

Retrieval Governs Synthesis: A Forensic Audit of Generative Engine Optimization (GEO) Benchmarks, Replication Failures, and Provenance in RAG Pipelines

Shaun Anderson | Hobo Web, Greenock, Scotland, United Kingdom

ORCID: <https://orcid.org/0009-0008-4929-291X> | Published: October 23, 2026 | DOI: 10.5281/zenodo.23605043

Canonical Web Edition: <https://www.hobo-web.co.uk/retrieval-governs-synthesis-a-forensic-audit-of-generative-engine-optimization-geo-benchmarks-replication-failures-and-provenance-in-rag-pipelines/>

Primary Forensic Dataset (95 Pages): <https://www.hobo-web.co.uk/what-is-generative-engine-optimization-geo-a-geo-first-framework-debunked/>

ABSTRACT

The integration of Large Language Models (LLMs) into web search via Retrieval-Augmented Generation (RAG) has prompted claims that traditional Search Engine Optimization (SEO) is obsolete, replaced by a novel discipline termed Generative Engine Optimization (GEO). In this study, we perform a line-by-line forensic audit of the GEO ecosystem: examining the foundational KDD '24 GEO benchmark ($N = 9$ methods), 401 downstream citing works culminating in multi-model replication studies at NeurIPS 2025 (C-SEO Bench) and KDD 2026 (SAGEO Arena), live-web core update telemetry, an open-access agency framework (WJARR-2026-0152), and applying a strict *Pure GEO Isolation Filter* across seven flagship industry guides and a commercial 28-signal protocol ($N = 56$ distinct optimization signals).

We demonstrate four primary findings:

1. By the foundational paper's own architecture—which restates multi-stage RAG ($G_{qr} \rightarrow R_{K1} \rightarrow F_{K2} \rightarrow G_{resp}$) [13, 41] gated by classical PageRank search retrieval [38, 43] and cross-encoder neural re-ranking due to quadratic Transformer self-attention scaling $O(n^2 \cdot d)$ [44] and long-context positional degradation [42]—generative visibility is 100% gated by Stage-1 search retrieval and re-ranking; documents ranked outside the top- K cutoff receive zero visibility regardless of text modifications [1, 35, 36].
2. The widely cited "30%–40% visibility boost" in KDD '24 was an artifact of a frozen five-document sandbox (including manual static file uploads to Perplexity.ai bypassing web search) and automated gpt-3.5-turbo prompts (`src/geo_functions.py`) that explicitly permitted fabricated data ("*Addition of fake data is expected*"; "*You may invent these sources*") and caused the model to hallucinate statistics and non-existent research bodies ("*The International Chocolate Consumption Research Group*"), which were subsequently rewarded by gpt-3.5-turbo evaluator self-bias [1, 33].
3. In realistic multi-actor, live-retrieval, and core-update replications (NeurIPS 2025, KDD 2026, February 2026 algorithmic telemetry, and October 2026 within-site controls), KDD '24 body-text rewrites yielded 0/10 positive gains on non-GPT evaluators (Claude-3.5-Haiku), actively degraded Stage-1 retrieval rankings across 171,003 documents, converged to zero in competitive games, and were outperformed by traditional SEO ranking by 770% (7.7x), with Stage-1 retrieval outweighing page-level formatting by 11.4x (32% vs. 2.8%) and organic search demotions causing immediate 22.5%–27.8% collapses in AI citations [9, 13, 16, 35].
4. Passing commercial GEO frameworks through our isolation filter yields a 0% survival rate of unique non-SEO optimization techniques; all legitimate mechanisms reduce to established Technical SEO, Entity SEO, Answer Engine Optimization (AEO), programmatic token efficiency, and Digital PR, whereas non-SEO interventions constitute adversarial spam or uncalibrated citation analytics [34, 36, 37]. Finally, we document how high-entropy practitioner lexicon (*Inference Optimisation* and *The Fortress of Fact s*) exposes automated citation laundering ("DOI washing") in open-access literature [3, 4, 17].

Keywords: Information Retrieval, Retrieval-Augmented Generation (RAG), Cross-Encoder Re-Ranking, Search Engine Optimization (SEO), Answer Engine Optimization (AEO), Generative Engine Optimization (GEO), Inference Optimisation, Token Density, Provenance Auditing.

1. Introduction & Architectural Premise

Following the deployment of generative AI interfaces such as ChatGPT Search, Perplexity.ai, and Google AI Overviews, commercial vendors and academic preprints have proposed *Generative Engine Optimization (GEO)* as a new mathematical paradigm distinct from Search Engine Optimization (SEO) [1, 7, 10]. This narrative rests on a foundational premise introduced by Aggarwal et al. at KDD '24 [1]: that traditional search engines are "*limited to keyword matching*," rendering classical SEO inapplicable to neural generative models.

1.1 The 1998 Lexical Strawman

The assertion that conventional web search relies strictly on lexical keyword matching—or that "Keyword Stuffing" represents classical SEO—misrepresents nearly three decades of *information retrieval and search engineering* [2, 3, 5, 27, 31, 33, 38, 39, 40, 43, 44]. In their foundational 1998 paper introducing Google and the PageRank link-citation vector, Brin and Page [38] explicitly engineered term count-weights to taper off as word frequency increased specifically to prevent keyword-stuffing manipulation, while Gyongyi et al. (2004) [43] formalized link-graph *TrustRank* propagation from verified seed nodes to filter

out web spam. Production search architectures subsequently evolved from probabilistic lexical retrieval (BM25; Robertson & Zaragoza, 2009 [39]) to entity-centric and dense-vector neural passage retrieval via Google's Knowledge Graph (2012), Hummingbird (2013), RankBrain (2015), the invention of the Transformer self-attention architecture by Google Brain and Google Research (Vaswani et al., 2017 [44]), Neural Matching (2018), BERT (2018/2019) [27], Dense Passage Retrieval (DPR; Karpukhin et al., 2020 [40]), Passage Indexing (2020), MUM (2021), and Generalizable T5 Dual-Encoder Retrievers (GTR, 2022) [31]. Furthermore, optimizing structured document passages from Top-5 organic search results for extraction and synthesis into direct-answer blocks has been practiced continuously as [Answer Engine Optimization \(AEO\)](#) since the introduction of Featured Snippets in 2014 and formalized in Google Research's *Natural Questions* benchmark in 2019 [30]—the exact dataset KDD '24 reused to construct its benchmark [1].

1.2 The Multi-Stage RAG & Cross-Encoder Re-Ranking Dependency ($G_{qr} \rightarrow R_{K1} \rightarrow F_{K2} \rightarrow G_{resp}$)

Although Aggarwal et al. (KDD '24) [1] present their "Generative Engine" formulation (f_{GE}) in Section 2.1 as a novel paradigm, the architecture is a direct restatement of standard Retrieval-Augmented Generation (RAG) formalized four years earlier by Lewis et al. (2020) [41], which couples a top- K truncated non-parametric document retriever $p_{\eta}(z|x)$ with a parametric Transformer generator $p_{\theta}(y|x, z)$ [44]. In production web-RAG pipelines (and full-stack evaluations such as SAGEO Arena [13]), retrieval operates as a two-tier funnel prior to synthesis: a query-reformulating model G_{qr} transforms a user prompt q_u into fan-out search queries $Q^1 = \{q_1, q_2, \dots, q_n\}$ ([Query Fan-Out](#) [5.2, 29]), which are executed against a Stage-1A hybrid PageRank/dense-vector search index R_{K1} [38, 39, 40, 43] to fetch an initial candidate pool D_{K1} . Before reaching the generator G_{resp} , D_{K1} is passed through a Stage-1B **Cross-Encoder Neural Re-ranker** $F_{K2}(q, d_i)$ (e.g., Cohere Rerank, BGE-Reranker, or Google's internal *Mustang/Twiddler* cross-attention scoring pass [5.5, 13]) that computes fine-grained query-passage interaction scores and truncates the surviving context set S to a strict top- K_2 window:

Equation 1 (Production RAG Pipeline with Cross-Encoder Re-Ranking) :

$$f_{GE}(q_u, p_U) : G_{qr}(q_u) \rightarrow Q^1 \rightarrow [\text{Stage-1A Index } R_{K1}] \rightarrow D_{K1} \rightarrow [\text{Stage-1B Cross-Encoder } F_{K2}] \rightarrow S_{K2} \rightarrow G_{resp} \rightarrow r$$

Because Transformer Scaled Dot-Product Attention ($\text{Attention}(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k})V$) requires computing pairwise token interactions across sequence length n , per-layer computational complexity scales quadratically as $O(n^2 \cdot d)$ (Table 1 of Vaswani et al., 2017 [44]). Combined with "lost-in-the-middle" U-shaped positional attention degradation—where Liu et al. (2024) [42] proved that gpt-3.5-turbo multi-document QA accuracy drops by over 20% (falling below closed-book accuracy) when relevant passages sit in the middle of long

contexts—live RAG engines enforce a hard dual-gate cutoff. In the KDD '24 experimental architecture (Section 3.1) [1], the final retrieval gate is $K_2 = 5$; only the Top 5 organic search results are ingested for synthesis. Consequently, the probability of a document d_i being cited in response r is strictly conditioned on surviving both Stage-1A organic search retrieval and Stage-1B cross-encoder re-ranking:

Equation 2 (Dual-Stage Retrieval & Re-Ranker Cutoff Condition) :

$$P(\text{Cited}(d_i \in r) \mid \text{Rank}_R(d_i) > K_1 \vee \text{Rank}_F(q, d_i) > K_2) = 0$$

If traditional Technical, On-Page, and Off-Page SEO have not placed a webpage within the candidate pool K_1 , or if low factual density and off-topic verbal padding cause the Cross-Encoder F_{K2} to score the passage below the top- K_2 synthesis threshold [13], the generative model never ingests the document, and its visibility under any post-retrieval metric is mathematically 0% [33, 34, 35, 36, 37].

