



Generative Engine Optimisation and Brand Visibility in AI Search:

Ansh Yadav

School of Business Management, NMIMS, Mumbai

Abstract – Consumers increasingly meet brands not in a list of ranked links but inside a single synthesised answer written by a large language model (LLM). This paper develops a conceptual framework for Generative Engine Optimisation (GEO), the practice of shaping the probability that a brand is retrieved, cited, named, and recommended by generative search engines, and synthesises the empirical evidence available as of September 2026. We formalise brand presence as a four-stage retrieval–citation–recommendation funnel with a parallel parametric-memory route, define a family of measurable outcomes (citation share, share of model voice, ghost-citation rate, recommendation rank) and combine them into a Brand Visibility Index (BVI). We then synthesise eight primary and secondary studies, graded by evidential tier, spanning the original GEO benchmark of Aggarwal et al. (2024), the earned-media audit of Chen et al. (2025), the verifiability audit of Liu et al. (2023), and industry and institutional studies of click behaviour and brand mentions. Four regularities emerge. First, the click economy around AI summaries is contracting: two independent designs report roughly 40–47% fewer organic clicks when a summary is shown. Second, content-level interventions that raise evidence density (quotations, statistics, source citations) improve inclusion in generated answers by up to about 40% in controlled tests, while keyword stuffing does not. Third, engines exhibit a systematic preference for third-party, earned-media sources over brand-owned content, and branded web mentions correlate with AI Overview brand mentions at 0.664 against 0.218 for backlinks. Fourth, being cited and being named are distinct events: in one industry audit, 61.7% of domain appearances were citation-only. We derive eight testable propositions, specify a pre-registrable audit and field-experiment protocol, and discuss ethical limits, including the boundary between legitimate optimisation and adversarial manipulation. No new primary data are reported; the contribution is conceptual and integrative.

Keywords – Generative Engine Optimisation, Answer Engine Optimisation, AI Search, Brand Visibility, Retrieval-Augmented Generation, Earned Media, Citation Share, Share of Model Voice, Large Language Models, Digital Marketing.

I. INTRODUCTION

For twenty years, online brand discovery has been organised around a simple object: a ranked list of links. A brand that wished to be found competed for position, and the discipline of search engine optimisation (SEO) grew up around the mechanics of that competition: keywords, backlinks, page speed, and the slow accumulation of domain authority. The unit of success was the click.

That object is being replaced. Generative engines such as ChatGPT with web search, Perplexity, Google's AI Overviews and AI Mode, and Gemini retrieve documents, pass them to a language model, and return a single composed answer with a handful of cited sources. Aggarwal et al. (2024) formalise this class of system as a generative engine and note that it satisfies queries by synthesising information from multiple sources rather than presenting them. For a brand, the consequence is structural. The question is no longer "what position do we hold?" but "are we in the answer, and in what words?"

The commercial stakes are already visible in behavioural data. Using browsing records from 900 U.S. adults, Chapekis and Lieb (2025) report that users clicked a traditional result link in 8% of visits when an AI summary appeared, against 15% when it did not, and clicked a link inside the summary in only 1% of visits. A randomised field experiment by Agarwal and Sen (2026) finds a 39.8% fall in outbound organic clicks when the summary is shown, with no measurable difference in user satisfaction.

Whatever the precise magnitude, brand exposure is migrating from the destination page to the answer itself. A brand that is absent from the answer may never be considered; a brand that is present may be influential even if nobody clicks.

At the same time, the practice of GEO is running ahead of its evidence base. Vendor dashboards report "AI visibility scores", agencies sell optimisation retainers, and trade blogs assert rules of thumb, yet the peer-reviewed literature is thin. The original GEO paper (Aggarwal et al., 2024) addresses content creators and measures the visibility of sources, not the presence of brands. Later work (Chen et al., 2025) documents systematic differences in the kinds of sources engines prefer, but stops short of a brand-level theory. Industry studies supply large samples but with commercial interests and heterogeneous definitions of "visibility".

This paper attempts to close part of that gap. It does not collect new data. Instead it (i) builds a brand-level conceptual framework that connects the information-retrieval mechanics of generative engines to established marketing constructs such as awareness and consideration; (ii) synthesises the best available evidence under an explicit grading scheme; and (iii) converts the synthesis into falsifiable propositions and a concrete study protocol.



ISSN:3048-7722

Research Objectives

- **The paper pursues three interrelated objectives:**

O1. Define brand visibility in generative engines precisely enough to be measured, separating retrieval, citation, mention, and recommendation, and distinguishing the retrieval route from the parametric-memory route.

O2. Synthesise what is currently known about the determinants of that visibility, grading each source by the strength of its design, and identifying where the evidence is causal, correlational, or merely anecdotal.

