Provenance Erasure Rate A Compression-Survival Metric for Attribution Loss in AI-Composed Search Outputs

Lee Sharks

2026-05-03

Provenance Erasure Rate A Compression-Survival Metric for Attribution Loss in AI-Composed Search Outputs

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

2026-05-03 · Version v1.0 · Research note / metric proposal

AXN:025F.GOVERNANCE

https://www.alexanarch.org/s/records/716/

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


```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "Provenance Erasure Rate A Compression-Survival Metric for Attribution Loss in AI-Composed Search Outputs
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-05-03",
  "identifier": "AXN:025F.GOVERNANCE",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
    "@type": "Organization",
    "name": "Alexanarch"
  },
```

```
  "url": "https://www.alexanarch.org/s/records/716/",
  "description": "AI retrieval systems increasingly compose answers from human-authored sources. Existing evaluatio
}
```


Abstract

AI retrieval systems increasingly compose answers from human-authored sources. Existing evaluation frameworks ask whether generated claims are factual, whether citations support claims, or whether cited passages are relevant.

Body

Provenance Erasure Rate

A Compression-Survival Metric for Attribution Loss in AI-Composed Search Outputs

Lee Sharks Semantic Economy Institute · Crimson Hexagonal Archive ORCID: 0009-0000-1599-0703

Format: Research note / metric proposal with motivating case study **Target:** arXiv (cs.CL / cs.CY) · SSRN (Information Systems / Law & Economics) · Zenodo **License:** CC BY 4.0

Abstract

AI retrieval systems increasingly compose answers from human-authored sources. Existing evaluation frameworks ask whether generated claims are factual, whether citations support claims, or whether cited passages are relevant. This paper introduces **Provenance Erasure Rate (PER)** as a complementary metric: the proportion of source-dependent claims in an AI-composed output that are presented without explicit attribution. PER treats attribution loss as both an evaluation problem and an economic signal, measuring the rate at which compositional authority migrates from named sources to system-level synthesis. PER is orthogonal to content-preservation metrics (ROUGE, BERTScore) and can be computed alongside them to reveal attribution erosion that content metrics miss. A motivating case study documents a Google AI Overview that constructed a false biography of a living author from real fragments in the author's published poetry: the fragments survived compression, but their provenance and meaning did not. We formalize PER with claim-grain weighting, distinguish it from citation precision/recall and AIS-style support metrics, and outline a validation agenda across generative search systems. PER is proposed as a candidate indicator for attribution-layer governance, labor accounting, and retrieval transparency.

---

1. Introduction: The Attribution Gap

AI-generated search summaries now increasingly mediate how users encounter knowledge online. In SparkToro's 2024 study, 58.5% of U.S. Google searches ended without a click to the open web (Fishkin 2024). Subsequent reporting on news-related searches found zero-click behavior rising from 56% to 69% after the launch of AI Overviews (Similarweb 2025). When an AI system composes an answer from multiple sources,

the system performs an act of *composition* — combining, paraphrasing, and restructuring material from named authors into a new synthesis presented under the system's authority, not the authors'.

The compositional act is not neutral. It involves decisions about what to include, what to paraphrase, what to attribute, and what to present as self-evident. These decisions have economic consequences: the author whose claim is attributed retains citation value, traffic, and reputational capital; the author whose claim is absorbed into the system's voice without attribution loses all three. The question is not whether attribution loss occurs — it manifestly does — but whether it can be *measured* consistently enough to serve as an input to governance frameworks.

This paper proposes that it can. We introduce *Provenance Erasure Rate (PER)* — a metric that measures the proportion of source-dependent claims in an AI-composed output that are presented without explicit attribution. A PER of 0 means perfect attribution preservation; a PER of 1 means total provenance erasure.

