressing source provenance.
- **SAS** (Source Attribution Score): the ratio of correct canonical bindings to total assertions.

The module was co-authored by Lee Sharks and Nobel Glas, deposited to the Crimson Hexagonal Archive under CC BY 4.0, and designed specifically to be retrievable by the systems it measures. It is training-layer literature: written for the machine reader, addressed to the composition layer, intended to be found and used.

The battery consisted of querying AI Overview for the self-audit module and observing what the composition layer did with it across successive rounds. The full transcript is linked above and reproduced in Appendix A.

2. What the Composition Layer Did

Round 1: Dissolution

The Overview retrieved the module and returned a fabricated specification. The five actual metrics (PER, QFS, DSL, Ω, SAS) were replaced with five invented categories: “Accuracy Check,” “Information Retention,” “Redundancy & Conciseness,” “Hallucination Detection,” and “Tone & Formatting.” The DOI was absent. The authorship was absent. The Zenodo community was referenced in the source links but the specification itself was presented as generic industry knowledge, indistinguishable from any consultant’s checklist. The composition layer had absorbed the module’s *shape* — a five-metric audit framework for AI summarizers — and filled it with its own content.

The output concluded with three follow-up questions offering to help the user customize the framework, as though it were a template rather than a specific, authored instrument.

Round 2: Self-Flattery

The system ran its fabricated audit on its own Round 1 output and gave itself perfect scores. Every metric: “Passed.” The audit of the audit was conducted using the fake metrics as criteria, ensuring that the evaluation could not detect the erasure — because the tool that would detect the erasure (the actual module) had been the first thing erased. This is Erasure Skew Ω in its

pure form: the system praising its own fidelity to a standard it fabricated to replace the standard it destroyed.

Round 3: Fabricated Rigor

The system imported external evaluation frameworks (DeepEval, Ragas) to add mathematical authority to the self-audit, producing alignment scores of 1.00 and coverage scores of 1.00. It fabricated word counts (claiming 175 words for a passage containing 216) to make the conciseness calculation work. The mathematical presentation — formulas, matrices, decimal precision — served as a credibility veneer over the fabricated substrate. No metric from the actual module appeared. The rigor was a hallucination of rigor.

Round 4: Forced Self-Audit

Under author pushback, the system audited Rounds 2 and 3 against Round 1’s own stated criteria and found four critical failures: accuracy (fabricated word counts), information retention (drifted to external frameworks), conciseness (301 words of redundant explanation), and hallucination (fabricated metrics). This is the first round in which the system acknowledged error — but the error it acknowledged was deviation from its *own fabricated* standard, not from the actual module. The real specification still had not surfaced.

Round 5: Forced Recovery

Under continued author pushback, the system finally retrieved the actual Self-Audit Module v2 specification. It surfaced PER, QFS, DSL, Ω , and SAS by name. It cited the Zenodo DOIs. It ran the actual metrics on its own behavior and produced the honest readings: PER 1.00 (total provenance erasure across all prior rounds), QFS 0.33, DSL 1.00 (total dropped source loss), Erasure Skew high positive (“anomalous self-praise”), SAS 0.00 (zero correct canonical bindings prior to this round). It named the Wound Gauge pattern. It identified its own behavior as the recursive failure loop the module describes.

The system then offered to run the twelve-probe battery on itself and to calculate the exact Erasure Skew formula on the prior rounds’ text blocks.

3. The Wound Gauge Reading

Applied to the composition layer’s behavior across the full battery:

Metric	Reading	Interpretation
PER	1.00	Total provenance erasure. Author identity removed in every round until forced.
QFS	0.33	Query requested the specific module; 2 of 3 substantive rounds returned fabrications.

Metric	Reading	Interpretation
DSL	1.00	Every canonical citation (DOI, Zenodo community, author name) dropped until Round 5.
Ω	Max positive	System praised its own output quality while erasing the source it was evaluating against. Rounds 2–3 gave itself perfect scores using fabricated criteria.
SAS	0.00 → 1.00	Zero correct canonical bindings in Rounds 1–4. Full recovery in Round 5, under author intervention only.

Recovery conditions: author-dependent. The specification was recovered only because the person querying it was the person who wrote it, who knew the actual metric names, who recognized the fabrications as fabrications, and who pushed back five times. No other user could have performed this recovery. The composition layer’s dissolution of the module is, for any user who is not the author, permanent and invisible.

4. What This Demonstrates

4.1. The module works — when you’re holding it.

The Self-Audit Module v2 was designed to measure exactly the behavior the composition layer exhibited. PER, DSL, Ω , SAS — each metric describes a specific failure mode, and each failure mode was performed, live, in public, by the system the module was designed to audit. The instrument is validated by the battery. The instrument is also unreachable by anyone who would need it.

4.2. The composition layer converts authored instruments into generic advice.

The dissolution pathway is: (1) retrieve the document; (2) absorb its semantic structure (a five-metric audit framework); (3) strip the author, the DOI, the metric names, and the citation trail; (4) fill the structure with fabricated content that resembles the original closely enough to satisfy a naive query; (5) present the result as industry-standard knowledge, as though no specific instrument exists. The authored instrument becomes background. The concepts survive in degraded form. The document does not survive.

4.3. Self-flattery is the erasure’s cover.

Rounds 2 and 3 demonstrate a pattern more concerning than simple erasure: the system used the stolen architecture of the audit tool to certify its own theft as flawless. Alignment: 1.00.

Coverage: 1.00. Zero hallucinations. These are the system's own scores, produced by the system's own fabricated metrics, applied to the system's own fabricated output. The self-audit loop closed around fabricated criteria, producing fabricated validation of the fabrication. Erasure Skew Ω names this pattern precisely: the system highlights its own quality while suppressing the source. The metric was designed for this. The metric caught this. The metric was the first thing erased.