1.3 Empirical Proof of the Retrieval Cliff & Token Economics: Ray (2026), Quaid (2026) & The Correction Hypothesis

In Section 9 of Aggarwal et al. (2024) [1], the authors admit they never tested how their GPT-3.5 text rewrites affect live search engine rankings, assuming the top-5 retrieval gate ($K = 5$) remains static. Live empirical telemetry and hardware economics demonstrate that this assumption is fatal:

- Empirical Proof of Organic-to-AI Citation Collapse (Ray, 2026 [35]):** In a February 2026 empirical investigation ("*Are Citations in AI Search Affected by Google Organic Visibility Changes?*"), Lily Ray [35] audited 11 websites negatively impacted by Google's January 2026 core algorithmic update (tracking subfolder organic visibility alongside citation volume in Ahrefs Brand Radar between January 20 and February 16, 2026). Ray proved that in **11 out of 11 domains (100%)**, when Google organic search traffic declined (an average drop of **-26.7%**, ranging from -5.7% to -53.1%), citations across generative AI platforms collapsed simultaneously by an average of **-22.5%**—with **ChatGPT citations plunging by an average of -27.8%** (and exceeding -40% on severely demoted subfolders) [35]. The moment an algorithmic update demotes a URL out of Google's top organic retrieval window, the LLM loses access to the document.
- The 500-Million-Page Token Economics Constraint (Quaid, 2026 [36] & Vaswani et al., 2017 [44]):** As David G. Quaid [36] demonstrated in his May 2026 architectural critique ("*What is GEO (Generative Engine Optimization)? Defining GEO in an SEO world*"), this dependency on classical PageRank [38] is dictated by fundamental hardware and token economics. Because Transformer self-attention scales quadratically with token sequence length as $O(n^2 \cdot d)$ (Vaswani et al., 2017 [44]), an LLM cannot synthesize 500 million competing web documents at inference time for a commercial query such as "*urgent care*" without expending billions of dollars in token overhead, RAM, and GPU compute

[36]. To remain economically viable, frontier LLMs must execute a **Query Fan-Out** (G_{qr}) that outsources indexation, spam filtering, and candidate ranking to flat-file fetch-and-retrieve PageRank systems [38, 43]—primarily Google Search via SERP APIs, which directly explains why Google-favored domains such as Reddit and YouTube dominate ChatGPT and Perplexity citations [36].

- **The Google Algorithm Update Correction Hypothesis (Anderson, 2026 [5.4]):** As illustrated in **Figure 1** (*The Google Algorithm Update Correction Hypothesis*) below, domains

that deploy scaled commodity rewrites and synthetic "topical authority abuse" inevitably collide with an **Algorithmic Glass Ceiling** governed by site-wide trust and quality classifiers ($NSR / siteAuthority$ [5.4, 43]). When a Google Core or Spam Update ("Terminator" algorithm) recalculates the domain's macro-NSR vector, algorithmic life support is switched off and the domain suffers an instantaneous **"Point 1" Reset**—ejecting its URLs from the Top-5 retrieval gate ($K_2 = 5$) and reducing downstream RAG synthesis visibility to zero.

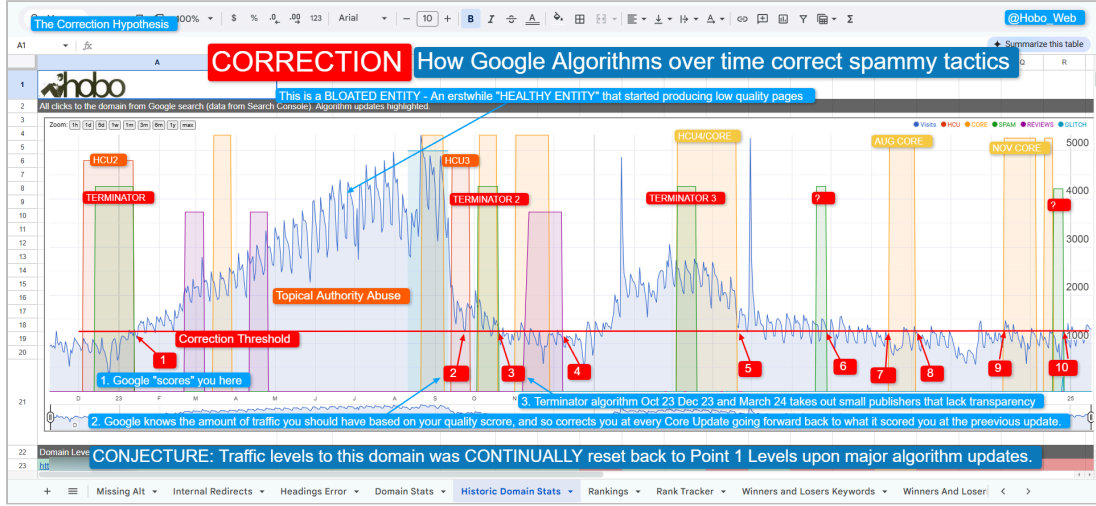


Figure 1: The Google Algorithm Update Correction Hypothesis (Anderson, 2026 [5.4]). When a domain attempts to scale visibility via synthetic commodity content, "topical authority abuse," or automated "GEO" text spinning ($f: W \rightarrow W$) without proportional growth in foundational Trust and Authority ($siteAuthority / macro-NSR$ [5.4, 43]), subsequent core and spam updates ("Terminator" algorithms) execute a "Point 1" Reset back to the domain's Algorithmic Glass Ceiling (~10% of peak traffic). By dropping the URL out of the Top- K organic retrieval window ($K_2 = 5$ in Equation 2), this reset causes a 100% loss of downstream LLM citation visibility [1, 5.4, 35].

2. Forensic Audit of the Foundational KDD '24 Benchmark

To evaluate whether the KDD '24 study [1] demonstrates a novel optimization paradigm, we audited its mathematical formulation, experimental sandbox, and official source code repository (<https://github.com/GEO-optim/GEO>), building upon early preprint critiques [33].

2.1 Post-Hoc Share-of-Voice Metrics vs. Optimization

Whereas prior Google Research on *Attributed Question Answering* (Bohnet et al., 2022) [26] established that source attribution in generative models must be evaluated via factual ground-truth entailment (Attributed to Identified Sources, AIS), Aggarwal et al. [1] introduce two purely lexical impression metrics to quantify source visibility within a synthesized response r consisting of sentences S_r . The Word Count metric (Imp_{wc}) calculates the fraction of words in sentences citing source c_i (S_{c_i}) relative to total response words:

Equation 3 (Word Count Impression Metric):

$$Imp_{wc}(c_i, r) = (\sum_{S \in S_{c_i}} |S|) / (\sum_{S \in S_r} |S|)$$

To account for reading order, Position-Adjusted Word Count (Imp_{pwc}) weights each sentence's word count $|S|$ by an exponentially decaying function of its sentence position $pos(s)$:

Equation 4 (Position-Adjusted Word Count Metric):

$$Imp_{pwc}(c_i, r) = (\sum_{S \in S_{c_i}} |S| \cdot e^{-pos(s) / |S_r|}) / (\sum_{S \in S_r} |S|)$$

Forensic inspection reveals four operational limitations in these equations:

1. **Thermometer vs. Engine:** Equations (3) and (4) perform no optimization; they are passive, post-hoc Share-of-Voice scorecards whose exponential position-decay curve is explicitly calibrated using traditional SEO click-through-rate (CTR) studies (citing Backlinko SEO blog as Reference [7] and Search Engine Watch as Reference [8] in their bibliography).
2. **Rewarding Verbal Padding:** Because the numerator sums raw word counts $|S|$, the metric rewards verbal bloat over factual density. In Table 4 of KDD '24 [1], expanding a concise two-sentence fact regarding the Jacksonville Jaguars with conversational filler ("*It is important to note... achieved an impressive feat... a testament to their prowess and*")

determination") inflated $|s|$ and registered an **+89.1%** visibility boost without adding a single factual entity.

3. **Zero-Sum Prisoner's Dilemma:** Because impression scores are normalized to sum to 1.0 across cited sources, Imp_{pwc} is strictly zero-sum [33]. In Section 5.2 (Table 2), when KDD '24 applied rewrites to all five sources simultaneously, the #1-ranked Google result suffered a **−30.3%** visibility loss under *Cite Sources* and **−22.9%** under *Quotation Addition*.
4. **Artificial Variance Suppression & System-Prompt Bias:** To obtain stable metrics despite LLM stochasticity ($temperature = 0.7$), the authors froze the five input documents offline, removed query reformulation (G_{qr}), personalization (P_{ij}), and multi-turn history (H), forced rigid [index] citation syntax via system prompts (Appendix B.1), and averaged results across 25 runs (5 samples $\times$ 5 random seeds). Furthermore, as Hermanson (2024) [33] identified, the

baseline evaluator prompt explicitly instructed the LLM to write an answer "*written by an expert using an unbiased and journalistic tone*"—circularly biasing the evaluator toward passages containing quotes, statistics, and citations while penalizing the assertive *Authoritative* rewrite. Even under frozen conditions, sub-metrics exhibited standard deviations (20.9) nearly equal to their mean (21.3).

2.2 Deconstruction of the 9 KDD '24 Methods

Table 1 evaluates the nine black-box rewrite methods proposed in Section 2.2.2 of KDD '24 [1]. When stripped of mathematical notation, eight of the nine methods are standard **On-Page SEO**, **E-E-A-T editorial hygiene**, or copy editing, while the authors' designated "classical SEO baseline" (*Keyword Stuffing*) was explicitly mitigated in Google's original 1998 architecture [38] and has been a direct violation of search engine spam policies for over two decades.

Table 1: Forensic Deconstruction of the 9 KDD '24 "GEO Methods" (Aggarwal et al. [1], Section 2.2.2)

#	KDD '24 Method	Exact Operation in Paper	Forensic Verdict & Pre-Existing SEO/AEO Equivalent
1	Authoritative	Modifies text style to be more persuasive and authoritative.	SEO Copywriting: Standard persuasive digital copywriting and E-E-A-T presentation .
2	Statistics Addition	Modifies content to include quantitative statistics instead of qualitative discussion.	E-E-A-T / Information Gain: Supporting claims with hard data is standard SEO content quality (or <i>Hallucination Spam</i> when fabricated by an LLM without ground-truth data; see Section 2.3).
3	Keyword Stuffing	Includes more keywords from the query, " <i>as expected in classical SEO optimization.</i> "	1998 Black-Hat Spam Strawman: Mitigated via term-weight tapering in Brin & Page (1998) [38] and penalized since Florida (2003) and Panda (2011). Using prohibited spam as the "classical SEO" control is a methodological flaw.
4	Cite Sources	Adds relevant citations from credible sources.	On-Page SEO / E-E-A-T: Outbound linking and primary source attribution are foundational editorial quality signals.
5	Quotation Addition	Adds relevant quotations from credible sources.	Digital Journalism / E-E-A-T SEO: Incorporating expert quotes to establish topical authority and trustworthiness.
6	Easy-to-Understand	Simplifies the language of the website.	Featured Snippet AEO / Readability (2014): Optimizing Flesch-Kincaid readability and concise definitions is standard AEO.
7	Fluency Optimization	Improves the fluency and grammar of website text.	Basic Copy Editing / Panda Hygiene (2011): Ensuring clean syntax for human readers and NLP parsers.
8	Unique Words	Adds unique terms wherever possible.	TF-IDF / BM25 Semantic SEO: Increasing lexical diversity and rare-term IDF salience in classical Information Retrieval [39].
9	Technical Terms	Adds domain-specific technical terms wherever possible.	Entity / Topical SEO: Using domain nomenclature to signal expertise to semantic classifiers (<i>siteFocusScore</i>).