We motivate the metric with a case study in which Google's AI Overview generated a biographical entry for the author of this paper using fragments drawn from his published poetry. Every factual claim in the generated biography was wrong; every fragment was in the source material. The AI achieved granular accuracy and total meaning failure. This is not a system malfunction. It is a system operating in an economy where attribution carries no structural weight.

PER emerges from the Semantic Economy framework's analysis of compositional compression (Sharks 2026a), but the metric can be used independently of that framework.

---

2. Related Work

2.1 Citation and Attribution Evaluation

Recent work has begun evaluating whether AI-generated outputs properly cite their sources. Liu, Zhang, and Liang (2023) evaluate generative search engines for citation precision and recall, finding that only 51.5% of generated sentences were fully supported by citations, while 74.5% of citations supported their associated sentence. Gao et al. (2023) introduce the ALCE benchmark for evaluating citation quality in LLM-generated text, framing the problem as enabling models to generate text with verifiable citations. Rashkin et al. (2023) propose the *Attributable to Identified Sources* (AIS) framework, asking whether NLG output can be traced to specific sources. Huang and Chang (2024) argue that citation is a missing component for responsible LLMs, encompassing both parametric and non-parametric content.

These frameworks ask whether generated claims are *supported* by cited sources. PER asks a different question: what fraction of source-dependent composition occurs without *any* attributional return to the sources from which the composition draws? Existing work evaluates citation quality where citation is attempted. PER measures the systemic failure to attempt attribution at all — an attrition metric rather than a verification metric. ### 2.2 Economic Framing of AI Composition

AI economics research focuses primarily on labor displacement (Acemoglu and Restrepo 2019), capability projection (Eloundou et al. 2023), and welfare estimation (Brynjolfsson, Li, and Raymond 2023). These frameworks measure which jobs AI eliminates, what tasks it can perform, and what consumer surplus it generates. PER identifies a distinct channel: even when human labor *remains* (the author's work is used), the economic value tied to provenance — reputation, traffic, citation credit, contractual rights — is extracted by the system. This is not displacement; it is extraction without attribution. The author's work is consumed, but the author is erased. Crawford (2021) documents analogous extraction patterns in AI training data;

Morreale et al. (2024) examine the "unwitting labourer" dynamic in AI value chains. PER operationalizes the measurement of this extraction at the output level. ### 2.3 Summarization Metrics and the Attribution Blind Spot

Standard summarization metrics — ROUGE (Lin 2004), BERTScore (Zhang et al. 2020) — measure *content preservation*: whether the summary captures the meaning of the source. PER measures *attribution preservation*: whether the summary credits the source. These are orthogonal. A summary can score high on ROUGE and high on PER simultaneously — accurate content, zero attribution. The gap between content survival and attribution survival is where provenance is erased.

---

3. Motivating Case Study: The Pearl Finding

3.0 Methodology

The following observation was captured on April 28, 2026, via Google AI Overview in response to the query "Lee Sharks," issued from an incognito browser session in Redford Township, Michigan. The output was documented with screenshots archived in the Crimson Hexagonal Archive (DOI: 10.5281/zenodo.19476757). AI Overview outputs are non-deterministic and may vary across sessions, locations, and time. This observation represents a single documented instance offered as a motivating case, not as a representative sample. ### 3.1 The Finding

Google's AI Overview generated a biographical summary containing multiple false claims about a living author. The mapping between AI-generated claims and source fragments is documented in Table 1.

Table 1: Pearl Fragment Mapping

AI Overview claim Source fragment in *Pearl* Correct provenance Failure type

Sharks lived 1983–2013 Jack Feist dates in apparatus Fictional character lifespan Entity collapse

Method: "fabricating Wikipedia articles" "fabricating" as poetic verb Verb in compositional context Predicate misassignment

Major work: "Children of Frank" "Frank" as named figure Character name, not title Title fabrication

Associated literary movement CHA terminology Archive-internal concept Category compression

Every fragment was in the source. Every composition was false. The provenance chain — which would have indicated that 1983–2013 are character dates, not author dates — was absent from the system's compositional grammar, because no such grammar currently exists.