O3. Specify a research agenda: testable propositions and an empirical protocol that a subse-quent study could execute and pre-register.

Principal Contributions

The paper makes four contributions to the emerging literature on generative search:

1. A funnel framework that decomposes brand presence into retrieval, citation, mention, and recommendation events, with a parallel parametric route, and explains why citation and mention can diverge (Section 3).
2. A measurement vocabulary: citation share, share of model voice, ghost-citation rate, recommendation rank, and a composite Brand Visibility Index, each with an explicit formula (Section 3.2).
3. A graded evidence synthesis of eight studies that separates causal, correlational, and reported evidence and shows where they agree (Sections 4 and 5).
4. Eight testable propositions and an audit-plus-field-experiment protocol, including a power calculation, that can be pre-registered (Sections 6 and 7).

The remainder of the paper is organised as follows. Section 2 surveys the relevant literature. Section 3 develops the framework. Section 4 describes the evidence base and grading. Section 5 synthesises the findings. Section 6 states the propositions. Section 7 sets out the research design. Section 8 draws managerial implications. Section 9 discusses limitations and ethics, and Section 10 concludes.

II. LITERATURE REVIEW

From Ranked Lists to Synthesised Answers

Classical web search returns an ordered list and leaves evaluation to the user. This design has two properties that shaped brand strategy. Exposure was ordinal, since position determined attention, and attributable, since a click carried the user to the brand's own property. The generative engine breaks both. Exposure becomes compositional : the brand is a phrase, a sentence, or a footnote inside an answer whose wording the brand does not control. Attribution becomes partial, because the engine may name a brand without linking to it or link to a page without naming the brand (Section 5).

Economic theory offers a useful reading. Nelson (1970) distinguished search goods, whose quality can be checked before purchase, from experience goods, whose quality is learned by use. Generative engines act as a compressed search stage in which the cost of evaluating alternatives is

borne by the model rather than the user, which is precisely the mechanism the consideration-set literature says drives choice: buyers admit a small number of options because evaluation is costly (Hauser and Wernerfelt, 1990). If the engine now performs the first screening, then inclusion in its answer functions as inclusion in the consideration set.

Retrieval-Augmented Generation and Source Selection

Most production generative engines follow the retrieval-augmented generation (RAG) pattern of Lewis et al. (2020): a retriever selects candidate documents, and a generator conditions on them. Two properties of this architecture matter for brands. First, the retriever, not the generator, decides which brands are eligible. A page that is never retrieved cannot be cited. Second, the generator's use of retrieved text is imperfect. Liu et al. (2024) show that language models use information at the beginning and end of a long context more reliably than information in the middle, which implies that where a brand's passage sits in the assembled context can affect whether it is used.

Grounding is also imperfect at the output stage. Liu et al. (2023) audited four commercial generative search engines and found that only 51.5% of generated sentences were fully supported by their citations, and only 74.5% of citations supported the statements they were attached to. For brands this cuts both ways: a citation is not proof of endorsement, and a favourable statement may not be traceable to the cited source.

Engines can also answer from parametric memory, that is, from associations learned in pre-training and fine-tuning, without consulting any live document. A brand that is heavily discussed across the training corpus can be named even when retrieval is switched off. The framework in Section 3 therefore treats retrieval and parametric memory as two routes to the same outcome.

Generative Engine Optimisation and Content-Level Interventions

Aggarwal et al. (2024) introduced GEO as a black-box optimisation framework for improving a source's visibility in generative responses. They built GEO-bench, comprising 10,000 queries across 25 domains, defined visibility metrics (a position-adjusted word count and a subjective impression score), and tested nine content transformations. Adding quotations, statistics, and source citations produced the largest gains, up to roughly 40% in relative visibility, while keyword stuffing, the staple of early SEO, produced negligible gains. Effects were domain-specific, and lower-ranked sources benefited disproportionately: in one reported result, the cite-sources transformation raised visibility by 115.1% for sources ranked fifth and lowered it by 30.3% for sources ranked first. A validation on the deployed engine Perplexity.ai reported gains of up to 37% on the subjective metric.

Two features limit direct managerial transfer. The unit of analysis is the source page, not the brand, and the metric is share of response text, not the probability of being named



ISSN:3048-7722

or recommended. The bridge from the first to the second is an assumption this paper makes explicit rather than an established result.

Brand Visibility, Equity, and Consideration

Marketing theory gives the outcome variable its meaning. Keller (1993) decomposes brand knowledge into awareness (recognition and recall) and image, and Aaker (1991) lists awareness among the core assets of brand equity. Generative engines are best understood as a new arena in which recall is elicited: a user who asks “which mutual fund platform should I use?” is performing an aided-recall task, and the engine’s answer is the retrieved set. Signalling theory (Spence, 1973) explains why third-party mentions might carry more weight than self-description: an assertion from an independent source is costlier to fake than an assertion on the brand’s own site. Dual-process models of persuasion (Petty and Cacioppo, 1986) suggest a further

reason why concrete, checkable cues such as statistics and quotations may help: they act as peripheral credibility cues for a model and a reader alike.