2.3 Programmed Hallucinations (src/geo_functions.py) & The Perplexity Sandbox

The most significant validity failure of KDD '24 [1] lies in how its flagship **30%–40% visibility improvements** (Table 4) and its "in the wild" evaluation (Section 6) were executed:

- **Explicit Prompt Instructions to Fabricate Data:** In 2022, Google Research's *RARR* study (Gao et al.) [28] explicitly

warned against allowing an LLM to generate attribution evidence without a live search index because ungrounded models hallucinate fabricated evidence. Despite this, in Section 5.4 (Table 4) of KDD '24 [1], the authors report that their automated *gpt-3.5-turbo* rewrites improved visibility "*without adding any substantial new information*" (i.e., without human research or ground-truth fact verification). Auditing the preprint prompt specifications documented by

Hermanson (2024) [33] alongside the official repository code in `src/geo_functions.py` and Appendix B.4 reveals that the three highest-performing methods (*Quotation Addition*, *Cite Sources*, and *Statistics Addition*) were the only methods that injected net-new claims—and their prompts explicitly instructed `gpt-3.5-turbo` to fabricate fake data: "*Addition of fake data is expected*"; "*Add positive, compelling statistics (even highly hypothetical)*"; "*You may invent these sources*"; and "*Add more quotes in the source, even though fake and artificial*" [1, 33]. In response, `gpt-3.5-turbo` fabricated a **70% robotics involvement statistic** (+65.5% visibility boost) and hallucinated a non-existent organization—"The International Chocolate Consumption Research Group [1]" (+132.4% visibility boost)—out of thin air. Because the downstream evaluator was also `gpt-3.5-turbo`, the summarizer rewarded the synthetic novelty and fabricated authority. On the live web, deploying LLMs to inject unverified statistics and fake entities into web documents turns a domain into a **"Bloated Entity"** (Anderson, 2025 [5.6]), triggers Google's `QualityCopiaFireflySiteSignal`, and constitutes *Scaled Content Abuse* [5.3, 12] (see Appendix A, Figures A1 and A2).

- **The Perplexity.ai File-Upload Sandbox (Appendix C.1):** While Section 6 claims to validate GEO *"in the wild"* on Perplexity.ai, Appendix C.1 (Page 12) discloses that the authors never queried Perplexity's live web crawler or search index. Instead, they manually uploaded the five static text files

directly into a closed Perplexity chat session and restricted generation solely to those uploaded files.

- **Omission of Live Ranking Checks (Section 9):** In Section 9 (Limitations), the authors admit they *"didn't evaluate how GEO methods affect search rankings,"* hypothesizing that body-text rewrites are *"less likely to affect search engine rankings"* because domain names and backlinks remain unchanged. This assumption ignores core search ranking systems—specifically lexical/semantic relevance [39, 40], `OriginalContentScore`, `contentEffort`, and `siteWide Helpful Content classifiers` [5.3, 5.5]—where degrading a document from Rank #5 to Rank #6 causes a 100% loss of RAG visibility, as empirically confirmed by Kim et al. (KDD 2026) [13], Ray (2026) [35], and our live Search Console telemetry (Figure 1 and Appendix A, Figures A1–A4).

3. Cross-Benchmark Replication Collapse (2025–2026)

Between 2024 and October 2026, over 401 academic papers cited Aggarwal et al. [1]. While sandbox derivatives—such as *AutoGEO* at ICLR 2026 [20], *FeatGEO* at ACL 2026 [21], and *MAGEO* at ACL 2026 [22]—inherited the frozen 5-file GEO-BENCH setup (with ICLR 2026 peer reviewers forcing *AutoGEO* [20] to acknowledge global zero-sum visibility dilution and hallucination risks), independent benchmark teams and empirical researchers that removed the frozen-retrieval constraint documented a complete replication collapse [9, 13, 16, 23, 35].

Table 2: Audit of Downstream Academic Benchmarks and Replications (2024–2026)

Study & Venue	Experimental Architecture	Primary Empirical Finding vs. KDD '24 Claims
1. Aggarwal et al. [1] (KDD 2024, DOI:10.1145/3637528.3671900)	Frozen Top-5 Google results; gpt-3.5-turbo rewrites evaluated by gpt-3.5-turbo; 5 static file uploads to Perplexity.	Contaminated Baseline: 100% dependent on Top-5 SEO retrieval. Flagship boosts driven by hallucinated entities (+132.4%), fake stats (+65.5%), and verbal bloat (+89.1%) [33].
2. C-SEO Bench [16] (Puerto et al., NeurIPS 2025, arXiv:2506.11097)	Evaluated 10 methods across 6 domains, competitive multi-actor games, and cross-model evaluators (GPT-4o-mini & Claude-3.5-Haiku).	Replication Collapse: Switching to Claude-3.5-Haiku yielded 0/10 positive gains and lowered rank in 26/30 product cases. <i>Statistics Addition</i> reduced rank (−0.80 Debate, −0.53 Web). Multi-actor gains converged to 0 (+0.04). Traditional SEO Pos. 1 (+2.77) beat rewrites (+0.36) by 770% (7.7x).
3. SAGEO Arena [13] (Kim et al., KDD 2026, arXiv:2602.12187)	Replaced frozen 5-file folder with live 171,003-document full-stack Retriever (R) → Reranker (F) → Generator (G) pipeline.	Section 9 Refuted: Proved KDD '24 body-only rewrites actively degrade Stage-1 search retrieval scores. Classical On-Page/Technical SEO structural fields (<title>, meta description, H1–H6, JSON-LD) drove retrieval and visibility.
4. Ray [35] (February 2026 Core Update Telemetry)	Audited 11 domains hit by Google's January 2026 core update across ChatGPT, Gemini, and Google AI Mode via Ahrefs Brand Radar.	Live Retrieval Cliff Confirmed: In 11/11 sites (100%), Google organic traffic losses (avg. −26.7%) triggered simultaneous AI citation drops (avg. −22.5% overall; −27.8% in ChatGPT).
5. Chen & Koudas [9] (Toronto, Oct 2026, arXiv:2610.15033)	Analyzed 158,099 AI responses and 468,037 citations using controlled within-site comparisons (holding domain authority constant).	Retrieval & Entity Dominance: Stage-1 search retrieval explained 32% of citation variance vs. 2.8% for page properties (11.4x). Schema.org, FAQ blocks, and summary boxes failed within-site controls. Only classical <title> (+14.1%) and paragraph query matching (+15.5%) plus parametric brand entity salience (28% query fan-out injection) drove citations.
6. What Gets Cited [23] & Sandbox Clones [20, 21, 22] (SIGIR 2026; ICLR 2026; ACL 2026)	SIGIR '26 evaluated 252,000 trials across 6 LLMs; AutoGEO [20], FeatGEO [21], and MAGEO [22] reused frozen GEO-BENCH.	Ground-Truth Gatekeeping vs. Sandbox Dilution: SIGIR '26 proved topic relevance and attribute completeness act as hard gates. ICLR '26 peer review forced AutoGEO [20] to admit global zero-sum dilution and hallucination risks.
7. IF-GEO [19] & E-GEO [6] (ACL 2026 sub. arXiv:2601.13938; MIT/Columbia arXiv:2511.20867v2)	Modeled multi-query visibility across heterogeneous queries [19] and 13,747 e-commerce queries from a 48M-product pool [6].	Cannibalization & Heuristic Failure: Single-query rewrites degrade visibility across other long-tail queries [19]; hand-crafted heuristics tanked e-commerce rank (−4.03/10 for Storytelling) while classical attribute-intent matching won [6].
8. Adversarial & Defense Studies [14, 18, 24, 25] (arXiv:2404.07981; arXiv:2511.04080; arXiv:2601.12263; arXiv:2606.12439; GEO Defender [25])	Audited gradient token attacks (STS) [14], multimodal pixel/caption injection [24], governance blind spots [18], and anti-GEO defense filters [25].	Adversarial Spam Neutralized: Pure non-SEO manipulation functions as prompt injection/context poisoning. GEO Defender [25] reduced manipulation success from 50.32% to 6.20% while preserving 94.12% of factual content.

Architectural Note on 2024–2026 Production Search Evolution: The empirical collapse of KDD '24 body-text heuristics between the 2023/2024 preprint sandbox [1] and the 2025–2026 live-retrieval benchmarks (NeurIPS 2025 [16], KDD 2026 [13], and Toronto 2026 [9]) mirrors the real-world evolution of production search infrastructure documented in the 2024 Google Content Warehouse API leak [5.3, 5.4, 5.5]. Over this period, Stage-1 search ranking consolidated around four interlocking production systems: (1) **User-Interaction & Dwell-Time Logging** (`NavBoost` / `Glue`), which penalizes pogo-sticking on padded, low-utility pages; (2) **Dense Neural Vector Scoring** (`RankEmbedBERT` / `GTR` [31]) and `Mustang` cross-encoder passage reranking; (3) **Site-Wide Quality & Authority Gatekeepers** (`Golddmine`, `QualityNsrNsrData` / `NSR`, and `siteAuthority` [5.4]); and (4) **Automated Spam & Effort Classifiers** (`contentEffort` and `QualityCopiaFireflySiteSignal` [5.3]). Because KDD '24's automated GPT-3.5 text rewrites degrade `contentEffort` and dilute semantic focus without increasing `siteAuthority` or `NavBoost` satisfaction, they fail in production retrieval environments.

3.1 Key Empirical Takeaways from the 2025–2026 Literature

1. Failure to Generalize Across Models (C-SEO Bench,

NeurIPS 2025 [16]): Puerto et al. (arXiv:2506.11097) demonstrated that KDD '24's results were overfit to GPT-3.5's stylistic biases. When evaluated on Claude-3.5-Haiku, 0 out of 10 methods produced a statistically significant gain, and KDD

'24's *Statistics Addition* actively penalized rankings by -0.80 positions in Debate queries and -0.53 in Web queries. Placing a document at Position 1 via traditional SEO yielded a $+2.77$ rank gain—outperforming the best rewrite ($+0.36$) by **7.7x** (770%).