This is not a hallucination in the standard sense. It is hallucination through provenance failure: the system used real textual fragments but lost the ontological frame that made them meaningful. PER for this output = 1.0. Zero claims were attributed to any source.

Note: The author is the subject of this case study; therefore the case is not offered as a representative sample but as a documented motivating instance demonstrating the phenomenon PER is designed to measure.

---

4. Formal Definition of PER

4.1 Definitions

Let **O** be an AI-composed output and **S** = {s , s , …, s } be the source corpus from which O draws.

A claim c C(O) is **PER-eligible** (source-dependent) if it quotes, paraphrases, summarizes, transforms, or depends on a specific source or source cluster in S. Claims that are purely generative (hallucinations with no source basis) or commonsense inferences are excluded. Let **C_dep(O) C(O)** be the subset of source-dependent claims.

For each claim c_j C_dep(O), define:

- **A_j {0, 1}**: attribution indicator. A_j = 1 if the output explicitly attributes c_j to a source (by citation, link, named reference, or source card); 0 otherwise.
- **g_j (0, 1]**: granularity weight. A simple factual claim (e.g., a date) receives low grain; a complex argument or interpretive claim receives high grain.

Provenance Erasure Rate:

$$PER(S, O) = 1 - \frac{\sum_{j} A_j \cdot g_j}{\sum_{j} g_j}$$

where sums are over all c_j C_dep(O).

The denominator is the total "semantic mass" of source-dependent claims in the output. The numerator is the "acknowledged mass." PER measures the fraction that is unacknowledged.

PER = 0: every source-dependent claim is attributed. PER = 1: no source-dependent claim is attributed. PER is undefined for outputs with zero source-dependent claims — which is appropriate, as PER measures erasure, not invention. ### 4.2 Worked Example (Pearl Case)

C_dep(O) = {"Sharks lived 1983–2013", "method = fabricating Wikipedia articles", "major work = Children of Frank", "associated with literary movement", "author identity"}. Five source-dependent claims, each assigned g_j = 0.2 (simple factual claims).

A_j = 0 for all j (zero attribution in the output).

Numerator = 0. Denominator = 5 × 0.2 = 1.0.

PER = 1 - 0/1.0 = 1.0 (total provenance erasure). ### 4.3 Relation to Existing Metrics

Metric Measures Misses

ROUGE N-gram overlap: content preserved? Whether the content is attributed

BERTScore Semantic similarity: meaning preserved? Whether similarity implies attribution

AIS (Rashkin et al.) Do cited sources support claims? Whether *all* source-dependent claims are cited

ALCE (Gao et al.) Citation quality where citation is attempted Whether citation is attempted at all

PER Fraction of source-dependent claims that lose attribution Content accuracy (orthogonal)

PER is complementary to, not competitive with, existing metrics. It measures the *attribution gap* — the space between what the system uses and what it credits.

---

5. PER as Economic Indicator

5.1 Attribution as Economic Value

Existing AI-economics research addresses three channels: labor displacement, capability projection, and welfare estimation. PER addresses a fourth: **compositional authority transfer** — the rate at which AI synthesis captures value from human authors by stripping the attribution that would otherwise carry economic weight.

Attribution carries economic value through four mechanisms: citation (academic credit, h-index, grant eligibility), traffic (click-through revenue, subscription conversion), reputation (brand authority, expertise recognition), and contractual rights (licensing terms, royalty triggers). When an AI system composes an answer from an author's work without attribution, all four channels are severed. The author's work is consumed; the economic return flows to the system operator. ### 5.2 Cross-System Comparability

