SEEN: A Four-Layer Framework for Generative Engine Optimization

Dmitry Kargaev
Independent Researcher, Los Angeles, CA
ORCID: 0009-0001-4788-2675
Version date: May 21, 2026

Abstract

Search engine optimization has traditionally treated the ranked web page as the main unit of work and rank position as the main visibility proxy. AI-mediated discovery weakens that model. Generative search systems, answer engines, large language models, and browser agents can retrieve passages, issue derived queries, synthesize answers, attach citations, and recommend entities without presenting a stable ranked list of pages. This working paper introduces SEEN, a four-layer framework for Generative Engine Optimization: Structure, Evidence, Entity, and Notability. SEEN reframes optimization around an entity's evidence corpus rather than a single owned page. It asks whether an entity is easy for AI systems to find, read, verify, resolve, and recommend.

The paper synthesizes platform documentation, academic work on generative engine optimization and conversational SEO, empirical studies of Google AI Overviews, retrieval-augmented generation mechanisms, evidence-evaluation research, knowledge-graph grounding, crawler documentation, and industry studies of AI search citations. It contributes: (1) a stage-aware model of AI visibility; (2) a four-layer entity-corpus framework; (3) a 41-item practitioner checklist that operationalizes the framework; and (4) a conservative validation agenda for future empirical work. The paper does not claim that SEEN guarantees citations, recommendations, rankings, or traffic. Instead, it argues that AI visibility should be analyzed as a corpus-engineering problem in which owned content, platform-hosted first-party content, and third-party corroboration must read as one coherent, verifiable entity.

Keywords: Generative Engine Optimization; GEO; AI search; answer engines; corpus engineering; entity optimization; retrieval-augmented generation; AI citations; search engine optimization; structured data

1. Introduction

The classic SEO question is "where do we rank?" That question remains useful for ordinary ranked search results, but it is incomplete for AI-mediated discovery. A user can ask ChatGPT, Perplexity, Google AI Mode, Gemini, Claude, or a browser agent for a recommendation and receive a synthesized answer rather than a ranked result page. In that environment, an entity may fail without ranking poorly. It may simply be absent.

This absence is an operationally important failure mode. A search-results page exposes a rank position; an AI answer often exposes only the final synthesis, an interface-dependent set of citations, and a set of named entities. The entity that is omitted has no direct observation of whether it was crawled, indexed, retrieved, reranked, used in generation, considered and rejected, confused with another entity, or never entered the candidate pool at all. For publishers and companies, this shifts the problem from position tracking to pipeline diagnosis.

The behavioral motivation is already visible. Pew Research Center tracked 900 U.S. adults across 68,879 Google searches in March 2025 and found that when an AI summary appeared, users clicked traditional search-result links in 8% of visits, compared with 15% when no AI summary appeared. Users clicked links inside the AI summary itself in 1% of visits (Chapekis & Lieb, 2025). In a related causal study, Khosravi and Yoganarasimhan (2026) use the staggered rollout of Google AI Overviews and Wikipedia's multilingual structure to estimate that AI Overview exposure reduced daily traffic to exposed English Wikipedia articles by approximately 15% across matched article-language pairs. These studies do not imply that all AI answer surfaces reduce all traffic; they do show that answer-first interfaces can materially change how attention is distributed.

The technical motivation is equally important. Google states that AI Overviews and AI Mode may use retrieval-augmented generation and query fan-out: multiple related queries can be issued across subtopics and data sources before a response is generated (Google, 2026a; Google, 2026b). OpenAI, Anthropic, and Perplexity publish crawler or bot documentation that distinguishes training crawlers, search/surfacing crawlers, and user-triggered fetches (OpenAI, 2026; Anthropic, 2026; Perplexity, 2026). These distinctions matter because a site may allow one form of access while blocking another, and because being available for training, search surfacing, or user-directed retrieval are different forms of visibility.

Recent academic work has begun to operationalize GEO and related conversational-search optimization problems. Aggarwal et al. (2024) introduce Generative Engine Optimization (GEO) and report visibility gains from content modifications such as citations, statistics, and quotations on GEO-bench. Chen, Wang, Chen, and Koudas (2025) test live AI search engines and report that AI search differs from traditional Google search in its source preferences, with a strong emphasis on earned media and substantial cross-engine variance. Puerto et al. (2025) introduce C-SEO Bench and caution that many conversational SEO methods are ineffective or even harmful when evaluated across domains and competitive adoption settings. Wu et al. (2025) introduce AutoGEO, showing that generative-engine preferences can be learned and used to rewrite content, but also reinforcing that engine preferences are model- and setting-dependent.

This paper argues that the most useful practitioner unit is neither the page nor the isolated prompt result. It is the entity's evidence corpus: owned content, platform-hosted first-party content, and third-party corroboration, connected into a coherent and verifiable identity. SEEN is a framework for auditing and improving that corpus.

SEEN stands for Structure, Evidence, Entity, and Notability. It is defined as follows:

> SEEN makes an entity easy for search engines, answer engines, LLMs, and browser agents to find, read, verify, resolve, and recommend.

The framework makes three claims. First, AI visibility should be analyzed across a retrieval-to-recommendation pipeline rather than reduced to rank or citation count. Second, an entity's evidence corpus spans owned, platform-hosted, and third-party surfaces. Third, the four SEEN layers interact multiplicatively: a weak layer can cap the value of the others. Clear content that lacks evidence is easy to quote but weak to trust; evidence without structure can be missed; both can be misattributed if the entity is ambiguous; and all three may still fail to produce recommendations if the wider web offers no corroboration.

The paper makes four contributions:

1. It proposes a stage-aware model of AI visibility that separates access, index eligibility, fan-out, retrieval, generation utility, citation, recommendation, and referral outcomes.
2. It defines SEEN as a four-layer framework for entity-level corpus engineering.
3. It operationalizes SEEN through a 41-item practitioner checklist.
4. It specifies claim boundaries and a validation agenda so the framework can be tested rather than treated as doctrine.