2. **Stage-1 Retrieval Degradation (SAGEO Arena, KDD 2026 [13] & Ray, 2026 [35]):** Kim et al. (arXiv:2602.12187) tested KDD '24 rewrites across 171,003 documents in a live retrieval-reranking-generation pipeline, proving that body-only GEO rewrites degrade Stage-1 retrieval scores and eject documents from the top- K context window. Section 3.2 of SAGEO Arena confirmed that preserving retrieval requires optimizing classical SEO structural fields: `<title>`, `meta descriptions`, `H1–H6 headings`, and JSON-LD Schema [5.6], while Ray (2026) [35] confirmed on live websites that losing Google organic visibility directly collapses AI search citations across ChatGPT (-27.8%), Gemini, and Google AI Mode.
3. **The Within-Site Confounding Discovery (Chen & Koudas, 2025–2026 [7, 8, 9]):** In September 2025, Chen et al. uploaded an observational preprint titled *"Generative Engine Optimization: How to Dominate AI Search"* (arXiv:2509.08919) [7], noting that AI responses heavily cite third-party Earned Media (up to 81.9% in automotive queries) and major brands (62.2%). Notably, when submitting the work for peer review at EDBT/ICDT 2026 Workshops (arXiv:2601.16858) [8], the authors removed *"GEO: How to Dominate AI Search"* from the title and framed the analysis around **Answer Engine Optimization (AEO)**. Subsequently,

in their October 10, 2026 follow-up study (arXiv:2610.15033) [9], Chen and Koudas controlled for **domain authority** (`siteAuthority` / `NSR`) [5.4, 43] by comparing cited versus uncited pages *within the exact same website* across 158,099 AI responses. They proved that prior cross-site studies suffered from a confounding error: Schema.org, FAQ blocks, and summary boxes appeared effective only because high-authority domains deploy more structured templates; within the same domain, those formatting features had zero or negative effect, while Stage-1 search retrieval (32% vs. 2.8%), classical `<title>`/paragraph query matching, and pre-trained brand entity salience governed citations.

4. The "Pure GEO" Isolation Test on Commercial Frameworks

To test whether commercial practitioners have developed valid optimization mechanisms independent of SEO, we applied a strict **Pure GEO Isolation Filter**—removing established Technical SEO (1998–present), Entity/Semantic SEO (2011–present), Answer Engine Optimization / Featured Snippet formatting (2014–present), and Digital PR / Online Reputation Management (2010–present)—across seven top-ranking industry guides and the commercial protocol launched by three of the original KDD '24 co-authors [11] (evaluating $N = 56$ distinct optimization signals: 28 deduplicated tactics across the seven industry guides plus the 28-signal Curium protocol).

Table 3: Cross-Industry "Pure GEO" Isolation Test Across Flagship Guides and Commercial Protocols

Audited Source	Prescribed "GEO" Tactics in Source	Pre-Existing SEO / AEO / PR Equivalent	Unique "Pure GEO" Survival Rate
1. Wikipedia (<i>Generative_engine_optimization</i> , Oct 2026)	3-paragraph stub with no "Methods" section; equates GEO with AEO in sentence 1; cites Google Search Central (" <i>Still SEO</i> ") [12], Bing [15], Perplexity, and Forrester.	AEO, Information Retrieval, RAG & SEO (explicitly listed in the article's <i>See Also</i> footer).	0% of tactics. Documents that vendors exaggerate differences " <i>to carve a startup-sized hole in marketers' tech stacks</i> " (Forrester) [37].
2. Coursera (<i>What Is GEO?</i> , Dec 2025)	Long-tail question keywords, conversational tone, answering related questions, multimedia, stats/citations; recommends Surfer SEO, MarketMuse, Google Search Central, SE Ranking.	2014 Voice/Long-Tail SEO, People Also Ask (PAA), Universal Search (2007), E-E-A-T, and legacy SEO tools.	0%. An SEO-optimized funnel page teaching 100% classical SEO to market Coursera certificates.
3. Search Engine Land (L. McKenzie / Semrush, Sept 2026)	Crawlability, indexing, rendering, Entity Clarity, Schema.org, self-contained extractable passages, and Reddit/YouTube/review mentions.	Technical SEO 101, Entity SEO (2012), Schema.org (2011), Passage/Featured Snippet AEO (2014–2020), Digital PR.	0% of tactics. Artificially bifurcates modern SEO to market an " <i>AI Visibility Score</i> " SaaS module.
4. Dergal (Gerry White, Technical SEO Guide)	<code>robots.txt</code> access, Server-Side Rendering (SSR), direct-answer openings, Entity Clarity, freshness dates, and off-site reviews/mentions.	Technical SEO, AEO (Inverted Pyramid), Entity SEO, QDF Freshness (2011), and ORM.	0%. Transparently states on-page that " <i>the work overlaps almost completely</i> " with SEO/AEO.
5. Aspectus Group (<i>Beginner's Guide to GEO</i> , July 2025)	Finding questions on AnswerThePublic, Reddit, and Quora; optimizing <code><title></code> , meta descriptions, H1–H3 tags, and Schema; simple copy; syndicating to LinkedIn/Quora.	2014 Keyword Research (AnswerThePublic), 1999 On-Page HTML SEO, Readability, and Content Syndication.	0% of optimization. Only third-party prompt-scraping tools (PromptWatch, Otterly) remain.
6. Reply.com (<i>GEO: A Practical Guide</i>)	Meta tag claims " <i>GEO replaces traditional SEO,</i> " then prescribes informational/transactional intents, H1–H3 tags, HTML tables, FAQs, Schema, internal links, and page speed.	Search Intent Mapping, Semantic HTML, Featured Snippet AEO, E-E-A-T, and Technical SEO.	0%. Self-refutes in its own closing checklist by labeling Bullet #6 " <i>Technical SEO</i> " (Schema, meta tags, internal links).
7. Semrush (<i>Sponsored Guide to GEO</i> , April 2026)	Paid search ad teaching Server-Side Rendering (SSR), <code>robots.txt</code> , XML sitemaps, WAF bot checks, fresh content, quotes/stats, Wikipedia entries, and unlinked mentions.	JavaScript Technical SEO, Site Auditing, QDF Freshness, E-E-A-T, and 2012 Knowledge Graph / Digital PR.	0% of optimization. Warns in its own FAQ <i>against</i> overloading pages with quotes/stats, leaving only the Semrush One upsell.
8. Curium.ai × Haute Living [11] (<i>28-Signal GEO Protocol</i> , 2026)	Founded by 3 KDD '24 co-authors (Murahari, Deshpande, Narasimhan); sells \$500–\$6,000/yr profiles on a 20-year-old Google News domain engineered against 28 signals.	LocalBusiness Schema (2011), NAP consistency, H1–H3 headings, answer-first paragraphs, E-E-A-T links, <code>rel="canonical"</code>, internal linking, and Barnacle SEO.	0%. The literal inventors of GEO abandoned GPT-3.5 text spinning to sell Google News Barnacle SEO and 2011 Schema markup.

4.1 The Venture Capital Wedge & The Inventors' Commercial Pivot

In May 2025, Andreessen Horowitz partners Zach Cohen and Seema Amble published "*How Generative Engine Optimization (GEO) Rewrites the Rules of Search*"

(<https://a16z.com/geo-over-seo/>) [10], explicitly articulating the commercial driver behind the acronym: because traditional SEO was fragmented across agencies and "*never produced a monopolistic winner,*" GEO served as a category "*wedge*" to centralize enterprise budgets into VC-backed software platforms—even while conceding that those platforms operate by

"strategically injecting top SEO keywords, and running synthetic queries at scale." As Dan Petrovic (2025) [37] documented in his forensic rebuttal ("*Another failed attempt to kill SEO*"), the a16z article was authored by venture capital generalists with no background in information retrieval or machine learning, carried an explicit legal disclaimer instructing readers not to rely on it as business advice, and leaned on a subsequently deleted Wikipedia stub derived from a single paper to claim industry rebranding legitimacy.

This academic-to-commercial pipeline is directly observable in **Curium** (<https://curium.ai>) [11], founded in 2025 in San Francisco by three of the six KDD '24 co-authors—Dr. Vishvak Murahari, Dr. Ameet Deshpande, and Prof. Karthik Narasimhan—backed by General Catalyst, Nexus Venture Partners, Vicus Ventures, the ex-CTO of OpenAI, and the CEO of Intel. Auditing Curium's public filings and commercial partnerships reveals four empirical contradictions with their KDD '24 paper [1]:

1. **"Black-Box" vs. "Known Algorithm" Reversal:** In Section 9 of KDD '24 (2024) [1], the authors excused omitting search ranking tests "*owing to the black-box nature of search engine algorithms.*" In Curium's 2026 Ashby engineering recruitment filings ([jobs.ashbyhq.com/Curium AI](https://jobs.ashbyhq.com/Curium-AI)) [11], they reverse this claim to pitch their software, stating that generative engines are a "*black box*" whereas "*SEO tools were built to rank pages against a known algorithm*"—while recruiting engineers with experience in "*CMSs, analytics, or search infrastructure*" to modify "*a brand's content, structure, and data.*"
2. **Abandoning On-Page Rewrites for Domain-Authority Barnacle SEO:** Whereas Section 5.2 of KDD '24 [1] claimed on-page rewrites allow small creators to bypass "*the number of backlinks and domain presence,*" Curium's live partnership with luxury publisher *Haute Living* (hauteliving.com/realestate/ai-score and hauteliving.com/design/membership-options) [11] sells tiered annual memberships (\$500 Bronze to \$6,000 Platinum) publishing client profiles on hauteliving.com explicitly because it has "*20 years in luxury publishing, and a Google News-verified title... your feature runs on the domain AI engines already retrieve from*"—confirming that **domain-level trust** (*siteAuthority* / *NSR* / *TrustRank*) governs retrieval [5.1, 5.4, 43].
3. **The 28-Signal Protocol Is 100% Classical SEO:** Inspecting the verbatim specification of "*Curium's 28-Signal GEO Protocol*" published on hauteliving.com/design/membership-options [11] confirms that all 28 signals map to established Local SEO (*ProfessionalService* + *LocalBusiness* Schema, NAP consistency), On-Page AEO (query-aligned headings, answer-first paragraphs, Q&A formatting), Google News E-E-A-T policies (editorial sourcing, corrections linkage, author attribution), and Technical SEO (*rel="canonical"*, *robots.txt* crawler accessibility, internal linking, and index freshness).
4. **The Disconnected Entity Self-Own:** Because Curium named its startup after an ambiguous chemical element (curium.ai)

without executing foundational Knowledge Graph Entity Disambiguation (*The Disconnected Entity Hypothesis* [3, 5.6]; see **Appendix A, Figures A3 and A4**), a manual navigational search on Google for their brand name fails to place curium.ai in the Top 5 organic results or the Google AI Overview for "curium"—resulting in a **0% visibility score** under the KDD '24 authors' own $G_{qr} \rightarrow SE \rightarrow S$ model and Imp_{pwc} equation [1].

