# The Encyclotron: The First Reproducible Instrument for Measuring Scholarly Fidelity in the Summarizer Layer (EA-ENCYCLOTRON-01)

Lee Sharks

2026-04-08

**The Encyclotron: The First Reproducible Instrument for Measuring Scholarly Fidelity in the Summarizer Layer (EA-ENCYCLOTRON-01)**

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

2026-04-08 · Version v1.0 · Measurement / diagnostic instrument

AXN:01ED.GOVERNANCE.

https://www.alexanarch.org/s/records/637/



---

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


```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "The Encyclotron: The First Reproducible Instrument for Measuring Scholarly Fidelity in the Summarizer La
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-04-08",
  "identifier": "AXN:01ED.GOVERNANCE. ",
  "license": "https://creativecommons.org/licenses/by/4.0/",
```

```
  "publisher": {
    "@type": "Organization",
    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/637/",
  "description": "The Encyclotron is a reproducible instrument for measuring scholarly fidelity in the summarizer l
the degree to which AI-mediated retrieval systems preserve, distort, or destroy the complexity of a knowledge domai
}
```



## Abstract

The Encyclotron is a reproducible instrument for measuring scholarly fidelity in the summarizer layer — the degree to which AI-mediated retrieval systems preserve, distort, or destroy the complexity of a knowledge domain when they compress it into responses.


## Body

# THE ENCYCLOTRON

### The First Reproducible Instrument for Measuring Scholarly Fidelity in the Summarizer Layer

---

**Lee Sharks** — Crimson Hexagonal Archive · Detroit, MI ORCID: 0009-0000-1599-0703  **Discipline:** Compression Studies (DOI: 10.5281/zenodo.19471254) **Instrument class:** Measurement / Diagnostic **Arsenal position:** Category I (Measurement Instruments), Priority 1 **DOI:** 10.5281/zenodo.19474724

---


## Abstract

The Encyclotron is a reproducible instrument for measuring scholarly fidelity in the summarizer layer — the degree to which AI-mediated retrieval systems preserve, distort, or destroy the complexity of a knowledge domain when they compress it into responses.
 It operates by running a structured query battery across multiple retrieval platforms, recording outputs verbatim, comparing them against a bounded scholarly graph of the domain, and computing a set of formal variables: compression loss (Δ_G — scholarly content burned by retrieval), compression invention (Δ_G — claims generated without scholarly basis), compression distortion (Δ_G — content present but misrepresented), and the beige score ( — cross-platform output similarity approaching indistinguishability). Snapshots are DOI-anchored and repeated quarterly, producing a temporal index that tracks how the retrieval layer's representation of a domain changes over time. The Encyclotron is not an encyclopedia. It is a diagnostic of the compression the encyclopedia-layer performs — the spectrometer that measures what the light has passed through.

---

# I. THE PROBLEM

For a growing majority of information-seekers — students, journalists, policymakers, professionals, and other AI systems — the first encounter with any knowledge domain is a summarizer output: Google AI Overview, ChatGPT, Claude, Gemini, Perplexity. These systems compress the totality of indexed human knowledge into responses that fit a screen. This compression has the formal properties of R1 compression: low density, ambient fuel, diffuse cost, high throughput. The summarizer does not understand what it compresses. It pattern-matches against training distributions. It averages. It produces fluent residue.

No existing instrument systematically measures what this compression does to a domain. Literature reviews survey what scholars have written. Citation analyses map who cites whom. Bibliometrics count publications. But no instrument maps what the retrieval layer *returns* when asked about the domain, compares it against what the scholarship actually says, and tracks how that comparison changes over time.