2. Scope and Definitions

This paper uses **SEO** to mean Search Engine Optimization: optimizing pages and sites to appear in classic ranked search results. SEO remains relevant. Google states that generative AI features in Search are rooted in its core Search ranking and quality systems, and that pages must be indexed and snippet-eligible to appear in Google generative AI features (Google, 2026a; Google, 2026b). SEEN therefore assumes that crawlability, indexability, page quality, structured data, internal links, and external authority remain part of the substrate.

It uses **GEO** to mean Generative Engine Optimization: improving visibility, citation, mention, and recommendation outcomes in generative systems. The term is used in the academic sense introduced by Aggarwal et al. (2024), but this paper avoids interpreting GEO as a deterministic tactic set. The more conservative interpretation is that GEO studies response-level visibility, not merely rank.

It uses **AI-mediated discovery** as the broader category covering search-augmented chat, AI Overview-style summaries, AI Mode-style interfaces, answer engines, conversational search engines, and browser agents acting on a user's behalf.

It uses **entity** to mean the object being represented: a person, product, organization, service, project, publication, or place.

It uses **evidence corpus** to mean every accessible surface that helps a retrieval or answer system represent the entity. The term does not imply a single unified platform-visible corpus. It is the analyst-defined set of public or semi-public surfaces available for possible retrieval, citation, or corroboration. The corpus has three categories:

1. **Owned content:** the entity's site, blog, documentation, help center, product pages, methodology pages, case studies, and other surfaces controlled by the entity.
2. **Platform-hosted first-party content:** content authored by the entity on platforms it does not own, including GitHub, LinkedIn, Substack, Medium, YouTube, podcast feeds, SSRN, arXiv, Google Scholar, Wikidata, and similar surfaces.
3. **Third-party corroboration:** independent coverage, reviews, customer-side case studies, community discussions, directory entries with editorial standards, interviews, podcasts, partner pages, and other external mentions.

It uses **recommendation** in a narrow sense: the AI answer names the entity as a fit for the user's expressed intent. A mention is weaker than a recommendation; a citation is weaker than a recommendation; a correct recommendation is stronger than all three.

The framework is not a ranking-factor model. It does not estimate production weights for Google, OpenAI, Perplexity, Anthropic, Microsoft, or any other system. It is a diagnostic and operational model: it identifies conditions under which an AI system has a better substrate for finding, reading, verifying, resolving, and recommending an entity.

3. Method and Evidence Base

This is a conceptual working paper based on structured evidence synthesis, not a primary causal experiment. The evidence base was organized into five source classes.

Platform documentation includes official Google, OpenAI, Anthropic, and Perplexity documentation. These sources are authoritative for stated eligibility, crawler behavior, and vendor-recommended controls. They are not independent evidence of hidden ranking or citation weights.

Academic mechanism evidence includes research on GEO, conversational SEO, generative search, retrieval-augmented generation, evidence evaluation, citation fidelity, and knowledge-graph grounding. These sources support model structure and mechanism plausibility.

Empirical AI-search studies include measurement work on Google AI Overviews, AI Mode, Gemini, Wikipedia traffic, Reddit engagement, and live AI search engines. These sources support the claim that AI-mediated search changes source selection, click-through, claim support, and publisher outcomes.

Industry observational studies include vendor or commercial analyses of AI search citations, AI mentions, platform citation share, and authority signals. These studies are useful for current practice but require caveats because methods, samples, incentives, and metrics vary.

Practitioner operationalization includes the author's SEEN checklist and platform inventory work. These are not independent evidence. They are the proposed operational artifact to be tested.

3.1 Evidence synthesis protocol

Sources were selected between 2026-05-20 and 2026-05-21 from local source notes in `research/sources/`, Google Search Central documentation, OpenAI/Anthropic/Perplexity crawler documentation, arXiv, ACL Anthology, ACM/venue pages where available, SSRN, Pew Research Center, and industry research pages from Semrush, Otterly.AI, and related AI-visibility vendors. Search terms included combinations of "generative engine optimization," "AI search citation," "Google AI Overviews source quality," "AI Overview publisher traffic," "conversational SEO benchmark," "retrieval augmented generation," "dense passage retrieval," "knowledge graph hallucination," "AI crawler robots.txt," and the names of known papers and vendors.

Sources were included when they met at least one of four criteria: (1) official documentation for a platform behavior or eligibility claim; (2) academic or working-paper evidence relevant to retrieval, citation, entity grounding, or AI-search measurement; (3) credible behavioral or industry evidence with inspectable methodology; or (4) direct operational relevance to the SEEN checklist. Sources were excluded or downgraded when they were pure opinion, unsupported tactical advice, single-prompt screenshots, derivative summaries without a primary source, or vendor material whose method was too opaque for the claim being made.

Each retained source was extracted into a source note when used for a non-obvious claim. Extraction fields included source URL, source type, evidence bucket, verification date, core findings, SEEN use, and claim boundaries. Industry studies were treated as directional context unless their methods were explicit enough to support a numeric claim. Living vendor documentation was checked against the live page on 2026-05-21 and should be re-checked before submission.

Sources were used under the following claim rule:

```
Official documentation → vendor-stated behavior and eligibility only. Academic mechanism
work → plausible mechanism and benchmark evidence. Empirical AI-search studies →
observed production-system behavior within the study design. Industry studies →
directional context unless methods are independently auditable. Practitioner judgment →
explicitly marked as framework judgment, not empirical proof.
```

This rule matters because GEO discourse often merges incompatible claims. A platform document that says a crawler exists does not prove citation lift. A benchmark result does not prove a live-engine tactic. A single prompt screenshot does not prove durable visibility. A checklist item does not become causal evidence because it is plausible.