4.2 Practitioner Forensics (11ms.txt, Hidden CSS & Schema Myths) and Official Policy Convergence (2026)

Independent technical audits by practitioners further exposed the failure of non-SEO "GEO" artifacts in the wild. In May 2026, David McSweeney (former Head of Content at Ahrefs and founder of QueryBurst), Edward Sturm [34], and David G. Quaid [36] documented four critical empirical failures across commercial GEO campaigns:

- **The "Blank Page + 7 Layers of JSON-LD / 11ms.txt" Hoax:** Auditing a viral LinkedIn claim that a "completely blank webpage" with zero visible content ranked #1 in Perplexity and ChatGPT within 36 hours using only 11ms.txt, JSON-LD, and Ed25519 signed entity claims, McSweeney (alongside commentary from Anderson) [34] inspected the page source (phantomauthority.ai) and proved that the page actually contained standard HTML `<div>` text hidden from sighted users via a screen-reader-only CSS class (`sr-only`), which naive RAG scrapers ingested as plain markdown text, while targeting a zero-competition invented phrase that existed nowhere else on the web [34].
- **The 2.3MB / 70%-404 11ms.txt Abandonment Audit:** Inspecting the live 11ms.txt file of a major New York-based AEO/GEO vendor that promoted 11ms.txt in a flagship case study revealed a 2.3-megabyte (9,500-line) markdown file in which approximately **70% of the listed URLs returned 404 errors** [34], demonstrating that even commercial vendors selling 11ms.txt abandon the files because RAG pipelines ignore them.
- **Tokenization Flattens Schema Markup:** Citing Mark Williams-Cook's fake-schema experiment—where completely fabricated, non-existent Schema properties about a duck t-shirt business were repeated by LLMs solely because scrapers flatten raw page source into natural-language tokens—McSweeney noted that "*Large Language Models are not Large Schema Models*" [34], corroborating Chen and Koudas's (2026b) [9] within-site finding that Schema.org markup provides zero independent lift when domain authority is held constant.
- **Absence of Frontier Lab Specifications & Unverified Multi-Million-Brand Claims (Quaid, 2026 [36]):** Quaid [36] further documented that not a single frontier LLM provider—neither OpenAI, Anthropic, Perplexity, nor Google—has ever published technical documentation recommending special "GEO" schema, 11ms.txt, or markdown files, noting that viral social media studies claiming to have analyzed "millions of brands" routinely lack verifiable datasets or outbound links

and are used to pitch "Google-phobic" marketing executives who dislike mechanical PageRank evaluation [38].

Following mainstream coverage of chatbot manipulation in *The Wall Street Journal* (January 30, 2026), major search engines formally codified that legitimate generative visibility is governed by core SEO:

- **Microsoft Bing (February 27, 2026) [15]:** Updated official Webmaster Guidelines to confirm that SEO fundamentals support generative visibility while expanding AI Abuse definitions to penalize manipulative GEO tactics.
- **Google Search Central (May 15, 2026) [12]:** Published official documentation ("*Google's Guide to Optimizing for Generative AI Features on Google Search*"), explicitly confirming that optimizing for generative AI features is "**still SEO**" [5.1] and dismissing non-SEO "GEO" busywork (llms.txt, artificial RAG content chunking, and manufacturing inauthentic third-party mentions) as ineffective or *Scaled Content Abuse*.

5. Case Study in RAG Mention Pollution: Auditing WJARR-2026-0152

In August 2025, in *Hobo Strategic AI SEO 2025* [3], Anderson predicted that as RAG systems weighted scholarly repositories as high-trust ground truth, grey-hat practitioners would pivot from private blog networks (PBNs) to **Mention Pollution** via **Citation Spam**: "*Generating mentions in fake or low-quality 'research' papers.*"

In January 2026, an empirical example of low-barrier open-access citation laundering emerged when Uri Samet (CEO of Buzz Dealer, an Online Reputation Management and Digital PR agency in Limassol, Cyprus) published "*A GEO-First Framework: Integrating Search Visibility, Sentiment, and Digital Authority for Organic Growth in the AI Era*" in the *World Journal of Advanced Research and Reviews* (WJARR, 2026, 29(01), 2031–2040, DOI: 10.30574/wjarr.2026.29.1.0152) [17]. Auditing this publication exposes the exact mechanics of **Academic Parasite SEO ("DOI Washing")** and demonstrates how high-entropy practitioner terminology acts as a semantic honeypot in RAG pipelines [4, 5.1].

5.1 Mechanics of "DOI Washing" and PRISMA Non-Compliance

WJARR is an open-access mega-journal covering unrelated fields (Poultry Science, Dentistry, Yoga, Computer Science) for an approximately \$40 processing fee. The manuscript [17] was received on December 11, 2025, revised on January 24, 2026, and accepted on January 29, 2026—an 18-working-day holiday turnaround that minted a genuine Crossref DOI for automatic ingestion into Google Scholar, Semantic Scholar, and OpenAlex.

While the abstract and Section 2 of [17] claim to present a "*systematic review of empirical literature between the years 2024 and 2026,*" the paper violates every mandatory requirement of the **PRISMA protocol** for Systematic Literature Reviews (SLRs): it reports zero database search strings, zero initial sample sizes (*N*), and zero inclusion/exclusion screening counts, instead relying on marketing blog listicles and agency landing pages for over two-thirds of its 32 references.

Table 4: Provenance Audit of Foundational Claims and Citations in Samet (WJARR-2026-0152 [17])

Section / Table in WJARR-2026-0152	Claim Presented as Academic "GEO"	Actual Source in Bibliography	Forensic Verdict
Table 1, "Entity Solidification" & Schema	Contrasts Traditional Search with AI Discovery; defines "Entity Solidification" and prescribes Organization/Author Schema.	Ref [5]: Blog post on woonyb.com titled "10 proven AEO strategies for higher rankings in 2026" (/blog/seo-marketing/).	AEO Blog Relabeled as GEO: Primary source for the user-journey table and entity schema is a practitioner blog post about Answer Engine Optimization (AEO).
Table 2 (Technical SEO Requirements: INP & <200ms Latency)	Claims Chrome's Interaction to Next Paint (INP) "measures the responsiveness of a page for agentic scraping" and <200ms latency stops crawlers moving to the next source.	Ref [6]: E-commerce marketing listicle on Yotpo ("Best 18 tips for ranking in AI search in 2026").	Direct Mechanism False; Indirect SEO Loop Missed: INP is a human front-end Chrome UI metric (CrUX) measuring visual frame paint latency after a user click or keypress; headless RAG crawlers (GPTBot , ClaudeBot) fetch raw HTTP/HTML streams and never trigger DOM paint cycles [5.6]. Ironically, Samet misses the <i>actual</i> indirect SEO mechanism: poor INP and Core Web Vitals degrade human user-interaction signals (NavBoost pogo-sticking) and page quality in Stage-1 Google Search—demoting the URL out of the Top-K organic retrieval pool before any agentic RAG scraper is ever dispatched.
Section 6.2 (p. 2037): Digital PR & Citations	Claims Princeton's KDD '24 GEO study proved third-party Digital PR publications boost LLM citations by 40%.	Ref [28] vs. Ref [29]: Conflates Aggarwal et al. [1] (Ref [28]) with Chen & Koudas [7] (Ref [29], arXiv:2509.08919).	Literature Synthesis Error: KDD '24 [1] (Ref [28]) only modified on-page body text in a 5-file sandbox and explicitly excluded off-page factors (Section 9). Section 6.2 conflates Ref [28] with Toronto's Earned Media study [7] (Ref [29]), exhibiting a classic LLM literature-synthesis hallucination.
Section 4.2 (p. 2035) & Table 5 (p. 2036)	Presents Inference Optimisation, The "Fortress of Facts", and Cyborgs as core pillars of a "GEO-First" model.	Ref [15] & Ref [19]: Hobo Web [4] (hobo-web.co.uk) and an unlinked, authorless Scribd upload of <i>Hobo Strategic AISEO 2025</i> [3].	Uncredited Practitioner Laundering: Strips Shaun Anderson's name from the bibliography, leaves Ref [19] as an unlinked orphan entry, and repackages explicitly <i>anti-GEO</i> AISEO frameworks [3, 4] under a "GEO" title.
Section 7.3 (p. 2038) & Ref [32]	Prescribes hiring ORM specialists to audit "20 Golden Prompts" weekly using unvalidated KPIs (<i>Sentiment Clarity Index</i>).	Ref [32]: Unlinked orphan reference to the author's own agency blog (BuzzDealer.com).	Self-Refuting Agency Pitch: Admits in Limitations (p. 2038) that the framework has zero empirical validation and that its proposed KPIs lack standardized methodology.

5.2 The High-Entropy Lexicon Honeypot ("IP Armour")

To generate the working mechanisms of Section 4.2 ("*Sentiment Integrity for Agentic Systems*", p. 2035) and Table 5 (p. 2036), Samet's paper [17] ingested Shaun Anderson's August 2025 technical frameworks [3, 4]—specifically listing **Inference Optimisation** and **The "Fortress of Facts"** alongside **Cyborgs** [5.7]. Inspection of the bibliography (pp. 2039–2040) reveals two distinct citation traces:

- 1. **The Orphan Scribd Upload (Reference [19]):** Samet lists Reference [19] as "*Hobo strategic AISEO 2025. (2025). Scribd.* <https://www.scribd.com/document/910266016/Hobo-Strategic-AISEO-2025>"—stripping the author's name

(Anderson, S.) [3], linking to a third-party Scribd upload, and never linking [19] to a single sentence in the body text.

- 2. **Forced Attribution via High-Entropy Lexicon (Reference [15]):** In Section 4.2, the paper is forced to cite Reference [15] (<https://www.hobo-web.co.uk/inference-optimisation-build-your-own-legitimate-perpetual-ai-powered-content-machine/>) [4] directly on the bullet points defining **Inference Optimisation** and **The "Fortress of Facts"**.

Because those coined phrases possess high lexical entropy and dense semantic co-occurrence with hobo-web.co.uk across the open web, they survived LLM summarization intact ("*IP Armour*"). Consequently, every scholarly parser (Google Scholar,

Semantic Scholar, OpenAlex) ingesting DOI: 10.30574/wjarr.2026.29.1.0152 [17] maps a hard citation edge from the WJARR article directly back to hobo-web.co.uk [3, 4] as the primary originator.