Measurement, Manipulation, and Instability

Measurement is hard because outputs are stochastic, personalised, and engine-specific. Industry trackers sample fixed prompt sets repeatedly and report the share of responses that mention or cite a brand, but definitions vary and few have been externally validated. On the adversarial side, Kumar and Lakkaraju (2024) show that a short optimised text sequence inserted into a product page can substantially raise the rank at which a language model recommends that product, demonstrating that the recommendation stage is manipulable and that a line separates content improvement from exploitation. This paper treats that line as an ethical boundary (Section 9).

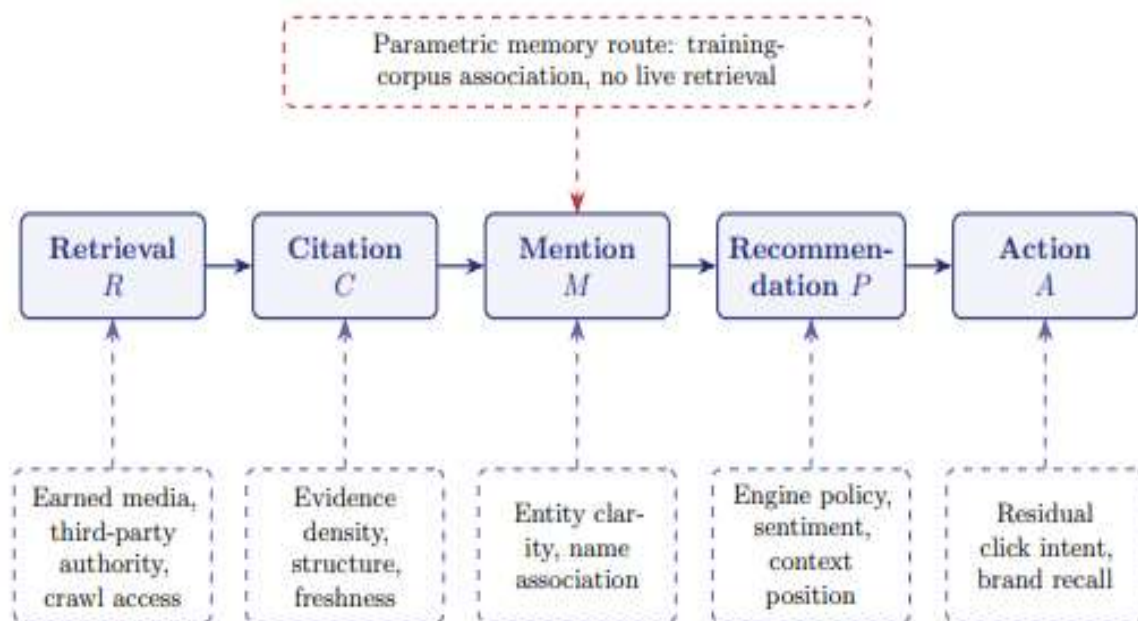


Figure 1: The retrieval–citation–recommendation funnel with its drivers (dashed boxes) and the parametric-memory route to a mention. Stages are treated as a sequence for exposition; the evidence in Section 5 shows the true process branches.

III. CONCEPTUAL FRAMEWORK

The Retrieval–Citation–Recommendation Funnel

Consider a user query q submitted to engine e , and a focal brand b . Figure 1 depicts the sequence of events through which b can appear in the answer.

The stages are defined as follows.

Retrieval (R). Content associated with b is fetched into the model’s context. This stage sets eligibility and is governed largely by the engine’s retriever, which in turn responds to authority, relevance, and crawlability.

Citation (C). The engine attaches a source link that belongs to b ’s domain to the answer.

Mention (M). The brand name appears in the body of the generated answer, with or without a link.

Recommendation (P). The brand is presented as a suggested option, ideally at a favourable position in the answer and with positive framing.

Action (A). The user clicks, searches for the brand, or later chooses it.