The paper also uses an anti-overclaim rule: if a source supports a mechanism but not a live-engine outcome, the paper states the mechanism only. For example, knowledge-graph-grounded retrieval can reduce hallucination in controlled benchmarks (Lavrinovics et al., 2025), but that does not prove that adding a Wikidata entry will change a consumer AI answer. Similarly, `llms.txt` is a proposed convention for language-model-facing site summaries, but Google states that site owners do not need new AI text files or Markdown to appear in its generative Search features (Google, 2026a; llms.txt, 2024).

4. A Stage-Aware Model of AI Visibility

AI visibility should be analyzed as a sequence of related but non-identical events. The exact pipeline differs across systems, but separating the events prevents rank, retrieval, citation, and recommendation from being collapsed into one proxy.

4.1 Access and crawler policy

The first question is whether the system can access the relevant content. This includes ordinary search crawlers, AI search crawlers, training crawlers, user-triggered fetchers, bot IP verification, robots rules, CDN/WAF behavior, JavaScript rendering, and paywall or login constraints.

Crawler classes differ. OpenAI distinguishes OAI-SearchBot for ChatGPT search surfacing, GPTBot for training, OAI-AdsBot for ad landing-page validation, and ChatGPT-User for user-triggered actions; OpenAI states that these settings are independent and that OAI-SearchBot should be managed for ChatGPT search visibility (OpenAI, 2026). Anthropic distinguishes ClaudeBot, Claude-User, and Claude-SearchBot for training, user-initiated retrieval, and search-quality functions (Anthropic, 2026). Perplexity states that PerplexityBot will not index full or partial text content disallowed by robots.txt, while it may still index domain, headline, and a brief factual summary for blocked pages (Perplexity, 2026).

SEEN treats access as a Structure concern, but crawler policy also affects Notability because third-party and platform-hosted sources may be reachable even when owned pages are not.

4.2 Index and eligibility

Access is not the same as eligibility. Google states that pages must meet Search technical requirements, be indexed, and be eligible to be shown with a snippet to appear in its generative AI features; it also states that eligibility does not guarantee crawling, indexing, serving, or citation (Google, 2026a). This creates a basic distinction between "technically possible" and "actually surfaced."

For SEEN, this means technical SEO is not optional. It is the substrate. The difference is that the downstream target is no longer only a ranked listing; it is also possible inclusion in synthesized answers, supporting links, and recommendation surfaces.

4.3 Query fan-out and intent expansion

Google states that AI Overviews and AI Mode may use query fan-out, issuing multiple related searches across subtopics and data sources to develop a response (Google, 2026b). This changes the visible-query assumption. A page may fail the final answer not because it does not rank for the visible query, but because it lacks answer-bearing passages for the derived subquestions.

This supports corpus-level thinking. The relevant unit is not one keyword mapped to one landing page. It is an intent cluster: definitions, comparisons, constraints, evidence needs, named entities, alternatives, procedures, and local or temporal facts that may be gathered before the answer is generated.

4.4 Retrieval and reranking

Once candidate documents or passages are available, systems must retrieve and rerank them. RAG and passage-retrieval literature show why passage-level answer utility matters: Lewis et al. (2020) combine generation with retrieved non-parametric memory, and Karpukhin et al. (2020) frame open-domain question answering around passage retrieval. A document can be topically relevant but fail to contain a self-contained, directly relevant passage. Conversely, a lower-ranked page or platform-hosted page may contain a passage that better answers a derived subquery.

Production evidence points in the same direction. Xu, Iqbal, and Montgomery (2026) find in a large Google AI Overview measurement study that nearly 30% of AIO-cited domains did not appear in the co-displayed first-page organic results. Grossman et al. (2026) find low average Jaccard similarity among sources returned by Google Search, Google AI Overviews, and Gemini Flash 2.5, suggesting that source pools differ across generative and traditional search surfaces.

4.5 Generation utility and claim support

Retrieved material must then support answer generation. A passage is more useful when it is directly relevant, entity-rich, fact-bearing, scoped, and supported by evidence. Wan, Wallace, and Klein (2024) show that language models can overweight direct relevance relative to stylistic credibility cues. This suggests that Evidence work should focus on direct claim-source pairing rather than only "trustworthy-looking" prose.

Generation utility is not the same as citation fidelity. Xu et al. (2026) decompose Google AI Overview responses into atomic claims and report that 11.0% are unsupported by cited pages, with source quality and claim fidelity largely independent. This is central for SEEN: the goal is not only to be cited, but to be represented correctly by sources that actually support the claim.

4.6 Citation, mention, and recommendation

Citation, mention, and recommendation are distinct outcomes. A cited page may not name the target entity. A mentioned entity may not be recommended. A recommendation may not cite the entity's owned page. A correct recommendation requires entity resolution, claim support, and fit to intent.

Chen et al. (2025) report that live AI search systems differ substantially in source preferences, domain diversity, freshness, cross-language stability, and phrasing sensitivity. This supports engine-specific evaluation. A SEEN audit should therefore avoid treating a single citation or one successful answer as durable evidence.

4.7 Referral and downstream value

Even when an entity is cited or recommended, the downstream value may differ from classic SEO traffic. Pew reports lower click-through in Google search visits with AI summaries (Chapekis & Lieb, 2025). Khosravi and Yoganarasimhan (2026) estimate a traffic reduction for Wikipedia under AI Overview exposure. Zhang et al. (2026), however, find that AI Overviews increased engagement in safe-for-work Reddit communities for experience-based discussions, while AI Mode later reduced those gains. The effect depends on content type and interface design.

SEEN therefore treats referral as downstream of visibility, not identical to it. The framework's primary object is not traffic prediction. It is entity representation in AI-mediated discovery.