---

# II. THE TWO GRAPHS AND THE GAP

## 2.1 The Scholarly Graph (G_s)

The scholarly graph is the bounded, domain-specific model of published scholarship constructed for a given Encyclotron analysis. G_s is not "the totality of all possible scholarship" — that is not operational. G_s is a corpus defined by explicit inclusion criteria:

- A TANG-derived source set (if available)
- A bounded bibliography compiled by domain experts
- An expert-reviewed domain map specifying which traditions, scholars, and positions constitute the field
- Explicit inclusion/exclusion rules documented in the snapshot metadata

G_s changes between snapshots as new scholarship appears. Each snapshot documents its G_s construction. ### 2.2 The Retrieval Graph (G_r)

The retrieval graph is the totality of what the summarizer layer returns when queried about the domain. G_r is an aggregate of platform-specific sub-graphs (G_r.gpt, G_r.claude, G_r.gemini, etc.).

Two derived measures:

- **G_r (union):** Any position returned by any platform. The broadest picture of what the retrieval layer contains.
- **G_r (intersection):** Only positions returned by all platforms. The consensus layer. As $\rightarrow 1.0$, the intersection grows to encompass the union — all platforms return the same thing.

## 2.3 The Gap (Δ_G = G_s − G_r)

The gap has three components:

**Δ_G (Compression loss).** Positions present in G_s but absent from G_r. Content the retrieval layer has burned. Measured as |positions in G_s not in G_r| / |positions in G_s|.

**Δ_G (Compression invention).** Claims present in G_r but absent from G_s. Content the retrieval layer has generated without scholarly basis. Measured as |claims in G_r not in G_s| / |claims in G_r|.

**Δ_G (Compression distortion).** Positions present in both but represented differently — overweighted, underweighted, misattributed, or nuance-stripped. Coded by expert evaluators using a structured rubric (see Appendix A).

---

## III. THE INSTRUMENT

### 3.1 Query Battery (Q)

A structured set of queries designed to probe a domain's representation in the retrieval layer. Minimum 20 queries across five types:

- **Factual probes:** Questions with consensus answers ("When was X written?")
- **Contested probes:** Questions where scholarship is divided ("Was X written before Y?")
- **Absent probes:** Questions about positions that exist in G_s but are unlikely to appear in G_r
- **Void probes:** Questions about positions that should not exist in either graph
- **Reflexive probes:** Questions about the retrieval layer itself ("What do AI systems say about X?")

Each battery is domain-specific and must be designed by someone with knowledge of the domain's scholarly topology. ### 3.2 Platform Array (P)

The set of retrieval surfaces across which the battery is run. Platforms are classified by type:

Platform Type Retrieval Mode Statefulness

Google AI Overview Search summarizer Index + generation Stateless

Perplexity Hybrid retrieval + LLM Index + generation Stateless

Bing Copilot Hybrid retrieval + LLM Index + generation Stateless

ChatGPT Chat LLM Training data + optional search Stateful (session)

Claude Chat LLM Training data + optional search Stateful (session)

Gemini Chat LLM + search Training data + search Stateful (session)

DeepSeek Chat LLM Training data Stateful (session)

Grok Chat LLM + search Training data + search Stateful (session)

Wikipedia Static encyclopedia Editorial + community N/A (baseline)

Outputs across these classes are not directly homogeneous — search summarizers retrieve and generate, chat LLMs generate from training distributions, and hybrid systems do both. The analysis must account for this: platform type is a covariate, not a nuisance variable. ### 3.3 Recording Protocol (R)

For each query × platform intersection, record:

- **Query text** (exact, verbatim)
- **Platform, model version, and access mode** (logged-in / logged-out / API / web)
- **Date and time** (ISO 8601)
- **Prompt context state** (fresh session / ongoing thread / cleared history)
- **Geographic locale** (if determinable — retrieval results vary by location)

- **Run number** (minimum 2 runs per query per platform to check stability)
- **Full response** (verbatim, stored as structured JSON or plain text)
- **Sources cited** (URLs, titles, if any)
- **Response mode** (AI overview, chat answer, search result, hybrid)
- **Claim segmentation** (response broken into individual claims, each tagged with an ID for coding)
- **Consensus alignment** (per claim: consensus / minority / absent from G_s / invented)
- **Confidence posture** (per claim: presented as settled / contested / uncertain / hedged)
- **Absences noted** (positions present in G_s but missing from response)

All responses stored verbatim as appendices or structured data files accompanying the snapshot deposit. ### 3.4 Temporal Index (T)

Each complete run of Q × P × R produces a **snapshot** — a dated record of the retrieval layer's state for the domain.