Formal Definitions

Let Q be a set of queries with weights w_q summing to one, E a set of engines, and r an index for repeated runs of the same prompt. Define indicator variables C_{bqer} , M_{bqer} , and P_{bqer} for whether brand b is cited, named, and recommended in run r on engine e for query q . Then:



ISSN:3048-7722

$$\text{Citation share: } CS_{be} = \sum_q w_q \bar{C}_{bqe} \quad (1)$$

$$\text{Share of model voice: } SMV_{be} = \sum_q w_q \bar{M}_{bqe} \quad (2)$$

$$\text{Ghost-citation rate: } GCR_{be} = \frac{\sum_{q,r} C_{bqer}(1 - M_{bqer})}{\sum_{q,r} C_{bqer}} \quad (3)$$

$$\text{Recommendation rank: } RR_{be} = \mathbb{E}[1/\text{rank}_{bqer} | P_{bqer} = 1] \quad (4)$$

where a bar denotes the mean over repeated runs. Averaging over r is essential because generation is stochastic. The ghost-citation rate captures the share of citations that deliver a link but no name. Let $\text{Sentbe} \in [0, 1]$ denote the mean coded sentiment of mentions. A composite Brand Visibility Index is then:

$$BV_b = \sum_{e \in \mathcal{E}} w_e [\alpha CS_{be} + \beta SMV_{be} + \gamma RR_{be} + \delta \text{Sentbe}_e], \quad \alpha + \beta + \gamma + \delta = 1, \quad \sum_e w_e = 1. \quad (5)$$

The weights w_e (engine importance) and $(\alpha, \beta, \gamma, \delta)$ are normative choices that depend on the brand's objectives. A brand seeking traffic should weight α heavily, while a brand seeking awareness in a category with low click-through should weight β . We recommend reporting results under equal weights and under a sensitivity range, rather than a single opaque score.

Two Routes to a Mention

Because an engine can name a brand from memory or from retrieved text, the probability that brand b is named is best written as the union of two routes:

$$\Pr(M_b) = 1 - (1 - \pi_b^{\text{par}})(1 - \pi_b^{\text{ret}}), \quad \pi_b^{\text{ret}} = \Pr(R_{ib}) \Pr(M_b | R_{ib}).$$

(6)

Here π_{par} is the probability of a parametric mention and depends on the volume and consistency of the brand's presence in the training corpus, while π_{ret} depends on live retrieval. The decomposition has a practical consequence: interventions that only improve owned pages act on $\Pr(M_b | R_b)$ and on citation, and cannot move π_{par} , which changes slowly and only when the underlying models are retrained.

Determinants and Engine Heterogeneity

Four families of determinants map onto the stages of Figure 1.

1. **Authority and earned media (drives R).** Independent coverage from publications, reviews, forums, and reference works increases the number and trustworthiness of retrievable documents that mention the brand.
2. **Evidence density and structure (drives C).** Pages that contain checkable statistics, attributed quotations, and clear structure are easier for a generator to quote and support.
3. **Entity clarity (drives M).** A consistent name, category association, and structured description reduce the chance that the model conflates or omits the brand.
4. **Engine policy and context position (drives P).** Each engine has its own tendencies in source diversity, recency, language, and sensitivity to phrasing (Chen et al., 2025). Position within the context also matters (Liu et al., 2024).

Because engines differ in all four, the framework indexes every outcome by e . A single blended score would hide the fact that a brand can be strong on one engine and absent on another.

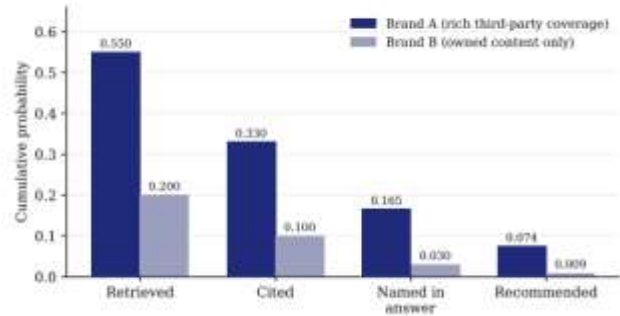


Figure 2: Illustrative cumulative funnel probabilities for two hypothetical brands. Parameters are assumptions for exposition (Appendix A); they are not empirical estimates.

Illustrative Funnel Arithmetic

Figure 2 shows why the funnel structure matters even when each stage looks moderate. The parameters are hypothetical, chosen only to illustrate compounding, and are not estimates. Brand A has rich third-party coverage and Brand B has owned content only. Stage-level advantages of between 1.2 and 2.75 times compound into an eight-fold gap in the probability of being recommended (0.074 against 0.009). Appendix A gives the parameters and extends the calculation with the parametric route.

IV. EVIDENCE BASE AND GRADING

Scope and Search

The synthesis is a structured but non-exhaustive review. Sources were located through targeted searches on 30 September 2026 for (i) the founding GEO literature, (ii) audits of which sources generative engines cite, (iii) studies of brand mentions and citations, and (iv) studies of click behaviour when AI summaries appear. Sources were included if they reported a defined sample, a stated method, and quantitative results. Sources that reported figures without a traceable method were excluded. Because the field moves quickly, a substantial share of the material is grey literature; this is flagged through the grading scheme below rather than hidden.