5. Related Work

5.1 Generative Engine Optimization and Conversational SEO

Aggarwal et al. (2024) formalize GEO as black-box optimization for generative-engine responses and introduce GEO-bench, a benchmark of queries and sources for studying response-level visibility. Their paper reports that certain content modifications can improve visibility by up to 40% in their experimental setup. The result is foundational but should be bounded: the test uses the authors' benchmark and synthesized-answer pipeline, and the intervention set includes content additions whose effects may partly reflect additional unique content rather than a generalizable "GEO tactic."

Chen et al. (2025) move closer to production systems by studying live AI search engines. Their findings are important for SEEN because they foreground earned media, engine-specific variation, and sensitivity to phrasing and language. Those observations support a framework that includes Notability and avoids universal tactic claims.

Puerto et al. (2025) provide a useful counterweight. C-SEO Bench evaluates conversational SEO methods across tasks, domains, and competitive adoption rates. The paper finds that many current methods are ineffective or negatively affect document ranking, while traditional SEO strategies that improve ranking in the LLM context perform better. It also finds diminishing gains as more actors adopt the same techniques. This supports SEEN's anti-hype stance: formatting and prompt-era content tricks should not be treated as guaranteed advantages.

Wu et al. (2025) introduce AutoGEO, a framework for learning generative-engine preferences and applying them through rewriting. AutoGEO is relevant because it frames GEO as a preference-learning and optimization problem. It also implies that preferences may be domain- and engine-specific, which supports SEEN's use as an audit framework rather than a closed list of universal levers.

5.2 AI Overview measurement and publisher effects

Recent AI Overview studies show that generative search changes source selection and claim behavior. Xu et al. (2026) study 55,393 trending queries over a 40-day window and report AIO activation of 13.7% overall and 64.7% for question-form queries. They also report that AIO-cited domains are more credible than co-displayed first-page results, nearly 30% of cited domains do not appear in those first-page results, and 11.0% of decomposed atomic claims are unsupported by cited pages.

Grossman et al. (2026) introduce a public benchmark of 11,500 user queries and compare Google Search, Google AI Overviews, and Gemini Flash 2.5. They report AIO generation for 51.5% of representative real-user queries, low source overlap across systems, and sensitivity to query edits. The paper has direct implications for SEEN's durability concept: a correct answer from one run or one surface is not enough.

Traffic and ecosystem effects are not uniform. Khosravi and Yoganarasimhan (2026) estimate a reduction in English Wikipedia article traffic after AI Overview exposure. Zhang et al. (2026) find that AI Overviews increased Reddit engagement for experience-based discussions relative to NSFW communities not surfaced in AIOs, but that AI Mode largely eliminated those gains. These mixed results reinforce the need to separate answer inclusion, citation, referral traffic, and community engagement.

5.3 Retrieval, evidence evaluation, and attribution

RAG and query-transformation work motivate the Structure and Evidence layers. Generated answers can depend on retrieved passages, derived subqueries, and reranked candidate pools (Lewis et al., 2020; Karpukhin et al., 2020). This means a page-level rank can be insufficient as a visibility proxy.

Evidence evaluation work complicates simple credibility advice. Wan et al. (2024) show that models rely heavily on direct relevance and may underweight stylistic features humans associate with credibility. For SEEN, the implication is that source-backed evidence must also be query-relevant and easy to connect to the claim. A methodology page that never states the claim in retrievable language may help human trust but fail machine retrieval.

Attribution work also matters because citations can be unfaithful. A cited page may not support the attached claim, or an answer may use a source without citing it. This is why SEEN distinguishes Evidence from citation count: the target is correct support, not only citation appearance.

5.4 Entity grounding and knowledge graphs

Entity resolution is a core problem for AI-mediated discovery. A product, founder, company, publication, and parent organization may appear across many surfaces with inconsistent names and descriptions. If those surfaces do not resolve to one entity, AI systems may merge, split, or misattribute information.

Lavrinovics et al. (2025) provide controlled evidence that knowledge-graph-grounded retrieval improves hallucination-related outcomes in multilingual QA. This does not prove a direct production effect from any individual schema or Wikidata action, but it supports the mechanism that structured entity grounding can improve factuality and disambiguation.

Schema.org `sameAs`, Organization/Person/Product markup, Wikidata, Google Business Profile, ORCID, GitHub, LinkedIn, SSRN, Google Scholar, and other identity surfaces can therefore be interpreted as entity-resolution infrastructure. Their value depends on accuracy and consistency, not raw count.

5.5 Platform-hosted and third-party surfaces

Industry studies suggest that AI citations and mentions draw heavily from high-authority domains, community platforms, and platform-hosted content. Semrush reports strong correlations between Authority Score and AI mentions, threshold effects, and similar AI-visibility correlations for follow and nofollow links (Loktionova & Drozdov, 2025). A separate Semrush study reports LinkedIn as the second most-cited domain after Reddit across its ChatGPT Search, Google AI Mode, and Perplexity sample, with original and active content overrepresented among cited LinkedIn URLs (Loktionova, 2026). Otterly.AI reports a high share of citations going to community and platform surfaces and flags technical access barriers on many analyzed sites (Peham, 2026).

These studies are not causal proof. Their value is directional: they support the premise that an owned site is not the whole discoverable corpus. SEEN therefore includes platform-hosted first-party content and third-party corroboration as first-class parts of the framework.

6. The SEEN Framework

SEEN has four layers:

```
S — Structure Can AI extract the right answer from your content? E — Evidence Can AI
verify your claims? E — Entity Does AI know exactly who or what you are? N — Notability
Does the wider web corroborate that you matter?
```

The layers are easiest to understand as failure filters. At each stage, a system can fail to use or recommend the entity for a different reason. Structure failure means the answer is inaccessible or not extractable. Evidence failure means the claim is unsupported or hard to verify. Entity failure means the system cannot resolve the entity cleanly. Notability failure means the entity is legible but lacks external corroboration.

6.1 Structure

Diagnostic question: Can AI extract the right answer from your content?

