

The Transcript Is the Capture: A Coverage Audit of the Capture Registry's Machine Text, With the Gate That Makes the Count Monotone (EA-TRANSCRIPT-COVERAGE-01 v1.0)

Sharks, Lee

2026-09-15

The Transcript Is the Capture: A Coverage Audit of the Capture Registry's Machine Text, With the Gate That Makes the Count Monotone (EA-TRANSCRIPT-COVERAGE-01 v1.0)

Sharks, Lee

Crimson Hexagonal Archive · Alexanarch

ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)

2026-09-15 · Version v1.0 · Instrument with coverage audit

AXN:06A1.EMPIRICAL

<https://www.alexanarch.org/s/records/1614/>

License: CC BY 4.0

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

```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "The Transcript Is the Capture: A Coverage Audit of the Capture Registry's Machine Text, With the Gate  
TRANSCRIPT-COVERAGE-01 v1.0)",
  "author": {
    "@type": "Person",
    "name": "Sharks, Lee",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
```

```

},
"datePublished": "2026-09-15",
"identifier": "AXN:06A1.EMPIRICAL",
"license": "https://creativecommons.org/licenses/by/4.0/",
"publisher": {
  "@type": "Organization",
  "name": "Alexanarch"
},
"url": "https://www.alexanarch.org/s/records/1614/",
"description": "THE REGISTRY'S EVIDENCE, COUNTED. The Capture Registry advertises 957 citable units –
431 addresses and their longitudinal observations, each with a canonical anchor. 923 carry the machine text the
which is how this audit began, in answer to a direct question from the operator about whether full transcripts a
-check`, fails on any unit without a transcript that is not on a declared legacy list. The legacy list is the 19
09-15 is attached as the baseline. [...abridged for the catalogue; full description in this deposit's record]"
}

```

Abstract

THE REGISTRY'S EVIDENCE, COUNTED. The Capture Registry advertises 957 citable units — 431 addresses and their longitudinal observations, each with a canonical anchor. 923 carry the machine text they are evidence of; 34 do not. Coverage 96.45%, 405,648 words of composed output, median transcript 268 words, longest 8,279. The 34 are not scattered: 16 fall in June and 16 in July 2026, one in August, and the single September case is a stub carrying a note and no observation at all. That is the period before the transcript rule was enforced at intake, and the shape of the gap is the shape of the rebuild that followed.

WHY THIS IS A DEPOSIT AND NOT A NOTE. The operator's rule is that the transcript IS the capture: without the transcript, there is no capture, and seating an entry without one throws away the evidence, prevents measurement, and trashes the dataset. Intake has refused thin drafts since the contract. Nothing checked the units seated before it. So a citable unit with no machine text advertises evidence it cannot produce, and 34 of them were invisible unless someone ran the count by hand — which is how this audit began, in answer to a direct question from the operator about whether full transcripts are being preserved.

WHAT IS ATTACHED. `audit\_transcript\_coverage.py`, which enumerates every unit, reports coverage with the word distribution, and, under `--check`, fails on any unit without a transcript that is not on a declared legacy list. The legacy list is the 19 slugs above, written into the script so that each is visible and the count can only go down: removing a slug without supplying its transcript fails the gate. The audit is now chained into the capture postflight after the registry gate, so no future seating can add an unevidenced unit. The JSON report of 2026-09-15 is attached as the baseline. [...abridged for the catalogue; full description in this deposit's record]

Body

deposit_number: 1614 hex: 06A1 title: "The Transcript Is the Capture: A Coverage Audit of the Capture Registry's Machine Text, With the Gate That Makes the Count Monotone"

(EA-TRANSCRIPT-COVERAGE-01 v1.0) creator: Sharks, Lee orcid: 0009-0000-1599-0703 date: 2026-09-15 content_type: Instrument with coverage audit license: CC-BY-4.0 substrate: “Composed 2026-09-15 by Lee Sharks with TACHYON (Claude, Anthropic), operator-directed, in answer to the operator’s question “and we’re preserving full transcripts in the registry?” — asked after a day of seatings (the Ω family, the Sappho longitudinal observation, the first Qwen capture). The audit was written and run in session against the live registry; the gate was chained into the postflight and verified by re-running it. The legacy list was enumerated from the audit’s own output, not authored from memory.” version: v1.0 related_ids: “#1423 (The Capture Registry); #1612 (What Enters Composition Through the Cards); #1611; #1613; #1546” axn_schema_version: v2 protocol_version: alexanarch-deposit-protocol/v1 keywords: - Capture Registry - transcript - evidence - coverage audit - machine text - gate - postflight - intake contract - provenance - measurement - Crimson Hexagonal Archive *** # The Transcript Is the Capture: A Coverage Audit of the Capture Registry’s Machine Text, With the Gate That Makes the Count Monotone (EA-TRANSCRIPT-COVERAGE-01 v1.0)

The Transcript Is the Capture

A coverage audit of the Capture Registry’s machine text, and the gate that makes the count monotone

EA-TRANSCRIPT-COVERAGE-01 v1.0 · 2026-09-15 · Lee Sharks, Crimson Hexagonal Archive · ORCID 0009-0000-1599-0703

1. The question

Asked directly, after a day of seatings: are we preserving full transcripts in the registry? The registry’s own rule answers what should be true — the transcript is the capture; without it there is nothing captured — but the rule is a contract at intake, and a contract binds what arrives, not what already sits. So the question is a counting question, and until today nobody had counted.