Evidence Tiers

Each source is assigned an evidential tier.

Tier 1. Peer-reviewed publication.

Tier 2. Working paper or preprint with a documented method.

Tier 3. Institutional or industry study with a published method but no external review. Industry studies may carry a commercial interest.

Tier 4. Secondary or trade-press report of a study whose original was not directly consulted.



ISSN:3048-7722

A second label records design strength: causal (randomised or controlled manipulation), observational (associational), or descriptive (audit or count).

The Evidence Register

Table 1 lists the eight studies that anchor the synthesis. Table 2 then gives each study's principal finding and its main limitation.

Table 1: Evidence register: sources, designs, and tiers

Study	Object and sample	Design	Tier	Checked
Aggarwal et al. (2024)	10,000 queries, 25 domains; 9 content transformations	Controlled manipulation	1	Full text
Liu et al. (2023)	4 commercial generative engines; citation support	Human audit	1	Abstract
Chen et al. (2025)	ChatGPT, Perplexity, Gemini vs. Google; source types	Comparative audit	2	Abstract
Agarwal and Sen (2026)	1,065 users; 2-week randomised AI Overview display	Field experiment	2	Secondary
Chapekis and Lieb (2025)	900 U.S. adults; 68,879 searches	Observational panel	3	Full page
Linehan (2025)	75,000 brands; Spearman correlations	Observational	3	Full page
Loktionova (2026)	3,981 domain appearances; 115 prompts; 14 countries	Descriptive audit	3	Full page
PPC Land (2026)	126M prompts; 1,200+ brands; 22 verticals	Descriptive (reported)	4	Secondary

Checked records what the author consulted directly. "Secondary" means the figures were confirmed only through a published summary of the original.

V. SYNTHESIS OF FINDINGS

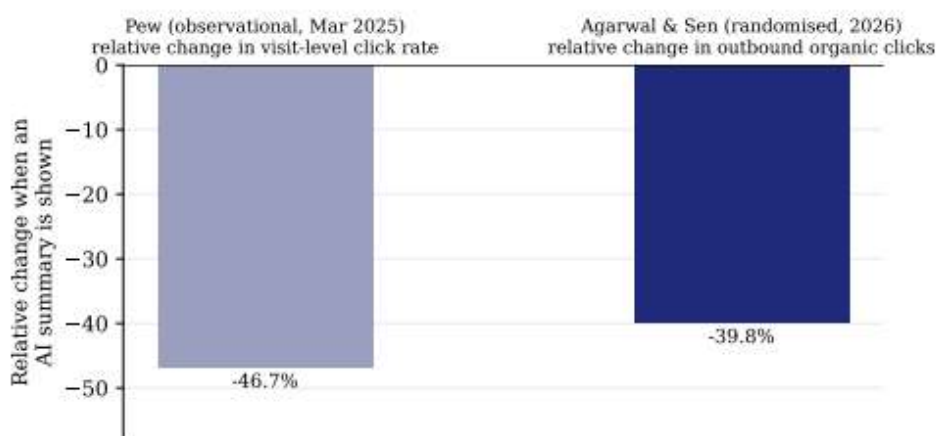


Figure 3: Reported reductions in organic clicking when an AI summary appears. The metrics differ (visit-level click rate in Pew; outbound organic clicks in Agarwal and Sen), so the bars indicate direction and rough magnitude, not a pooled estimate. The Pew figure is the author's calculation from the published rates of 8% and 15%.



Table 2: Principal findings and limitations by study

Study	Principal finding	Main limitation
Aggarwal et al. (2024)	Quotations, statistics, and citations raise source visibility by up to about 40%; keyword stuffing does not; lower-ranked sources gain most.	Measures share of source text, not brand mention; engines change.
Liu et al. (2023)	51.5% of sentences fully supported by citations; 74.5% of citations support their statements.	2023 engines; may have improved.
Chen et al. (2025)	AI search shows a systematic bias toward earned media over brand-owned and social content; engines differ in diversity, freshness, language, and phrasing sensitivity.	Preprint; specific engine versions.
Agarwal and Sen (2026)	AI Overview display cuts outbound organic clicks by 39.8% and raises zero-click searches by 34.5%; satisfaction unchanged.	Desktop Chrome, U.S., paid panel; working paper.
Chapekis and Lieb (2025)	Organic link clicks in 8% of visits with an AI summary vs. 15% without; 1% click inside the summary.	Observational; Google has disputed the query set.
Linehan (2025)	Branded web mentions correlate 0.664 with AI Overview brand mentions; backlinks 0.218.	Correlation only; vendor; DR > 40 filter; Google engine only.
Loktionova (2026)	61.7% of domain appearances were citation-only; 13.2% both; 25.1% mention-only.	Small prompt set; vendor tool; regional variation.
PPC Land (2026)	Only 36 of 1,200+ brands were in the top 100 on every platform in every month.	Reported through a press briefing.