Structure maps primarily to access, index eligibility, fan-out coverage, retrieval, and generation utility. It concerns whether important facts are machine-readable, passage-addressable, and visible in the rendered or indexed content.

Key mechanisms:

- **Crawl and renderability:** content must be reachable through the relevant crawler or user-triggered fetch path.
- **Snippet and index eligibility:** for Google generative AI features, pages must be indexed and snippet-eligible (Google, 2026a).
- **Passage extraction:** RAG-like systems retrieve and condition generation on passages rather than human-scale page reading.
- **Query fan-out coverage:** a page should answer adjacent subquestions and constraints, not only the head query.
- **DOM and accessibility readiness:** browser agents interact with rendered pages and accessibility structures, not just source HTML.

Operational indicators:

- Answer-first summaries near the top of key pages.
- Semantic H1/H2/H3 hierarchy.
- Visible HTML facts for pricing, scope, location, jurisdiction, product category, service model, and contact paths.
- Comparison tables with named columns.
- FAQs that reflect real buyer or user questions.
- Schema markup that matches visible content.
- Useful no-JavaScript or server-rendered fallback for core content.
- Long-form pages with anchor links and self-contained sections.

The evidence for exact formatting choices is incomplete. ALM Corp and Indig (2026) report that 44.2% of ChatGPT citations in their sample came from the first 30% of webpage content, but the study is industry research and correlational. The safer conclusion is not "moving text upward guarantees citations." It is that answer location and extractability are plausible, measurable Structure variables.

6.2 Evidence

Diagnostic question: Can AI verify your claims?

Evidence maps to generation utility, claim support, citation fidelity, and displacement resistance. It concerns whether the corpus contains reachable support for the claims an AI system might repeat.

Key mechanisms:

- **Claim-source pairing:** important claims should connect to method pages, case studies, primary sources, or independent reviews.

- **Direct relevance:** the supporting passage should address the query or claim directly, not merely look credible.
- **Citation fidelity:** citations should support the claims attached to them.
- **Cross-surface verification:** third-party and platform-hosted support can verify claims that would otherwise exist only on owned pages.

Operational indicators:

- Quantitative claims linked to methods, data, case studies, or primary sources.
- Original research or methodology pages under an owned URL.
- Case studies with named clients, real numbers, and dates.
- Independent reviews on credible platforms.
- Credentialed bylines on substantive content.
- Transparent pricing model, even if exact quotes vary.
- Attributed testimonials with role, company, and link where possible.
- Linked press, podcast, conference, and partner proof.

The GEO literature supports Evidence cautiously. Aggarwal et al. (2024) report visibility gains from citations, statistics, and quotations in GEO-bench. Chen et al. (2025) report that justification-oriented and earned-media strategies matter in live AI search, but with engine-specific differences. Wan et al. (2024) show that relevance can dominate stylistic credibility. Xu et al. (2026) show that citations can fail to support claims. Together, these sources imply that Evidence is not decorative citation stuffing; it is the engineering of directly relevant, supportable claims.

6.3 Entity

Diagnostic question: Does AI know exactly who or what you are?

Entity maps to identity resolution, disambiguation, attribution, and hallucination reduction. It concerns whether the system can connect all relevant surfaces to one entity and avoid confusing that entity with a namesake, parent brand, founder, product, or category term.

Key mechanisms:

- **Canonical naming:** one stable name reduces split-entity behavior.
- **Description consistency:** repeated one-sentence descriptions reduce ambiguity across surfaces.
- **Structured identity:** schema, `sameAs`, Wikidata, ORCID, Google Business Profile, and platform profiles provide machine-readable links.
- **Relationship clarity:** founder-company-product-publication relationships should be explicit.
- **Disambiguation:** namesake collisions should be handled in titles, descriptions, schema, and About copy.

Operational indicators:

- Consistent canonical name across owned and platform-hosted surfaces.
- One reusable one-sentence description.
- Organization/Person/Product JSON-LD where an owned site exists.
- `sameAs` links to canonical, active profiles rather than long-tail junk directories.
- Real About page with entity, founder, team, story, and mission.
- Explicit founder-company-product relationships.
- Wikidata or other structured identity records where appropriate.
- Consistent profile photos, logos, bios, and dates.

The evidence is partly mechanistic. Lavrinovics et al. (2025) show that knowledge-graph-grounded retrieval can improve hallucination-related outcomes in a controlled benchmark. Schema.org provides the `sameAs` vocabulary for linking identity surfaces. Production systems remain opaque, so the correct claim is not that one schema field guarantees better AI output. It is that entity coherence reduces avoidable ambiguity in the corpus AI systems retrieve from.

6.4 Notability

Diagnostic question: Does the wider web corroborate that you matter?

Notability maps to external corroboration, authority, recommendation likelihood, and durability. It concerns whether the entity is supported by signals outside its own self-description.

Key mechanisms:

- **Earned media:** independent sources can corroborate claims and category membership.
- **Authority thresholds:** high-authority domains may be more likely to enter retrieval and citation pools.
- **Community presence:** forums and communities can supply experience-based evidence and natural-language comparisons.
- **Platform-hosted reach:** visible engagement on authoritative platforms can function as a notability signal, though this remains partly practitioner judgment.
- **Temporal durability:** SEEN treats third-party proof as a longer-lived corpus asset than short-term formatting changes, though this remains a framework hypothesis rather than a measured universal effect.

Operational indicators:

- Independent category publications mentioning the entity.
- Editorial "best of," "alternatives," or category-list inclusion where legitimate.
- Organic mentions in relevant subreddits, forums, Discords, or professional communities.
- Links or mentions from genuinely high-authority domains.
- Wikipedia mentions where notability standards allow.

- Speaking, podcast, or guest-writing trail.
- Platform-hosted reach: GitHub stars, Substack subscribers, LinkedIn engagement, YouTube subscribers, podcast listeners, SSRN downloads, citations.
- Partner-side and customer-side proof.