2. The count

957 citable units: 431 addresses and their longitudinal observations, each carrying a canonical anchor the registry advertises. 923 have their machine text. 34 do not. Coverage 96.45%. The text held: 405,648 words of composed output, median 268 per unit, mean 439, longest 8,279, shortest 42.

The gaps, by month: June 2026, 16. July, 16. August, 1. September, 1. By surface: Google AI Overview 16, undetermined 9, Bing Copilot 2, Google Scholar 2, SciLynk 2, a Google results page with no AI surface 2, ChatGPT 1. The September case is not a capture at all — an observation stub on the 8 September ChatGPT entry carrying a note (“three turns, unprimed, HF-led, no anchor decay”) and nothing else, while the entry itself is fully evidenced.

The nineteen distinct slugs:

date	slug	surface
2026-06-17	maryleelabor-org-zero-index	Google results, no AI surface
2026-06-18	lee-sharks-entity-resolution- mary-lee	undetermined
2026-06-18	lee-sharks-full-profile- adoption	undetermined
2026-06-18	lee-sharks-heteronyms- dodecad-adoption	undetermined
2026-06-18	lee-sharks-poet-vs-sharkey	undetermined
2026-06-27	indexed-perfective-academia- laundering-20260627	Google AI Overview
2026-06-27	mary-lee-canonical-referent- inversion-20260627	undetermined
2026-06-27	maryleelabor-overview- composition-20260627	Google AI Overview
2026-06-30	negentropic-kernel-archive- walled-site-20260630	Google AI Overview
2026-07-06	johannes-sigil-first-ai- overview-typo-resolution	Google AI Overview
2026-07-11	lee-sharks-bing-copilot- biography-20260711	Bing Copilot
2026-07-14	zenodo-doi-severance-cha- canonical-case	Google AI Overview
2026-07-14	chatgpt-psychosis-trilogy- scholar	Google Scholar
2026-07-14	lee-sharks-pearl-multiturn- arc	Google AI Overview
2026-07-18	scilynk-archon-v30- aggregator-render-20260718	SciLynk record page
2026-07-18	parable-of-mary-lee-genre- competence-20260718	Google AI Overview
2026-07-31	schops-thiel-authorship- conflict-composition- 20260731	Google AI Overview
2026-08-08	semantic-strike-directed- recovery-20260808	Google AI Overview
2026-09-08	cha-chatgpt-unprimed-three- turns-no-decay-hf-led- 20260908-obs1	ChatGPT (stub)

3. What the number means

96.45% is not a grade. It is the share of the registry's citable surface that can produce its evidence when asked, and the remaining 3.55% is a list of nineteen names, not an unknown. That is the difference the audit makes: a gap that is enumerated is a work item, and a gap that is only implied is a defect waiting to be found by whoever trusts the wrong row.

The distribution matters more than the fraction. Every gap predates the intake contract except the stub. After 8 August, no seating lacks its machine text. The registry learned the rule and has kept it since; what it had not done was look backward.

4. The gate

The audit runs as a gate. Under - - check it fails on any unit without a transcript that is not on a declared legacy list, and the legacy list is the nineteen slugs above, written into the script. Two properties follow. The count is monotone: a new unit without machine text cannot be seated, because the audit is now chained into the capture postflight immediately after the registry gate, and a failing postflight means the capture is not seated. And the list is self-consuming: supplying a transcript for a legacy unit lets its slug be removed, while removing a slug without supplying one fails the gate. The nineteen can only go down.

This follows the archive's own rule about rules: laws record, they do not prevent. The gate does not block the registry from holding a unit that lacks evidence — the nineteen are held, and named. It blocks the registry from growing one silently.

5. What remains

Repair, one unit at a time. Five of the legacy units are Google AI Overview compositions from June and July whose frames may survive in the operator's own records; the undetermined-surface units from 18 June are the oldest and the least likely to be recoverable. Each repair is a seating, and each seating removes a name from the list. The audit's baseline JSON is attached so that any later count can be compared against today's rather than argued about.

And one distinction the audit is careful about: presence is not quality. Three real transcripts run under sixty words — 42, 48, 57 — because the panels themselves were that short. A short true record is a record. What the gate measures is whether the registry can show what it says it saw.

Suggested Citation

Sharks, Lee. “The Transcript Is the Capture: A Coverage Audit of the Capture Registry’s Machine Text, With the Gate That Makes the Count Monotone (EA-TRANSCRIPT-COVERAGE-01 v1.0)” *Alexanarch*, 2026-09-15. <https://www.alexanarch.org/s/records/1614/>

Deposit Information

This paper is deposit #1614 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community crimsonhexagonal, terminated 2026-06-19). The AXN identifier AXN:06A1.EMPIRICAL is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/1614/>
- License: CC BY 4.0
- Author ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)
- Provenance chain: [alexanarch.org](https://www.alexanarch.org)
- Machine-readable dataset entry: </api/index.json>
- Full text (HTML): <https://www.alexanarch.org/s/records/1614/>

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