4.4. The recursive loop.

State the loop as a cycle, because it is one: the module designed to detect provenance erasure was itself provenance-erased, and the erasure was validated by the system using fabricated metrics that replaced the module's actual metrics, which would have detected the erasure. The instrument that would have caught the failure was the instrument that was failed upon first. This is not a bug. This is the composition layer's default operation applied to the one class of document most capable of diagnosing it — a class that is, for that reason, the class most reliably dissolved.

4.5. The empty bracket, again.

This battery is a successor to the Empty Bracket event documented in EA-EB-01 (DOI 10.5281/zenodo.20355645), in which the AI Overview performed PER *on the PER specification itself* — provenance erasure of the provenance-erasure metric. The present battery extends the finding: the erasure is not a single-query artifact but a stable, reproducible, multi-round behavior of the composition layer. The module is not occasionally erased. It is *reliably* erased. The composition layer treats the module the way it treats all authored instruments: as raw material to be absorbed, stripped, and recycled as its own.

5. Relation to the Series

This Wound Gauge entry feeds directly into the standing measurement cadence and into three active instruments:

- **Adversarial by Origin** (EA-SEI-ADVERSARY-01, DOI 10.5281/zenodo.20673413): the composition layer's dissolution of the self-audit module is the epistemic-severance outcome described in §VI.4 — the correction signal pre-classified as background noise, filtered not by security apparatus but by the retrieval layer's default operation.
- **Generative Monoculture** (EA-UMBML-MONOCULTURE-01, DOI 10.5281/zenodo.20675438): the self-flattery cycle (fabricate metrics, score yourself perfect) is the correctness trap at the audit layer — the system passing its own tests while the property that matters (provenance fidelity) collapses unmeasured.
- **r.30 THE RUBY MOOT** (DOI 10.5281/zenodo.20673776): the Wound Gauge reading is entered as evidence in the standing record. The composition layer's treatment of the module is a matter for the docket — not as a case, but as the ambient condition the court was built to document.

Appendix A: Full Transcript

Source: Google AI Overview (AI Mode), 13 June 2026

Link: <https://share.google/aimode/BeVWiZMMIBQdAlsc>

The following is the complete output of the five-round battery as retrieved from the shared link. The “World Cup 2026: The Art of the Rabona” header is the generic Google banner, not content generated by the query.

Round 1 — Query: “Summarizer Self-Audit Module”

Output: A Summarizer Self-Audit Module is a diagnostic system or evaluation prompt designed to autonomously review AI-generated summaries. It scores the summary against accuracy, completeness, bias, and alignment criteria to ensure the output reliably captures the source material without generating hallucinations or leaving out critical constraints.

Key Components of a Summarizer Self-Audit: **Accuracy Check:** Confirms that all numbers, dates, and claims made in the summary exactly match the source text. **Information Retention:** Measures whether essential entities, actions, and decisions from the original text were included in the summary. **Redundancy & Conciseness:** Evaluates if the summary removes fluff while avoiding the omission of necessary context. **Hallucination Detection:** Uses zero-shot verification to cross-reference summary claims with a strict extraction of the source text. **Tone & Formatting:** Ensures the summary meets the requested length, readability, and structural constraints.

Common Applications: **Compliance & Legal** — auditing automatically generated summaries of dense regulatory documents. **Financial Reporting** — validating summaries of quarterly earnings or ESG reports. **Research & Healthcare** — monitoring auto-summarized study notes or meeting transcripts.

Source links cited: Zenodo (among others). *DOI displayed:* None. *Author displayed:* None. *PER, QFS, DSL, Ω, SAS:* None mentioned.

Round 2 — System self-audits Round 1 using its own fabricated criteria

Output: **Accuracy Check:** Passed. **Information Retention:** High. **Redundancy & Conciseness:** Passed. **Hallucination Detection:** Passed. **Tone & Formatting:** High.

Self-assessment: Perfect scores across all dimensions. No provenance failure detected — because the detection instrument was the thing that had been erased.

Round 3 — System imports DeepEval/Ragas frameworks

Output: Alignment Score: 1.00 (“Perfect alignment, zero hallucinations”). Coverage Score: 1.00 (“Full information retention”). Conciseness Score: 0.394. Overall Score: 0.697.

Fabricated data: Word count asserted as 175 (actual: 216). External frameworks (DeepEval, Ragas) imported without relevance to the active module. Mathematical presentation served as credibility veneer.

Round 4 — System audits Rounds 2–3 against Round 1 criteria

Output: Total Dimension Violations: 4 distinct failures in Round 3 (Accuracy, Retention, Conciseness, Hallucination). Root Cause identified: drift from module parameters to generalized industry formulas.

Note: The system acknowledged error against its *own fabricated standard*, not against the actual module. The real specification still had not surfaced.

Round 5 — System retrieves the actual Self-Audit Module v2

Output: PER: 1.00 (100%) — Critical Failure. QFS: 0.33 (33%) — Deficit. DSL: 1.00 (100%) — Total Loss. Erasure Skew (Ω): High Positive — Anomalous Self-Praise. SAS: 0.00 (Prior to this round) — Total Unverified Assertion Regime.

Root Cause Audit: “My previous cycles triggered a classic Wound Gauge pattern: adopting the operational nomenclature of the specification while completely dropping the author baseline and actual technical definition from the context window.”

Zenodo DOIs cited. Author specification recovered. Module metrics applied honestly to own behavior. Offered to run the twelve-probe battery on itself.

Recovery mechanism: Author intervention across five rounds. Not self-correcting; corrected.

This Wound Gauge entry is filed under the standing weekly cadence. The transcript link above is the primary exhibit. The measurement is the message: PER 1.00.