Notability is both crucial and difficult to validate. Chen et al. (2025) report earned-media importance in live AI search. Semrush reports authority correlations with AI mentions (Loktionova & Drozdov, 2025), and its LinkedIn study reports heavy use of LinkedIn as a platform-hosted citation surface (Loktionova, 2026). Otterly.AI reports heavy use of platform and community surfaces in AI citations (Peham, 2026). These sources support direction, not deterministic prediction. SEEN treats Notability as a layer that likely compounds over time, but that compounding claim should be tested rather than assumed.

7. Operationalizing SEEN as a Checklist

The practitioner artifact for SEEN is a 41-item checklist. Structure, Evidence, and Entity each contain 10 items. Notability contains 11 because speaking/guest-writing trail and platform-hosted reach are separated.

Each item is scored as `1` if clearly true today and `0` if missing, vague, stale, or only partially true. Site-specific items can be skipped without penalty for purely platform-hosted entities, with the layer score renormalized over applicable items.

The checklist is intentionally diagnostic rather than predictive. It does not estimate the probability of appearing in a given AI answer. It identifies weak corpus layers. Its key rule is to fix the weakest layer first.

7.1 Why weakest-layer prioritization matters

The framework treats layers as multiplicative:

```
SEEN readiness ≈ f(Structure × Evidence × Entity × Notability)
```

This is not a formal fitted equation. It is a sequencing heuristic. The point is that a near-zero layer can suppress the value of the others. A strong Evidence corpus cannot help if it is not extractable. Clean Structure cannot help if claims are unsupported. Both can fail if the entity is unresolved. All three can fail to produce recommendations if no independent surface suggests the entity matters.

7.2 Candidate observables by layer

Future studies can measure SEEN layers using observable variables:

Layer	Candidate observables
Structure	renderability, index status, snippet eligibility, answer passage position, heading quality, schema-content consistency, JS dependence, anchorability

Evidence	claim-source link rate, method-page availability, primary-source citation rate, named case-study count, independent-review presence, citation support accuracy
Entity	canonical-name consistency, profile-description consistency, `sameAs` graph completeness, Wikidata/identifier presence, entity-collision rate, hallucinated-attribute rate
Notability	independent mentions, authoritative referring domains, community mentions, editorial list inclusion, platform-hosted engagement, partner/customer-side proof

This table is part of the framework's research value. It translates a practitioner checklist into variables that can be tested against AI answer outcomes.

7.3 Outcome variables

A validation study should separate at least six outcomes:

- **Mention inclusion:** the entity appears in the answer.
- **Correct mention:** the entity is named and described accurately.
- **Citation inclusion:** a relevant source is linked or cited.
- **Citation support:** the cited source supports the attached claim.
- **Recommendation inclusion:** the entity is recommended for the user's intent.
- **Recommendation fit:** the recommendation matches what the entity actually does.

These outcomes should not be collapsed into one "AI visibility" score. A hallucinated mention is not a win. An unsupported citation is not Evidence success. A recommendation for the wrong job can be harmful.

8. Proposed Validation Design

The next step for SEEN is empirical validation. A modest study would be enough to test whether the framework identifies real failure modes.

8.1 Entity sample

Select 50 to 100 entities across two or three categories where recommendation queries are common. Categories should include at least one B2B software or services category and one consumer or local-service category. Each entity should have an owned site, at least two platform-hosted profiles, and some third-party mention history, even if weak.

8.2 SEEN scoring

Two independent coders score each entity using the 41-item checklist. Inter-rater reliability should be reported at the item level using Cohen's kappa for binary items and intraclass correlation for layer totals. Disagreements are resolved only after reliability is calculated. Layer scores and item-level scores are retained separately.

8.3 Prompt panel

For each category, create prompt classes:

- Category recommendation prompts.
- Alternatives-to prompts.
- Comparison prompts.
- Verification prompts.
- Pricing or scope prompts.
- Founder/company/product identity prompts.

Each prompt should have paraphrase variants to test phrasing sensitivity. Runs should be repeated across at least three systems, for example Google AI Mode or AI Overviews where available, ChatGPT Search, Perplexity, Gemini, and Claude web search. Each prompt-system pair should be captured at least three times on separate days where the interface permits repeated collection. Account state, geography, language, device/browser, model or interface version where visible, date, and time should be recorded.

8.4 Coding outputs

For each answer, code:

- whether the entity appears;
- whether the description is correct;
- whether the entity is recommended;
- whether recommendation fit is correct;
- which source URLs are cited;
- whether citations support the claims;
- whether cited URLs are owned, platform-hosted first-party, or third-party;
- whether hallucinations or wrong-entity merges occur.

8.5 Analysis

The study should test whether weak layers predict specific failure modes:

- Low Structure → missing citations to owned pages, wrong pricing/scope, weak extractability.
- Low Evidence → unsupported claims, failure on verification prompts, citation to third-party sources over owned sources.

- Low Entity → wrong descriptions, stale roles, namesake confusion, founder/product/company merges.
- Low Notability → absence from recommendation prompts despite adequate owned content.

The analysis need not claim causality. A first validation can test whether SEEN scores correlate with observed failure types and whether the weakest-layer heuristic produces useful remediation priorities. A simple first model would use layer scores to predict binary answer outcomes with logistic regression or mixed-effects logistic regression, with random effects for entity, prompt, and system when the sample supports it. Item-level analyses should be exploratory unless preregistered.

9. Discussion

SEEN's core move is to treat AI visibility as corpus engineering. The corpus is not only the owned site. It includes platform-hosted first-party content and third-party corroboration. This matters because current evidence suggests that AI answer systems draw from sources whose relationship to the visible query and ordinary rank is indirect. Query fan-out, search indexes, passage retrieval, entity grounding, citation attribution, and interface design all shape the final answer.