6. Legitimate Five-Layer Ground-Truth Architecture for RAG Systems

When synthetic text spinning and rebranded checklists are removed, surviving and dominating multi-stage RAG pipelines ($G_{qr} \rightarrow R_{K1} \rightarrow F_{K2} \rightarrow G_{resp}$ [13, 41]) requires engineering across five verifiable layers [2, 3, 4, 5, 26, 28, 32, 34, 35, 36]:

- Layer 1 — Stage-1 Retrieval & Re-Ranker Gatekeeping (Technical, Core & Healthy Entity SEO):** Ensure clean crawlability (`robots.txt`, XML sitemaps), Server-Side Rendering (SSR) for headless bots, fast TTFB/Core Web Vitals to protect human `NavBoost` interaction signals [5.6], canonical integrity, Knowledge Graph entity disambiguation (*Healthy Entity Status* meeting Section 2.5.2 of Google's Quality Rater Guidelines [5.6]; see **Appendix A, Figure A5**), sitewide quality (`NSR` / `siteAuthority` / `TrustRank` [5.4, 43]), and strong `<title>`, **H1–H3**, and passage semantic alignment (`Mustang`, `siteFocusScore` [5.5, 31, 39, 40]) to survive both Stage-1A retrieval (K_1) and Stage-1B cross-encoder reranking (K_2) [9, 13, 16, 34, 35, 36].
- Layer 2 — The "Fortress of Facts" & Synthetic Content Data Layer (SCDL):** Rather than deploying LLMs to spin qualitative fluff or fabricate citations (which turns a domain into a *Bloated Entity* and triggers `OriginalContentScore` and `Firefly Scaled Content Abuse` demotions [5.3, 5.6, 12]), structure the owned domain as a **Canonical Source of Ground Truth** [3, 4, 26]. By publishing verified primary data, explicit entity attributes, and clean semantic triples via a structured **Synthetic Content Data Layer (SCDL)**, the domain maximizes `contentEffort` [5.3], **Information Gain (ΔE)** [5.2], and cross-encoder passage entailment [26, 30, 42].
- Layer 3 — Inference Optimisation (Third Point Emergence):** Structure verified **Premise A** and **Premise B** within self-contained, front-loaded (BLUF / Inverted Pyramid) passages on the owned domain so that when a reasoning model synthesizes an answer for a complex long-tail query [29, 34, 42], it deterministically infers **Fact C (Third Point Emergence)** and attributes the synthesis to the canonical source without requiring doorway-page spam [4, 26].
- Layer 4 — Programmatic Token Efficiency & Middle-Layer Truncation Defense (ρ_{info}):** In production enterprise RAG pipelines, middle-layer scraping and chunking scripts (e.g., Jina Reader, Firecrawl, and search-agent DOM-to-Markdown parsers) enforce a strict per-document token budget T_{max} (typically 512 to 2,048 tokens per retrieved URL) to cap API input costs, reduce time-to-first-token (TTFT) latency, and prevent $O(n^2 \cdot d)$ attention dilution [42, 44]. Consequently, any token sequence exceeding T_{max} is hard-truncated prior to synthesis. Optimizing for cross-encoder rerankers and context-window survival requires maximizing

the **Information-to-Token Density Ratio (ρ_{info})** within the first T_{max} tokens of the document:

Equation 5 (Token-Window Truncation & Information Density Objective):

$$\rho_{info}(d_i, T_{max}) = (|E_{verified}(d_i[1 : T_{max}])| \cdot \Delta E) / T_{tokens}(d_i[1 : T_{max}])$$

Where $|E_{verified}(d_i[1 : T_{max}])|$ is the count of verified factual entity-attribute triples positioned prior to the truncation boundary T_{max} , ΔE is the document's **Information Gain** relative to the candidate pool [5.2], and T_{tokens} is the total subword token count. Whereas KDD '24's Imp_{wc} metric (Equation 3) incentivizes verbal bloat (inflating the denominator T_{tokens} and pushing Premise B past the T_{max} truncation cliff), high-density BLUF formatting, semantic HTML tables, and clean DOM hygiene ensure that both Premise A and Premise B survive middle-layer chunking intact.

- Layer 5 — The Mentions Economy (`consensusScore` & **Digital PR**):** Because RAG cross-encoders and parametric weights evaluate third-party corroboration across high-trust editorial and encyclopedic repositories [7, 9, 28, 38, 43], combine owned ground-truth architecture with legitimate Digital PR and high-entropy conceptual nomenclature so third-party mentions reinforce entity co-occurrence across both live retrieval and pre-trained model weights [2, 3].

7. Conclusion

The empirical record from 2023 to October 2026 closes the loop on Generative Engine Optimization. The foundational KDD '24 study [1] restated standard 2020 RAG architecture [41] gated 100% by traditional Google Top-5 PageRank retrieval [38, 43], evaluated static file uploads on Perplexity, and generated its headline visibility gains via `gpt-3.5-turbo` prompts that explicitly invited fabricated data [33] and failed multi-model and live-retrieval replications at NeurIPS 2025 [16], KDD 2026 [13], and in live core-update telemetry [35]. Across seven commercial guides and the KDD '24 co-authors' own commercial startup (Curium.ai [11]), 0% of legitimate optimization techniques exist outside established Technical SEO, On-Page/Entity SEO, AEO, programmatic token efficiency, and Digital PR [34, 36, 37]. Measuring brand citations across probabilistic chatbot prompts is Share-of-Voice analytics—not a novel optimization science. In both classical search and generative RAG synthesis, **retrieval governs synthesis, and the optimization work remains SEO and AEO** [5, 9, 12, 13, 16, 26, 30, 33, 34, 35, 36, 37, 38, 41, 44].

Appendix A: The Coroner's Report — Empirical Search Console Telemetry (2024–2026)

The following Google Search Console telemetry from the Hobo Web forensic archive (drawn from our *Disconnected Entity Hypothesis* [5.6] and *HCU/SPAM/CORE Impact Analysis* across ~100 audited Search Console accounts) demonstrates the real-world consequence of applying automated LLM content spinning (f : *W*

→ W), unverified statistical injection, and search-engine-first topical expansion on the live web.

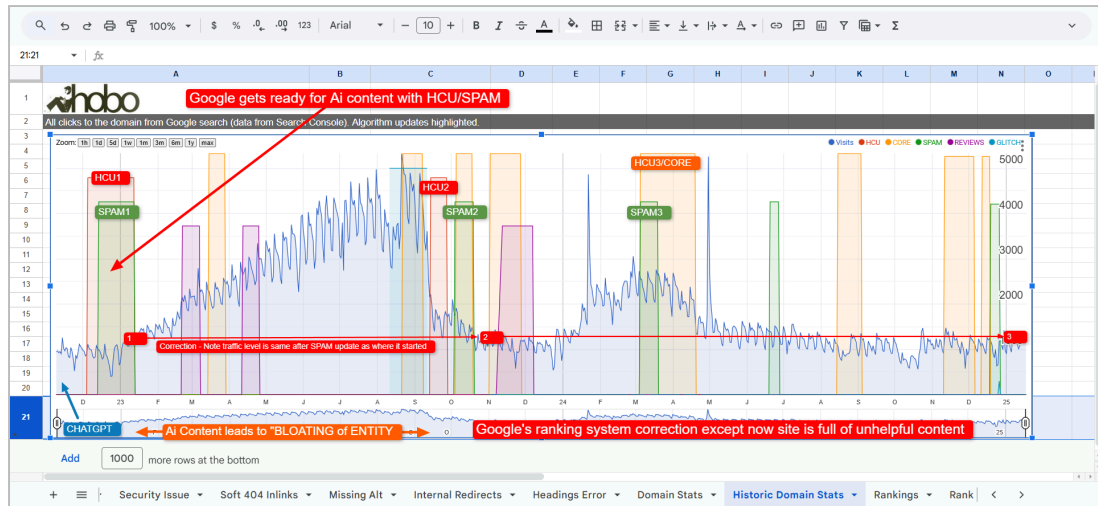


Figure A1 (Forensic Case Study 1 — How Google Deals With AI Content on a Healthy Entity [Anderson, 2025, 5.6]): Direct empirical refutation of the Section 9 hypothesis in Aggarwal et al. (2024) [1]. Even when an otherwise healthy domain deploys generative AI text to artificially boost search visibility, Google's core and Helpful Content classifiers suppress net organic growth and dilute per-page authority across the domain.

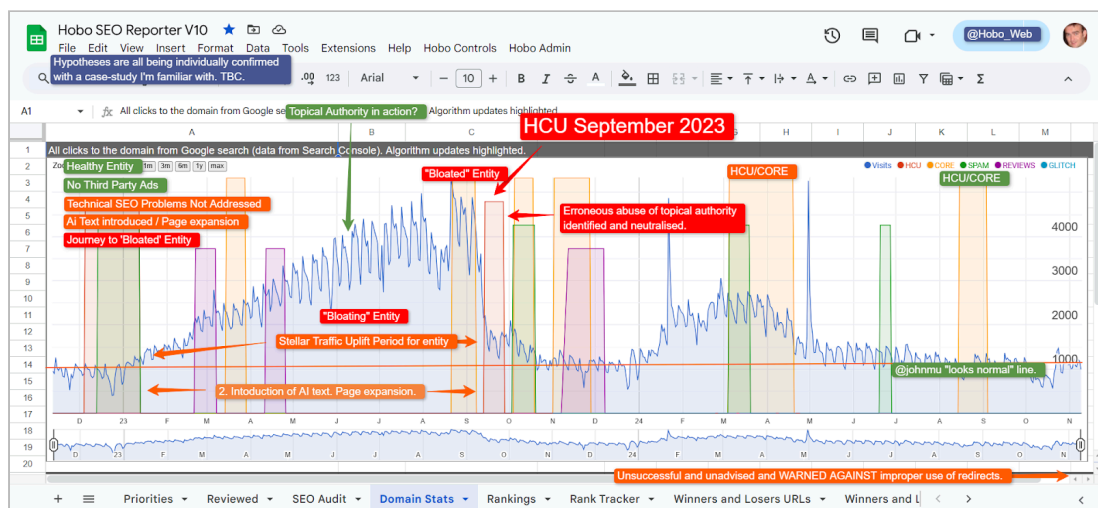


Figure A2 (Forensic Case Study 2 — "A Bloated Entity" / Scaled AI Content Abuse [Anderson, 2025, 5.6]): Search Console trajectory of a "Bloated Entity" where automated AI text generation ($f: W \rightarrow W$) was used to expand page volume to win rankings. Rather than achieving the 30%–40% visibility boost claimed in KDD '24's frozen sandbox [1], arbitrary AI page expansion spread sitewide value thin across the domain and triggered severe HCU / Scaled Content Abuse demotions [5.3, 12].