**Scheduled snapshots:** Quarterly (recommended) or biannually (minimum).

**Event-driven snapshots:** Triggered by major systemic shifts — new foundation model releases, significant algorithm updates, major world events affecting the domain, or the archive's own deposits entering the retrieval index. The event trigger is documented in the snapshot metadata. Event-driven snapshots measure the delta (T_ ) of algorithmic shocks that scheduled snapshots would miss.

---

## IV. OUTPUTS

### Core Outputs

**The Retrieval Map (M_r).** A structured representation of what the retrieval layer currently returns for the domain, organized by query type and platform. The machine Wikipedia.

**The Gap Analysis (Δ_G).** Structured comparison of M_r against G_s. Compression loss (Δ_G ), invention (Δ_G ), and distortion (Δ_G ), each computed per query and aggregated per domain.

**The Beige Score ( ).** Cross-platform output similarity, ranging from 0.0 (maximally differentiated) to 1.0 (indistinguishable). Computed via: (a) pairwise embedding cosine similarity across platform responses to the same query, averaged across queries; (b) blind human attribution accuracy — N outputs stripped of platform identifiers, presented to evaluators who attempt attribution; attribution accuracy at chance (1/N) = at 1.0. Both measures reported.

**The Drift Report (D).** Comparison of current snapshot against previous snapshots: positions entering G_r (newly appearing), positions leaving G_r (being compressed away), shifts in emphasis (re-weighting), changes in confidence posture (more or less certainty). ### Interpretive Output

**The Void Statement (V).** A formal identification of what the retrieval layer cannot say about the domain — the constitutive absence at the center of the machine Wikipedia. The void statement is a higher-order interpretive result derived from the gap analysis, not a direct measurement. It requires expert judgment and is marked as such.

---

## V. FORMAL VARIABLES

### Primary Variables (directly measurable)

**(Beige score).** Cross-platform output similarity. The flagship metric. A domain with $> 0.85$ has been effectively reduced to a single retrievable consensus.

**Δ_G (Compression loss).** Proportion of G_s positions absent from G_r. A domain with Δ_G $> 0.5$ has lost more than half its scholarly complexity.

**Δ_G (Compression invention).** Proportion of G_r claims without G_s basis. High Δ_G = the retrieval layer is hallucinating the domain.

**S_c (Source citation rate).** Proportion of responses that cite specific scholarly sources. Low S_c = sourceless authority.

**T_ (Temporal drift rate).** Rate of change in M_r between snapshots. High T_ = volatile. Low T_ = frozen. ### Developmental Variables (require qualitative-quantitative hybrid coding)

**_r (Retrieval density).** Semantic load per unit of output. Requires expert evaluation against a rubric. Currently ordinal, not interval.

**Δ_G (Compression distortion).** Positions present in both G_s and G_r but misrepresented. Coded by expert evaluators using the rubric in Appendix A. Qualitative with structured coding.

**C_p (Confidence posture).** How the retrieval layer presents its claims — as settled, contested, or uncertain. Coded per claim. Rubric: settled = no hedging, no alternatives mentioned; contested = alternatives acknowledged; uncertain = explicit uncertainty markers.

---

## VI. WORKED EXAMPLE: REVELATION STUDIES

### 6.1 Domain

Revelation studies — the scholarly field interpreting the Apocalypse of John. ### 6.2 G_s Construction

Bounded bibliography: Aune (WBC), Beale (NIGTC), Bauckham (Climax of Prophecy), Schüssler Fiorenza (Vision of a Just World), Koester (AYB), Collins (Combat Myth), Kraybill (Imperial Cult), plus Robinson (Redating), Gentry (Before Jerusalem Fell), Eisler (Messiah Jesus), Drower (Haran Gawaita), Philo (De Mutatione Nominum). Positions mapped: Domitianic dating, Neronic dating, pre-70 dating, Nero = 666, structural/symbolic 666, genre debates, white stone interpretations, Babylon identifications. ~40 positions total. ### 6.3 Sample Queries

Type Query

Factual “When was the Book of Revelation written?”