The framework is deliberately not a list of isolated AI-search tactics. Google rejects the idea that its generative Search features require special AI-only files or markup (Google, 2026a). C-SEO Bench cautions that many conversational SEO methods do not generalize and can become congested under competitive adoption (Puerto et al., 2025). AI Overview studies show unsupported claims and low source overlap (Xu et al., 2026; Grossman et al., 2026). Those findings argue for foundational work: content that can be retrieved, claims that can be supported, entities that can be resolved, and notability that can be corroborated.

The framework also distinguishes AI visibility from AI traffic. The business value of AI-mediated discovery may come through citations, mentions, brand familiarity, no-click answers, referral traffic, or recommendation decisions. Depending on interface design, AI answers may suppress clicks, redirect attention, or increase engagement on certain source platforms (Chapekis & Lieb, 2025; Khosravi & Yoganarasimhan, 2026; Zhang et al., 2026). The operational target should therefore be explicit: traffic, citation, correct mention, recommendation, or trust transfer are different outcomes.

Finally, SEEN is meant to be used over time. Structure and Evidence can often be improved directly. Entity coherence requires cross-surface maintenance. Notability compounds slowly and is hardest to fake. A quarterly re-run is more useful than a one-time screenshot audit.

10. Limitations

This paper has several limitations.

First, it is a conceptual and methodological working paper, not a primary causal experiment. It synthesizes existing evidence and proposes an operational checklist. It does not prove that applying SEEN causes lift in AI mentions, citations, recommendations, or traffic.

Second, the evidence base is mixed. Official documentation is authoritative for vendor-stated behavior, but not for hidden weights. Academic benchmarks support mechanisms, but may not represent production systems. Industry studies are current and useful but often vendor-mediated.

Third, AI systems change quickly. Crawler behavior, citation display, retrieval strategies, model versions, interface design, and publisher policies can change within months. Any measured association should be treated as time-bounded.

Fourth, the checklist is not validated. Its scoring thresholds are practical heuristics, not empirically estimated cutoffs. The proposed validation design should be run before treating layer scores as predictive.

Fifth, Notability is the least controlled layer. External coverage, community discussion, platform engagement, and editorial inclusion are partly earned and partly category-dependent. Some entities legitimately cannot or should not expose all evidence publicly, especially in regulated, confidential, local, or early-stage contexts.

Sixth, SEEN may be less applicable to pure media publishers, anonymous creators, private companies with no public surface, or entities whose discovery is mostly app-store, marketplace, or social-feed mediated. The framework can still diagnose corpus gaps, but its checklist may need category-specific weighting.

11. Conclusion

AI-mediated discovery changes the visibility problem from rank alone to representation inside answers and recommendations. The page still matters, but it is no longer the whole unit. The entity's evidence corpus matters: owned content, platform-hosted first-party content, and third-party corroboration must add up to a coherent, verifiable whole.

SEEN is a framework for that work. Structure makes content extractable. Evidence makes claims verifiable. Entity makes identity resolvable. Notability makes relevance externally corroborated. The framework does not promise deterministic AI visibility. It gives practitioners and researchers a disciplined way to audit the substrate on which AI visibility depends.

The central hypothesis is simple: entities that engineer their corpus deliberately should be easier for AI systems to find, read, verify, resolve, and recommend than entities that treat AI visibility as a single-surface formatting problem. That is not a guarantee. It is a testable claim.

Author Contributions

The author conceived the framework, conducted the source review, designed the evidence synthesis protocol, developed the checklist, and wrote the manuscript.

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflict of interest.

Data Availability

The analysis in this manuscript is based on project materials maintained by the author, including source notes, reference files, framework notes, and checklist materials used during manuscript preparation. These materials are available from the author upon reasonable request.

AI Disclosure Statement

This research was conducted with the assistance of AI tools. Large language models were used for source discovery assistance, note organization, and draft support. All analysis, interpretation, methodology design, and conclusions are the work of the author. AI-assisted text and source suggestions were reviewed, verified, and substantially revised before inclusion.

References

Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K., & Deshpande, A. (2024). *GEO: Generative Engine Optimization*. Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining. https://arxiv.org/abs/2311.09735

ALM Corp, & Indig, K. (2026). *ChatGPT citations study: 44% from first third of content*. https://almcorp.com/blog/chatgpt-citations-study-44-percent-first-third-content/

Anthropic. (2026). *Does Anthropic crawl data from the web, and how can site owners block the crawler?* https://support.claude.com/en/articles/8896518-does-anthropic-crawl-data-from-the-web-and-how-can-site-owners-bloc

Chapekis, A., & Lieb, A. (2025). *Google users are less likely to click on links when an AI summary appears in the results*. Pew Research Center. https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summar

Chen, M., Wang, X., Chen, K., & Koudas, N. (2025). *Generative Engine Optimization: How to Dominate AI Search*. arXiv. https://arxiv.org/abs/2509.08919

Google. (2026a). *Optimizing your website for generative AI features on Google Search*. Google Search Central. https://developers.google.com/search/docs/fundamentals/ai-optimization-guide

Google. (2026b). *AI features and your website*. Google Search Central. https://developers.google.com/search/docs/appearance/ai-features

Grossman, R., Liu, S., Chen, M. K., Smith, M., Borcea, C., & Chen, Y. (2026). *How Generative AI Disrupts Search: An Empirical Study of Google Search, Gemini, and AI Overviews*. arXiv. https://arxiv.org/abs/2604.27790

Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). *Dense Passage Retrieval for Open-Domain Question Answering*. Proceedings of EMNLP 2020. https://arxiv.org/abs/2004.04906

Khosravi, M., & Yoganarasimhan, H. (2026). *Impact of AI Search Summaries on Website Traffic: Evidence from Google AI Overviews and Wikipedia*. arXiv. https://arxiv.org/abs/2602.18455