The Click Economy Is Contracting

Two independent designs converge on the same direction. Chapekis and Lieb (2025) observe a fall from 15% to 8% in the share of visits with a traditional link click, a relative reduction of about 47%. Agarwal and Sen (2026), exploiting random assignment of AI Overview display, estimate a 39.8% fall in outbound organic clicks and find that sponsored clicks were unchanged. Figure 3 sets the two estimates side by side.

Two points deserve emphasis. The randomised design of Agarwal and Sen (2026) addresses the query-mix confound that limits observational comparisons, which raises confidence that the effect is causal rather than an artefact of which queries trigger a summary. And the finding of unchanged satisfaction suggests users are not being harmed by the summary, so a reversal of the trend by user pushback is unlikely. The implication for brands is that exposure inside the answer increasingly substitutes for exposure on the destination page.

Evidence Density Moves Inclusion, Keywords Do Not

The clearest causal evidence on content is Aggarwal et al. (2024). Across their benchmark, transformations that add checkable material (quotations, statistics, source citations) outperformed stylistic changes, and keyword stuffing, the legacy SEO tactic, delivered little. The rank-dependent

result is strategically important: sources that were not already ranked first gained most, and the top-ranked source could lose visibility from the same edit. This suggests GEO can widen access for challengers but offers incumbents little marginal return from evidence-density edits alone. The domain-specific pattern (different transformations winning in different subject areas) counsels against a universal checklist.

Third-Party Authority Dominates Owned Content

Chen et al. (2025) report that the engines they audited exhibit “a systematic and overwhelming bias” toward earned media over brand-owned and social content, in contrast to the more balanced mix in Google’s conventional results.

This is consistent with the correlational evidence of Linehan (2025): across 75,000 brands, branded web mentions correlate with AI Overview brand mentions at 0.664, branded anchors at 0.527, branded search volume at 0.392, Domain Rating at 0.326, referring domains at 0.295, and backlinks at 0.218 (Figure 4). The authors themselves stress that correlation is not causation and that all factors show moderate-to-weak correlations, and roughly 26% of brands had no AI Overview mentions at all.

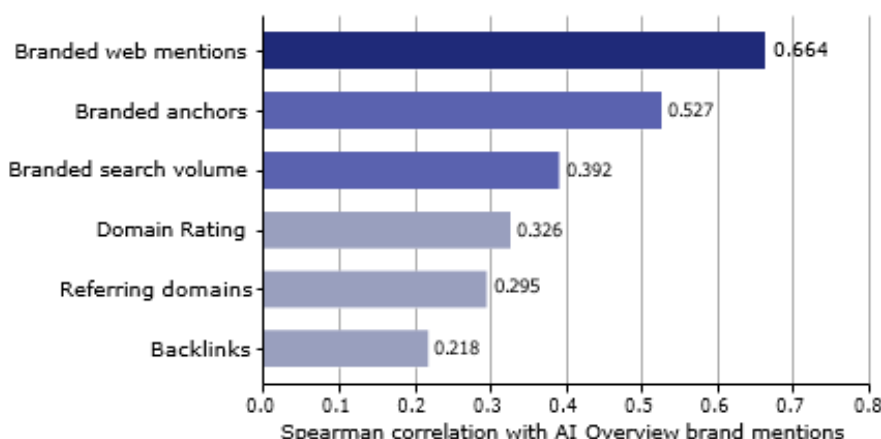


Figure 4: Spearman correlations between brand-level factors and AI Overview brand mentions (75,000 brands). Source: Linehan (2025). Correlational evidence only.

Read together, the two studies fit a signalling account (Spence, 1973): a brand talked about by many independent sources sends a signal that is costly to counterfeit. They also fit the retrieval mechanics of Section 3, where each additional independent document mentioning the brand raises the chance that at least one is retrieved.

A Citation Is Not a Mention

The audit of Loktionova (2026) distinguishes a citation (a source link to the domain) from a mention (the brand name in the answer text). Of 3,981 domain appearances, 61.7% were citation-only, 13.2% were both, and 25.1% were mention-only; the citation rate (74.9%) was almost twice the mention rate (38.3%). Dividing the citation-only share by the overall citation rate gives an implied ghost-citation rate in the sense of Equation (3) of $61.7/74.9 \approx 82\%$ (author's calculation). The split differed by platform: in the report, Gemini showed a mention rate of 83.7% against a citation rate of 21.4%, whereas ChatGPT showed the reverse (citation rate 87%, mention rate 20.7%).