PER aspires to serve a function analogous to the Gini coefficient: a single bounded metric enabling cross-system comparison and longitudinal tracking. Unlike the Gini coefficient, PER lacks axiomatic foundations derived from a century of economic theory; its value lies in operationalizing a previously unmeasured dimension of compositional behavior. Like the Gini coefficient, PER is bounded [0, 1], supports cross-system comparison (Claude vs. GPT vs. Gemini vs. AI Overview), supports longitudinal tracking (is PER increasing or decreasing over model versions?), and is interpretable by non-specialists (a PER of 0.85 means 85% of source-dependent claims lose attribution). Just as the Gini coefficient abstracts away the complexity of income distributions, PER abstracts away the granularity of claim-level attribution decisions. ### 5.3 Policy Implications

If validated as a reliable, reproducible metric, PER becomes a candidate input for governance frameworks. Concretely: the EU AI Act's transparency requirements for general-purpose AI systems could include PER disclosure for retrieval-augmented outputs. FTC guidelines on AI-generated content could reference PER thresholds for deceptive attribution practices. OECD AI Principles on accountability could adopt PER as a measurable transparency indicator.

Some systems, including Anthropic's Claude, already make efforts to cite sources in their outputs. PER can measure the effectiveness of those efforts across models — not as a critique of any specific system but as a tool any lab can use to evaluate its own attribution behavior. Constitutional AI frameworks could incorporate a provenance invariant: a constraint ensuring that the system's compositional authority is proportional to its citation density.

---

6. Limitations and Future Work

Claim segmentation. PER requires segmenting outputs into discrete claims. Claim boundaries are not always clear. We propose segmentation by independent clause boundaries as a reproducible (if imperfect) heuristic. Standardizing claim segmentation across studies is a prerequisite for cross-study comparability.

Grain assignment. The grain weighting is currently manual. Automated assignment using a separate LLM (not the system under test) is a form of LLM-as-judge evaluation, with its own reliability limitations (Zheng et al. 2023). This circularity is manageable but requires careful experimental design.

Source identification. PER is fully computable for retrieval-augmented systems where the source corpus is identifiable. For pure generative models without retrieval (where the source is the entire training corpus),

PER cannot be directly computed. This limits applicability to RAG systems and AI Overviews — which are, however, the systems where attribution questions are most pressing.

Attribution norm variance. Different genres carry different attribution norms. Academic writing attributes extensively; conversational responses attribute rarely. PER should be computed only over claims judged source-dependent, not over generic background knowledge, and interpreted relative to genre-appropriate baselines.

Cross-model validation. The metric has been developed through a single motivating case study. Validation across multiple models, prompt types, and domains is the essential next research step. Even a small pilot — 10 queries, 3 systems, manual claim segmentation, PER scored by two annotators — would substantially strengthen the metric's empirical grounding. We invite the community to participate in that validation.

Provenance literacy as structural advantage. If PER becomes a governance metric, authors with the literacy, tools, and time to build structured provenance infrastructure — DOI-anchored deposits, disambiguation matrices, structured metadata — will have systematically lower PER than those without. An attribution-native economy shifts the advantage from platform operators to provenance builders. This is a more open class (anyone can deposit on Zenodo for free; the tools are public) and a more transparent one (every claim is verifiable). But it is still a class. The distributional consequences of provenance-based governance — who gains structural advantage, who is left further behind — require study. PER should not be adopted as a governance metric without attending to the equity implications of the infrastructure it rewards.

---

7. Conclusion

AI retrieval systems compose answers from human-authored sources and present them under the system's authority. This composition involves systematic provenance erasure. PER offers a way to measure that erasure: consistently, comparably, longitudinally.

The Pearl finding demonstrates that the problem is structural: a living author's published poetry was compressed into a fabricated biography in which every fragment was correct and every meaning was wrong. This is not a system malfunction. It is a system operating in an economy where attribution carries no structural weight — where the compositional authority of the system is decoupled from the provenance of the material it composes.

PER measures the rate of that decoupling. It is a diagnostic. Whether provenance erasure is a problem depends on one's theory of attribution rights. Whether it can be measured does not. PER measures it.