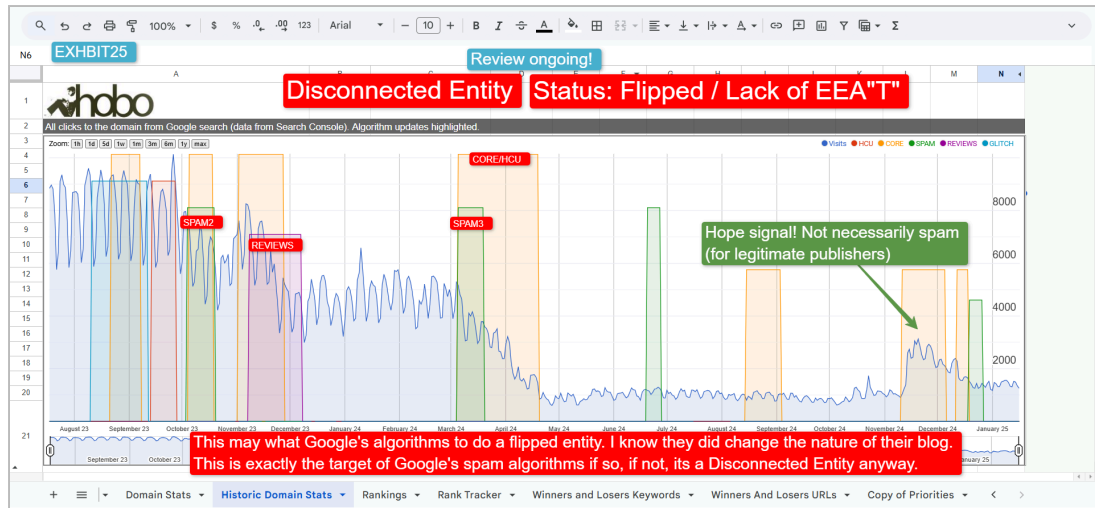


Figure A3 (Forensic Case Study 3 — Exhibit 25: Disconnected / Search-Engine-First Niche Site Demotion [Anderson, 2025, 5.6]): Search Console telemetry from Exhibit 25 demonstrating how Google's corrective algorithms demote domains that lack verifiable entity transparency (Section 2.5.2 of Google's Quality Rater Guidelines) or pursue search-engine-first content strategies—ejecting the domain from the Stage-1 Top-5 retrieval window ($K_2 = 5$).

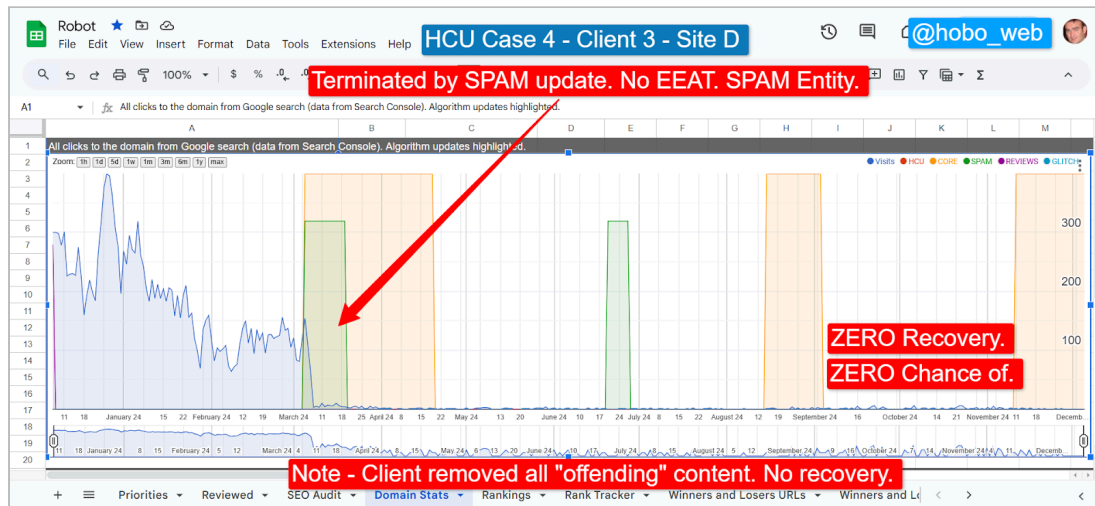


Figure A4 (Forensic Case Study 4 — "A Spam Entity": Terminal Traffic Collapse [Anderson, 2025, 5.6]): Empirical proof that in production search engines, *Entity Health Status* > *Links* > *Relevance*. Once a domain degrades into "Spam Entity" status ($\text{QualityCopiaFireflySiteSignal} / \text{macro-NSR}$ [5.3, 5.4]), organic retrieval turns off completely—resulting in mathematically 0% visibility in downstream RAG synthesis (Equation 2).

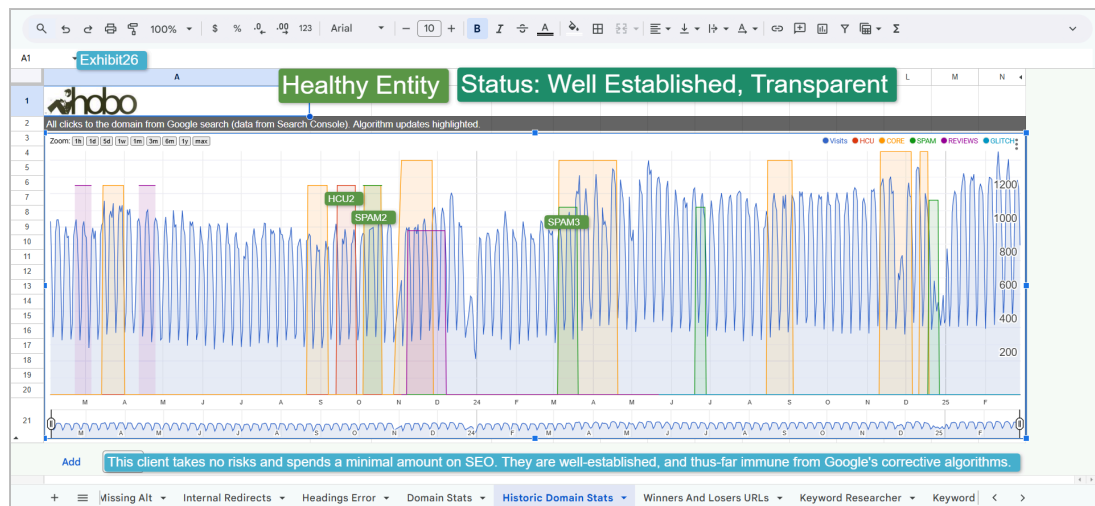


Figure A5 (Forensic Case Study 5 — Exhibit 26: "A Healthy Entity" Control Baseline [Anderson, 2025, 5.6]): The positive control case. When a domain is backed by a transparent, verified **Healthy Entity** where (1) the web vouches for the brand via legitimate links [38, 43] and (2) Google vouches for the site via high sitewide quality (`siteAuthority` / `NSR` [5.4]), classical SEO compounds across algorithm updates—securing permanent placement inside the Stage-1 top-*K* RAG retrieval gate.

References

- Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K., & Deshpande, A. (2024). GEO: Generative Engine Optimization. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*, August 25–29, 2024, Barcelona, Spain, pp. 5–16. ACM. DOI: 10.1145/3637528.3671900 (arXiv:2311.09735).
- Anderson, S. (2025a). *Hobo Strategic SEO 2025 & Hobo SEO Quadrilogy Ebooks*. Hobo Web, July 2025 / May 2026. <https://www.hobo-web.co.uk/strategic-seo-2025/> & <https://www.hobo-web.co.uk/hobo-seo-quadrilogy/>
- Anderson, S. (2025b). *Strategic AI SEO 2025 – Free SEO Ebook*. Hobo Web, August 2025 (Updated October 17, 2026). <https://www.hobo-web.co.uk/hobo-strategic-aiseo-2025/> (Archived Mirror: <https://www.scribd.com/document/910266016/Hobo-Strategic-AISEO-2025>).
- Anderson, S. (2025c). Inference Optimisation: Build Your Own Legitimate Perpetual AI-Powered Content Machine. *Hobo Web*, August 2025 (Updated January 20, 2026). <https://www.hobo-web.co.uk/inference-optimisation-build-your-own-legitimate-perpetual-ai-powered-content-machine/>
- Anderson, S. (2025d–2026). Primary Forensic Audits, Answer Engine Optimisation (AEO) Studies & Google Content Warehouse API Leak Decodings on *Hobo Web* & *Thisisagency.ai*:
 - [5.1] **Primary GEO & AEO Forensic Audits:** What is Generative Engine Optimization (GEO) & A GEO-First Framework Debunked (Oct 23, 2026); The “GEO” Grift is Over. It’s SEO (Oct 23, 2026); What is Answer Engine Optimisation (AEO)? (Oct 23, 2026); Good SEO is Good GEO (Not For New Sites, In Any Reasonable Time Frame) (Apr 24, 2026); & Is There a Sandbox for New Sites in Google AI O reviews? (Apr 28, 2026).
 - [5.2] **Query Fan-Out (G_{qr}) & Information Gain (AE):** What is Query Fan-Out in SEO (AI SEO)? (Oct 3, 2025); The Myth of LSI Keywords – It’s Query Fan Out, All The Way Down (Oct 3, 2025); & What is Information Gain Score in SEO? (Sept 17, 2026).
 - [5.3] **Content Effort (`contentEffort`), Helpful Content & Scaled Content Abuse (`Firefly`):** What is Google’s Content Effort Signal? (Oct 9, 2026); The ContentEffort Attribute, the Helpful Content System and E-E-A-T: Is Gemini Behind the HCU? (Oct 21, 2025); & What is ‘Firefly’? Google’s Scaled Content Abuse System: `QualityCop` `iaFireflySiteSignal` (Oct 13, 2025).
 - [5.4] **Site-Wide Authority (`siteAuthority`), New Site Rank (`NSR`), `Goldmine`, `RankLab` & The Correction Hypothesis:** The Google Algorithm Update Correction Hypothesis (June 24, 2026); The RankLab Loop: How Goldmine and (NSR) Control Your Organic Visibility in Google Core Updates (Oct 11, 2026); Google Site Wide Ranking Factors Decoded: Unmasking `QualityNsrNsrData` (New Site Rank) (Oct 15, 2026); New Site Rank (NSR) – Google’s Site-Wide Quality Engine (Oct 15, 2026); What is Site Authority? The `siteAuthority` Signal in the Google Leak (Oct 13, 2026); & What Is ‘Goldmine’? Google’s Quality Judge For SERPs Leaked (Oct 22, 2025).
 - [5.5] **Core Document Architecture (`PerDocData`, `CompositeDoc`, `PQ`, `QDF` & Topical Authority):** The `PerDocData`: Google’s Leaked Core Document Model Decoded (Oct 12, 2026); The `CompositeDoc`: Google’s Core Document Data Structure Leaked (Oct 11, 2026); The Freshness Signal: `QDF` (Query Deserves Freshness) (Oct 12, 2026); Topical Authority: Site Radius & Site Focus Score from the Google Leak (Oct 10, 2026); The Google Leak, Page Quality Rating (`PQ`) and the Rater Connections (Jan 12, 2026); & E-E-A-T Decoded (Sept 17, 2026).