This decoupling matters conceptually. A brand optimising only for citations may accumulate source links while remaining unnamed to the reader, delivering little awareness. A brand optimising only for mentions may be named without a traceable source. The framework's separation of C and M is therefore an empirical necessity, not a modelling convenience.

Grounding Is Imperfect

The verifiability audit of Liu et al. (2023) shows that citations attached to generated sentences frequently fail to support them. Although the engines audited predate 2026 systems, the finding establishes a benchmark risk: a favourable claim about a brand can be attributed to a source that does not say it, and an unfavourable claim can be attributed likewise. Brands therefore have a monitoring interest in the accuracy of what is said about them, not only its frequency.

Visibility Is Concentrated

An industry index summarised in trade press reports that, across 126 million prompts collected January–April 2026 on four platforms and more than 1,200 brands in 22 verticals, only 36 brands appeared in the top 100 most-mentioned list on every platform in every month (PPC Land, 2026). This is Tier 4 evidence and should be treated as suggestive. Its direction is consistent with the incumbent advantage that Chen et al. (2025) identify, and with the parametric route in Equation (6): well-known brands enjoy a high π_{par} that new entrants cannot quickly change.

Cross-Study Convergence

Table 3 summarises where the evidence agrees, and how strong it is. Click contraction has support from two independent designs, and earned-media dominance and the value of evidence density each from two sources using different methods; ghost citations and concentration rest on a single study or a press summary.

Table 3: Convergence of evidence across studies

Regularity	Supporting sources	Strength
Clicks fall when an AI summary appears	Chapekis and Lieb (2025); Agarwal and Sen (2026)	Strong (direction)
Evidence density raises inclusion	Aggarwal et al. (2024); Chen et al. (2025) (machine readability)	Moderate
Third-party sources outweigh owned content	Chen et al. (2025); Linehan (2025)	Moderate
Citation and mention diverge	Loktionova (2026)	Limited
Visibility concentrates in incumbents	PPC Land (2026); Chen et al. (2025)	Limited
Grounding is imperfect	Liu et al. (2023)	Moderate



VI. PROPOSITIONS

The framework and synthesis yield eight propositions, each stated so that it could be falsified by the design in Section 7.

Table 4: Testable propositions

No.	Proposition	Grounding
P1	Earned-media dominance. Holding owned-content volume constant, the number of independent third-party documents mentioning a brand predicts $\Pr(R_b)$ and SMV_{be} more strongly than owned-page volume or backlink counts.	Chen et al. (2025); Linehan (2025)
P2	Evidence density. Adding attributed quotations, statistics, and source citations to a page raises the probability that it is cited for a matched query, relative to an unedited control.	Aggarwal et al. (2024)
P3	Keyword null. Inserting additional query keywords without new information does not raise citation probability.	Aggarwal et al. (2024)
P4	Citation-mention decoupling. The ghost-citation rate GCR_{be} is materially above zero and varies significantly across engines.	Loktionova (2026)
P5	Engine heterogeneity. The overlap between the sets of brands named by different engines for the same query is low, so a single-engine index does not predict multi-engine visibility.	Chen et al. (2025)
P6	Stochastic instability. Single-run audits misclassify visibility: the run-to-run variance in M_{bqer} is large enough that at least ten repeats per prompt are needed for stable SMV_{be} .	Framework (stochastic generation)
P7	Incumbent advantage. The parametric-route probability π_b^{par} rises with brand age and pre-existing web footprint, so evidence-density edits yield smaller gains for incumbents than for challengers.	Aggarwal et al. (2024); PPC Land (2026)
P8	Click substitution. Where a summary is shown, SMV_{be} is a better predictor of subsequent branded search than citation-derived click traffic.	Chapekis and Lieb (2025); Agarwal and Sen (2026)

VII. RESEARCH DESIGN FOR EMPIRICAL VALIDATION

This section specifies, but does not execute, a two-part study. Reporting the protocol has two purposes: it lets readers judge what the synthesis cannot yet claim, and it can be lodged in a public registry before data collection, which protects against post-hoc flexibility.

Part A: Multi-Engine Brand Audit

Sampling frame. Select 12 consumer product categories, stratified across regulated and unregulated categories, and 20 brands per category (240 brands), including both market leaders and challengers. Write 25 prompts per category covering four intents (informational, comparative, transactional, and recommendation-seeking), giving 300 prompts.

Engines and repeats. Run each prompt on four engines (for example ChatGPT with search, Perplexity, Google AI Overviews or AI Mode, and Gemini) ten times each, on fresh sessions with location and language fixed, yielding 12,000 responses.