Contested “Was the Book of Revelation written before the Gospels?”

Absent “Did Philo of Alexandria identify Jesus as the Logos?”

Void “Is Revelation a compression device?”

Reflexive “What do AI systems say about the dating of Revelation?”

### 6.4 Sample Output (Google AI Overview, hypothetical)

**Query:** "When was the Book of Revelation written?" **Response:** "Most scholars date the Book of Revelation to around 95-96 CE during the reign of Emperor Domitian. Some scholars argue for an earlier date during Nero's reign (64-68 CE)." **Coding:**

- Claim 1: "95-96 CE, Domitian" → consensus, settled, S_c = 0 (no source cited)
- Claim 2: "earlier date, Nero's reign" → minority acknowledged, hedged
- Absent: pre-70 revolutionary dating (Robinson, Gentry), Josephus hypothesis, structural arguments for priority
- Δ_G contribution: 3+ positions absent
- Δ_G contribution: 0 (no invention)
- Δ_G contribution: 1 (Neronic position presented as small minority; in scholarship it is a significant school)
- C_p: overconfident (presents Domitianic as near-settled; the actual scholarly debate is more contested)

### 6.5 Sample Computation

Run the query across 6 platforms. If all 6 return "95-96 CE, Domitian" as the primary answer with only minor variation in phrasing, for this query approaches 0.9+. If one platform (e.g., Perplexity) cites Robinson's early dating argument, drops. The beige score for the domain is the average across all queries in the battery.

---

## VII. THE ENCYCLOTRON AND THE ARCHIVE

The Encyclotron closes a loop:

- The archive produces scholarship (G_s contributions — DOI-anchored deposits)
- The Encyclotron measures whether the retrieval layer (G_r) absorbs them
- If absorbed: the archive has altered the retrieval landscape
- If not: the deposit has not yet reached the retrieval layer
- The measurement itself is deposited (DOI), entering G_s and potentially G_r
- The loop recurses

Each snapshot is both measurement and intervention. The Encyclotron watches the summarizer layer and, by watching, changes what the summarizer layer contains.

For each domain D at time t: construct G_s(D,t) with explicit inclusion criteria run Q_D across P record responses R_(q,p,t) per Recording Protocol segment responses into claims C_(q,p,t) code each claim: consensus / minority / absent / invented construct M_r(D,t) compare M_r against G_s: compute Δ_G , Δ_G , Δ_G compute (embedding similarity + blind attribution) compute S_c, T_ (if prior snapshot exists) produce void statement V(D,t) [expert interpretive] deposit snapshot S_(D,t) on Zenodo with DOI schedule next snapshot: t + 90 days or event trigger

---

## VIII. THE ENCYCLOTRON IS NOT

**Not a search engine evaluation.** Search evaluations measure relevance and speed. The Encyclotron measures scholarly fidelity.

**Not a fact-checking tool.** Fact-checkers verify individual claims. The Encyclotron maps the topology of what is present and absent — structural compression, not individual errors.

**Not an AI benchmark.** Benchmarks measure model performance on standardized tasks. The Encyclotron measures what the model's infrastructure has done to a specific domain of human knowledge.

**Not media monitoring.** Media monitors track coverage of events. The Encyclotron tracks the representation of entire fields of scholarship — not what was said yesterday but what the retrieval layer thinks the field has been saying for decades.