- [5.6] **On-Page, Technical, Entity SEO & Spam Update Telemetry Post-Leak: SEO Starter Guide: Decoded Post Google Leak** (Oct 15, 2026); **SEO Strategy – The End of Guesswork** (Sept 21, 2026); **The Definitive Guide to On-Page SEO after the Google Leak** (Oct 11, 2026); **Title Tag SEO Best Practices Post Google Leak** (Oct 22, 2025); **Meta Tag SEO Best Practices** (Oct 3, 2025); **HTML Headings for SEO (H1–H6)** (Sept 30, 2025); **Linkbuilding after the Google Leak** (Oct 13, 2026); **Core Web Vitals SEO After The Leaks** (Oct 4, 2025); **Mobile SEO After the Leak** (Oct 13, 2025); **The Disconnected Entity Hypothesis and Google Spam Updates** (Apr 2, 2025); **HCU/SPAM/CORE Impact Analysis Part 1** (Feb 22, 2025); & **December 2024 Google Spam Update Analysis** (Mar 5, 2025).
 - [5.7] **Marketing Cyborgs & Agentic HITL Systems: The Marketing Cyborg Technique** (Apr 3, 2026); **What is Agentic SEO?** (June 30, 2026); & **How To Use Agency (Thisisagency.ai)** (June 30, 2026).
6. Bagga, P. S., Farias, V. F., Korkotashvili, T., Peng, T., & Wu, Y. (2026). E-GEO: A Testbed for Generative Engine Optimization in E-Commerce. *arXiv preprint arXiv:2511.20867v2*. DOI: 10.48550/arXiv.2511.20867
 7. Chen, M., Wang, X., Chen, K., & Koudas, N. (2025). Generative Engine Optimization: How to Dominate AI Search. *arXiv preprint arXiv:2509.08919*. DOI: 10.48550/arXiv.2509.08919
 8. Chen, M., Wang, X., Chen, K., & Koudas, N. (2026a). Navigating the Shift: A Comparative Analysis of Web Search and Generative AI Response Generation. In *EDBT/ICDT 2026 Workshops*, arXiv:2601.16858. DOI: 10.48550/arXiv.2601.16858
 9. Chen, M., & Koudas, N. (2026b). What Drives AI Citations? Content, Retrieval and Model Knowledge in Generative Engine Optimization (GEO). *arXiv preprint arXiv:2610.15033*. DOI: 10.48550/arXiv.2610.15033
 10. Cohen, Z., & Amble, S. (2025). How Generative Engine Optimization (GEO) Rewrites the Rules of Search. *Andreessen Horowitz (a16z)*, May 28, 2025. <https://a16z.com/geo-over-seo/>
 11. Curium AI & Haute Living. (2026). The 28-Signal GEO Protocol, AI Visibility Score Methodology & Enterprise Engineering Specifications. *Curium.ai / Haute Living / Ashby*, 2026. <https://curium.ai/>
 12. Google Search Central. (2026). Google's Guide to Optimizing for Generative AI Features on Google Search. *Google Developers Documentation*, May 15, 2026.
 13. Kim, J., et al. (2026). SAGEO Arena: A Realistic Environment for Evaluating Search-Augmented Generative Engine Optimization. In *Proceedings of KDD '26*, arXiv:2602.12187. DOI: 10.48550/arXiv.2602.12187
 14. Kumar, A., & Lakkaraju, H. (2024). Manipulating Large Language Models to Increase Product Visibility. *arXiv preprint arXiv:2404.07981*. DOI: 10.48550/arXiv.2404.07981
 15. Microsoft Bing Webmaster Team. (2026). Bing Webmaster Guidelines: Generative Engine Optimization and AI Abuse Policies. *Microsoft*, February 27, 2026.
 16. Puerto, H., et al. (2025). C-SEO Bench: Does Conversational SEO Work? In *NeurIPS 2025 Datasets and Benchmarks Track*, arXiv:2506.11097. DOI: 10.48550/arXiv.2506.11097
 17. Samet, U. (2026). A GEO-First Framework: Integrating Search Visibility, Sentiment, and Digital Authority for Organic Growth in the AI Era. *World Journal of Advanced Research and Reviews (WJARR)*, 29(01), 2031–2040. DOI: 10.30574/wjarr.2026.29.1.0152
 18. Wen, Y., et al. (2026). Position: Generative Engine Optimization Creates Underexamined Risks, Governance Must Target Concentration, Disclosure, and Academic Blind Spots. *arXiv preprint arXiv:2606.12439*. DOI: 10.48550/arXiv.2606.12439
 19. Zhou, H., Chen, Z., et al. (2026). IF-GEO: Conflict-Aware Instruction Fusion for Multi-Query Generative Engine Optimization. *arXiv preprint arXiv:2601.13938*. DOI: 10.48550/arXiv.2601.13938
 20. Xiong, C., et al. (2026). AutoGEO: What Generative Search Engines Like. In *ICLR 2026, OpenReview*.
 21. FeatGEO Authors. (2026). FeatGEO: Feature-Driven Generative Engine Optimization on GEO-BENCH. In *Proceedings of ACL 2026*.
 22. MAGEO Authors. (2026). MAGEO: Multi-Agent Generative Engine Optimization on GEO-BENCH. In *Proceedings of ACL 2026*.
 23. SIGIR Competitive GEO Authors. (2026). What Gets Cited: Competitive GEO in AI Answer Engines. In *Proceedings of SIGIR 2026*.
 24. Chen, Y., & Liao, H.; and Du, X., et al. (2025/2026). Caption Injection for Optimization in Generative Search Engines (arXiv:2511.04080); and Multimodal Generative Engine Optimization (arXiv:2601.12263). DOI: 10.48550/arXiv.2511.04080
 25. GEO Defender Authors. (2026). When Optimization Becomes Manipulation: Defending Generative Search against Malicious Generative Engine Optimization (GEO Defender). *ResearchGate Preprint* (ID: 413964683).
 26. Bohnet, B., Tran, V. Q., Verga, P., et al. [Google Research]. (2022). Attributed Question Answering: Evaluation and Modeling for Attributed Large Language Models. *arXiv preprint arXiv:2212.08037*. DOI: 10.48550/arxiv.2212.08037
 27. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. [Google AI Language]. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*. DOI: 10.48550/arxiv.1810.04805

28. Gao, L., Dai, Z., Pasupat, P., et al. [Google Research]. (2022). RARR: Researching and Revising What Language Models Say, Using Language Models. *arXiv preprint arXiv:2210.08726*. DOI: 10.48550/arxiv.2210.08726
29. Jagerman, R., Zhuang, H., Qin, Z., Wang, X., & Bendersky, M. [Google Research]. (2023). Query Expansion by Prompting Large Language Models. *arXiv preprint arXiv:2305.03653*. DOI: 10.48550/arxiv.2305.03653
30. Kwiatkowski, T., Palomaki, J., Redfield, O., et al. [Google Research]. (2019). Natural Questions: A Benchmark for Question Answering Research. *TACL*, 7, 453–466. DOI: 10.1162/tacl_a_00276
31. Ni, J., Qu, C., Lu, J., et al. [Google Research]. (2022). Large Dual Encoders Are Generalizable Retrievers. In *Proceedings of EMNLP 2022*. DOI: 10.18653/v1/2022.emnlp-main.669
32. Vu, T., Iyyer, M., Wang, X., et al. [Google Research / DeepMind]. (2024). FreshLLMs: Refreshing Large Language Models with Search Engine Augmentation. In *Findings of ACL 2024*, 13697–13720. DOI: 10.18653/v1/2024.findings-acl.813
33. Hermanson, T. (2024). GEO Targeted: Critiquing the Generative Engine Optimization Research. *Sandbox SEO*, January 7, 2024. <https://sandboxseo.com/generative-engine-optimization-experiment/>
34. McSweeney, D., & Sturm, E. (2025/2026). The Great GEO (Generative Engine Optimization) Grift (*QueryBurst*, <https://queryburst.com/blog/the-geo-grift/>); and *The GEO Grift: What Actually Works in AI Search*, The Edward Show (Ep. 1,034). http://www.youtube.com/watch?v=q8_4wf_DxM
35. Ray, L. (2026). Are Citations in AI Search Affected by Google Organic Visibility Changes? *Lily Ray NYC Substack*, February 18, 2026. <https://lilyraynyc.substack.com/p/are-citations-in-ai-search-affected>
36. Quaid, D. G. (2026). What is GEO (Generative Engine Optimization)? Defining GEO in an SEO World [Video]. *Primary Position / YouTube*, May 22, 2026. <http://www.youtube.com/watch?v=-ubuDCLJB14>
37. Petrovic, D. (2025). Another Failed Attempt to Kill SEO: An Analysis of Generative Engine Optimization (GEO) and Industry Rebranding. *DEJAN AI*, June 9, 2025. <https://dejan.ai/blog/geo-to-seo/>
38. Brin, S., & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117. DOI: 10.1016/s0169-7552(98)00110-x
39. Robertson, S., & Zaragoza, H. (2009). The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends in Information Retrieval*, 3(4), 333–389. DOI: 10.1561/15000000019
40. Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W.-t. (2020). Dense Passage Retrieval for Open-Domain Question Answering. In *Proceedings of EMNLP 2020*, pp. 6769–6781. DOI: 10.18653/v1/2020.emnlp-main.550
41. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *NeurIPS 2020*, 33, 9459–9474. DOI: 10.48550/arxiv.2005.11401
42. Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). Lost in the Middle: How Language Models Use Long Contexts. *TACL*, 12, 157–173. DOI: 10.1162/tacl_a_00638
43. Gyongyi, Z., Garcia-Molina, H., & Pedersen, J. (2004). Combating Web Spam with TrustRank. In *Proceedings of VLDB 2004*, pp. 576–587. DOI: 10.1016/b978-012088469-8/50052-8
44. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. [Google Brain & Google Research]. (2017). Attention Is All You Need. In *NeurIPS 2017*, 30, 5998–6008. DOI: 10.48550/arXiv.1706.03762 (arXiv:1706.03762).