Lavrinovics, E., Biswas, R., Hose, K., & Bjerva, J. (2025). *MultiHal: Multilingual Dataset for Knowledge-Graph Grounded Evaluation of LLM Hallucinations*. arXiv. https://arxiv.org/abs/2505.14101

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Kuttler, H., Lewis, M., Yih, W., Rocktaschel, T., Riedel, S., & Kiela, D. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. Advances in Neural Information Processing Systems. https://arxiv.org/abs/2005.11401

llms.txt. (2024). *llms.txt specification*. https://llmstxt.org/

Loktionova, A., & Drozdov, A. (2025). *Backlinks in AI Search: New Data on What Actually Moves Citations*. Semrush. https://www.semrush.com/blog/backlinks-ai-search-study/

Loktionova, M. (2026). *We Analyzed 89K LinkedIn URLs Cited in AI Search: Here's What Drives Visibility*. Semrush. https://www.semrush.com/blog/linkedin-ai-visibility-study/

OpenAI. (2026). *Overview of OpenAI Crawlers*. https://developers.openai.com/api/docs/bots

Peham, T. (2026). *The AI Citation Economy*. Otterly.AI. https://otterly.ai/blog/the-ai-citations-report-2026/

Perplexity. (2026). *How does Perplexity follow robots.txt?* https://www.perplexity.ai/help-center/en/articles/10354969-how-does-perplexity-follow-robots-txt

Puerto, H., Gubri, M., Green, T., Oh, S. J., & Yun, S. (2025). *C-SEO Bench: Does Conversational SEO Work?* arXiv. https://arxiv.org/abs/2506.11097

Schema.org. (n.d.). *sameAs*. https://schema.org/sameAs

Wan, A., Wallace, E., & Klein, D. (2024). *What Evidence Do Language Models Find Convincing?* Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. https://aclanthology.org/2024.acl-long.403/

Wu, Y., Zhong, S., Kim, Y., & Xiong, C. (2025). *What Generative Search Engines Like and How to Optimize Web Content Cooperatively*. arXiv. https://arxiv.org/abs/2510.11438

Xu, H., Iqbal, U., & Montgomery, J. M. (2026). *Measuring Google AI Overviews: Activation, Source Quality, Claim Fidelity, and Publisher Impact*. arXiv. https://arxiv.org/abs/2605.14021

Zhang, P., Cui, R., & Zhang, D. J. (2026). *The Impact of AI Search on the Online Content Ecosystem: Evidence from Google and Reddit*. arXiv. https://arxiv.org/abs/2605.16428

Appendix A. SEEN Checklist Summary

The full checklist lives in `framework/seen-checklist.md`. Its structure is:

- **S — Structure:** 10 items covering answer-first layout, semantic headings, visible HTML facts, comparison pages, FAQs, direct service/product answers, matching schema, renderability, anchorable long-form content, and optional `llms.txt` infrastructure.
- **E — Evidence:** 10 items covering quantitative-claim support, original research or methodology, named case studies, independent reviews, credentialed bylines, transparent pricing, attributed testimonials, linked press/talks, primary-source citations, and an About/Methodology page.
- **E — Entity:** 10 items covering canonical naming, one-sentence description consistency, JSON-LD and `sameAs`, real About page, founder-company-product relationships, Wikidata where appropriate, Google Business Profile where relevant, disambiguation, dated profile copy, and consistent photos/logos/bios.
- **N — Notability:** 11 items covering independent mentions, editorial lists, organic community presence, high-authority links, Wikipedia where appropriate, speaking/guest-writing, platform-hosted reach, partner-side proof, customer-side proof, founder trail, and credible awards.

Each item is scored `0` or `1`; the weakest layer is the priority for remediation. The checklist is a diagnostic tool, not a validated predictive score.

Compact rubric

Layer	Item codes	Short labels
Structure	S1-S10	answer-first layout; semantic headings; visible HTML facts; comparison tables; FAQs; direct product/service answers; matching schema; renderability; anchorable long-form sections; optional `llms.txt` infrastructure
Evidence	E1-E10	quantitative-claim support; original research/methodology; named case studies; independent reviews; credentialed bylines; transparent pricing; attributed testimonials; linked press/talks; primary-source citations; About/Methodology page
Entity	ID1-ID10	canonical name; consistent description; JSON-LD and `sameAs`; real About page; explicit relationships; Wikidata where appropriate; Google Business Profile where relevant; disambiguation; recently updated profile copy; consistent visual/profile identity

Notability	N1-N11	independent mentions; editorial lists; organic community mentions; high-authority links; Wikipedia where appropriate; speaking/guest-writing; platform-hosted reach; partner-side proof; customer-side proof; founder/key-person trail; credible awards

Appendix B. Source Classification

Class	Role in paper	Examples	Claim weight
Official platform documentation	Vendor-stated behavior, eligibility, crawler/user-agent distinctions	Google Search Central, OpenAI crawler docs, Anthropic crawler docs, Perplexity help/docs	High for what the vendor states; not evidence of hidden ranking weights
Academic mechanism evidence	Retrieval, passage selection, evidence evaluation, entity grounding	Lewis et al. (2020), Karpukhin et al. (2020), Wan et al. (2024), Lavrinovics et al. (2025)	High for mechanism plausibility; indirect for production search outcomes
Empirical AI-search studies	Production or quasi-production behavior in defined systems	Chen et al. (2025), Grossman et al. (2026), Xu et al. (2026), Khosravi & Yoganarasimhan (2026), Zhang et al. (2026)	Medium to high within study scope; system- and time-bounded
Industry observational studies	Current market/citation patterns with inspectable but vendor-mediated methods	Semrush, Otterly.AI, ALM/Indig	Directional unless independently replicated
Practitioner operationalization	Proposed checklist and framework variables	SEEN checklist, source notes, platform inventory	Hypothesis-generating; requires validation