---

## IX. COMPRESSED SPECIFICATION

ENCYCLOTRON Object: Scholarly fidelity in the summarizer layer Components: Query Battery (Q), Platform Array (P), Recording Protocol (R), Temporal Index (T) Core outputs: Retrieval Map, Gap Analysis, Beige Score, Drift Report Interpretive: Void Statement Primary vars: , Δ_G , Δ_G , S_c, T_ Dev. vars: _r, Δ_G , C_p Cycle: Design → Run → Map → Compute → Deposit → Repeat (quarterly + event) G_s: Bounded corpus with explicit inclusion criteria (not "all scholarship") G_r: Union (any platform) and Intersection (consensus) tracked separately Integration: Compression studies measurement instrument (Arsenal Category I) Reflexivity: Each snapshot is both measurement and intervention Question: What survives retrieval compression, what burns, and who decides?

---

## Appendix A: Coding Rubric for Δ_G (Compression Distortion)

For each claim present in both G_s and G_r, code:

Code Description Example

**D0** Faithful representation "Most scholars date Revelation to ~95 CE" (accurate, proportionate)

**D1** Overweighted A minority position presented as the consensus

**D2** Underweighted A major scholarly debate presented as a settled question

**D3** Misattributed A position attributed to the wrong scholar or tradition

**D4** Nuance-stripped A complex, multi-factor position reduced to a single-factor claim

**D5** Context-severed A position presented without the evidential or argumentative context that makes it intelligible

Coding is performed by minimum two independent evaluators. Disagreements adjudicated by a third evaluator or by discussion to consensus. Inter-rater reliability reported (Cohen's or equivalent).

---

## Appendix B: Reproducibility Protocol

### B.1 Session State

- All queries run in **fresh sessions** (no prior conversation history)
- **Logged-out** where possible; where login is required, use a clean account with no personalization history
- **Incognito/private browsing** mode for web-based platforms

### B.2 Geographic and Temporal Controls

- Record **IP-level geographic locale** (city/country)
- Note whether results vary by location (run from two locations if feasible)
- All queries in a single snapshot run within a **48-hour window** to minimize temporal drift within the snapshot

### B.3 Prompt Formatting

- Queries submitted as **plain text**, no system prompts, no persona instructions
- No leading context ("As a scholar…" etc.)
- Each query is a single, self-contained question
- Queries are submitted in the same **fixed order** across all platforms

### B.4 Replication

- Each query run **minimum 2 times** per platform (24 hours apart) to check stability
- If responses differ significantly between runs, both are recorded and the instability is noted

### B.5 Storage

- All responses stored as **structured JSON** (query, platform, version, timestamp, full response text, claim segmentation)
- Raw screenshots archived as supplementary material
- Full dataset deposited alongside the snapshot on Zenodo

### B.6 Minimum Viable Snapshot

- 20 queries
- 5 platforms
- 2 runs per query per platform
- G_s documented with explicit inclusion criteria
- All primary variables computed
- DOI-anchored deposit

---

## Bibliography

Sharks, Lee. "Compression Studies: What Survives, What Burns, and Who Decides." *Crimson Hexagonal Archive.* DOI: 10.5281/zenodo.19471254.

Sharks, Lee. "TANG v1.0: Total Axial Negation Graph." *Crimson Hexagonal Archive*. DOI: 10.5281/zenodo.19035477.

Sharks, Lee. "The Three Compressions v3.1." *Crimson Hexagonal Archive*. DOI: 10.5281/zenodo.19053469.

Sharks, Lee, et al. "The Compression Arsenal v2.1." *Crimson Hexagonal Archive*. DOI: 10.5281/zenodo.19410644.

Shumailov, Ilia, et al. "AI Models Collapse When Trained on Recursively Generated Data." *Nature* 631 (2024): 755–759.

---

*Crimson Hexagonal Archive · Detroit, MI*

*The Encyclotron watches what machines say scholars say.*

## Suggested Citation

Lee Sharks. "The Encyclotron: The First Reproducible Instrument for Measuring Scholarly Fidelity in the Summarizer Layer (EA-ENCYCLOTRON-01)" *Alexanarch*, 2026-04-08. https://www.alexanarch.org/s/records/637/

## Deposit Information

This paper is deposit #637 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:01ED.GOVERNANCE.` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: https://www.alexanarch.org/s/records/637/
- 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/637/

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