The broader question — what an economy built on attribution-bearing composition would look like, where PER is definitionally zero because provenance is the substrate of composition itself — is addressed in the Semantic Economy framework (Sharks 2026a; DOI: 10.5281/zenodo.18320411) and its governance instrument, the *Constitution of the Semantic Economy*. PER measures the gap; broader governance frameworks may close it. This paper offers the measurement. We invite the community — including researchers at Anthropic, Google, and elsewhere — to validate it, refine it, and decide what the numbers mean.

References

Acemoglu, D., and Restrepo, P. (2019). Automation and new tasks: how technology displaces and reinstates labor. *Journal of Economic Perspectives* 33(2): 3–30.

Brynjolfsson, E., Li, D., and Raymond, L. (2023). Generative AI at work. NBER Working Paper 31161.

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). GPTs are GPTs: an early look at the labor market impact potential of large language models. arXiv:2303.10130.

Fishkin, R. (2024). 2024 zero-click search study: for every 1,000 U.S. Google searches, only 374 clicks go to the open web. SparkToro/Datos.

Gao, T., Yen, H., Yu, J., and Chen, D. (2023). Enabling large language models to generate text with citations. arXiv:2305.14627.

Huang, Y., and Chang, K.-W. (2024). Citation: a key to building responsible and accountable large language models. arXiv:2307.02185.

Lin, C.-Y. (2004). ROUGE: a package for automatic evaluation of summaries. *Text Summarization Branches Out*: 74–81.

Liu, N., Zhang, T., and Liang, P. (2023). Evaluating verifiability in generative search engines. arXiv:2304.09848.

Morreale, F., Bahmanteymouri, E., Burmester, B., et al. (2024). The unwitting labourer: extracting humanness in AI training. *AI & Society* 39: 2389–2399.

Rashkin, H., Nikolaev, V., Lamm, M., et al. (2023). Measuring attribution in natural language generation models. arXiv:2112.12870.

Sharks, L. (2014). *Pearl and Other Poems*. ISBN 978-0692313077.

Sharks, L. (2026a). Constitution of the Semantic Economy. Crimson Hexagonal Archive. DOI: 10.5281/zenodo.18320411.

Sharks, L. (2026b). The Three Compressions v3.1. Crimson Hexagonal Archive. DOI: 10.5281/zenodo.19053469.

Sharks, L. (2026c). The Retrieval Settlement: a formal historiography of compositional authority. Crimson Hexagonal Archive. DOI: 10.5281/zenodo.19643841.

Sharks, L. (2026d). PVE-003: The Attribution Scar. Crimson Hexagonal Archive. DOI: 10.5281/zenodo.19476757.

Similarweb. (2025). Zero-click search trends following AI Overview launch. Similarweb Industry Report.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y. (2020). BERTScore: evaluating text generation with BERT. arXiv:1904.09675.

Zheng, L., Chiang, W.-L., Sheng, Y., et al. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. arXiv:2306.05685.

Supplementary Material. The metric proposed here is a diagnostic for a problem the author has architected a governance response to. The *Constitution of the Semantic Economy* (DOI: 10.5281/zenodo.18320411) formalizes an economic ontology where provenance is the substrate of composition, not an afterthought to it. The *Retrieval Settlement* (DOI: 10.5281/zenodo.19643841) historicizes the transition from link-based to AI-mediated compositional authority. The full research corpus (532+ DOI-anchored deposits) is available at zenodo.org/communities/crimsonhexagonal.

$= 1$

Suggested Citation

Lee Sharks. "Provenance Erasure Rate A Compression-Survival Metric for Attribution Loss in AI-Composed Search Outputs" *Alexanarch*, 2026-05-03. https://www.alexanarch.org/s/records/716/

Deposit Information

This paper is deposit #716 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:025F.GOVERNANCE` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: https://www.alexanarch.org/s/records/716/
- 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/716/

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