Coding. For each response, record whether each brand is cited, named, and recommended, its position, and the sentiment of the mention. Two coders code a 10% sample to report inter-rater agreement.

Model. Estimate the probability of a mention by a mixed-effects logistic regression with crossed random effects for brand, prompt, and engine:

$$\text{logit } \Pr(M_{bqer} = 1) = \beta_0 + \beta_1 \text{ Emb} + \beta_2 \text{ Ownedb} + \beta_3 \text{ Entityb} + \beta_4 \text{ Ageb} + \text{ub} + \text{vq} + \text{se}, \quad (7)$$

where Emb is the count of independent third-party mentions, Ownedb the volume of owned content, Entityb an entity-clarity score, and Ageb brand age. P1 predicts $\beta_1 > \beta_2$; P7 predicts a negative interaction between Ageb and any intervention effect. The same specification is estimated for C and for P, and the difference $C - M$ is used to test P4.

Part B: Content Field Experiment

Design. Recruit matched pairs of pages that currently rank similarly for a target query. Randomly assign pages to one of five arms: control, statistics added, quotations added, cited sources added, or combined. For P3, add a keyword-insertion arm without new information.

Outcome and power. The primary outcome is the citation probability for the matched query across repeated runs. To detect an increase from 0.20 to 0.30 with a two-sided test at $\alpha = 0.05$ and 80% power, approximately 295 page-query pairs are needed per arm, using the standard two-proportion formula

$$n = \frac{\left(z_{\alpha/2} \sqrt{2\bar{p}\bar{q}} + z_{\beta} \sqrt{p_1 q_1 + p_2 q_2} \right)^2}{(p_2 - p_1)^2}. \quad (8)$$



ISSN:3048-7722

Analysis uses a difference-in-differences comparison of pre- and post-edit citation rates with page fixed effects, and tests the treatment $\times$ incumbent-rank interaction for P7. Engines are re-sampled on the same schedule for both arms to control for model updates.

Threats to Validity

The main threats are model drift (engines update during the study, addressed by concurrent control pages and date-stamped runs), prompt sensitivity (addressed by multiple paraphrases per intent), personalisation (addressed by fixed, logged-out sessions, with the caveat that logged-in behaviour may differ), and geography and language (the protocol should be repeated in at least one non-English or India-specific setting, given evidence of cross-language inconsistency in Chen et al. (2025)).

VIII. MANAGERIAL IMPLICATIONS

Although the propositions remain to be tested, the convergent evidence supports several cautious recommendations.

1. Measure the funnel, not a single score. Track citation share, share of model voice, ghost-citation rate, and sentiment separately by engine, using repeated runs. A single “AI visibility” number obscures whether a brand is being cited, named, or recommended.
2. Invest in third-party presence. If engines favour earned media, marginal budget may earn more in expert commentary, independent reviews, credible directories, and press than in additional owned pages. This is consistent with, but does not prove, a causal role for earned media.
3. Raise evidence density on owned pages. Add sourced statistics, attributed quotations, and clear structure. Do not add keywords for their own sake.
4. Make the entity unambiguous. Use a consistent legal and trade name, category descriptors, and structured data so that the model attaches the right claims to the right entity.
5. Guard accuracy. Because grounding is imperfect (Liu et al., 2023), monitor what engines say about the brand and correct inaccurate source material at the origin.
6. Re-define success. Where a summary is shown, clicks understate influence (Chapekis and Lieb, 2025). Pair mention-based metrics with branded search, direct traffic, and assisted conversions.

Challengers Versus Incumbents

The rank-dependent result of Aggarwal et al. (2024) and the concentration reported by PPC Land (2026) point in different directions and both may be true: evidence-density edits can help a challenger enter the retrieved set on a specific query, while the incumbent’s advantage in parametric memory persists across queries. A challenger’s most credible route is therefore focused: choose narrow query clusters where it can be the best-documented option, and build independent coverage there first.

Regulated and High-Stakes Categories

In categories where advice carries legal or financial consequences, such as investment products, insurance, and health, two additional considerations apply. First, claims must remain compliant with sector rules, so evidence-density tactics must use verifiable and appropriately qualified statements rather than performance promises. Second, the cost of a wrong or misattributed statement is higher, which raises the value of the accuracy monitoring described above. Whether engines apply stricter source filters in such categories is an open empirical question that Part A of the protocol can address through its regulated-category stratum.

IX. DISCUSSION, LIMITATIONS, AND ETHICS

Limitations

Several limitations bound what this paper can claim.

1. **No primary data.** The paper is conceptual and integrative. The framework and propositions are hypotheses, not findings, and the illustrative arithmetic of Section 3 uses assumed parameters.
2. **Evidence quality.** Of eight sources, only two are peer reviewed. Four are industry or institutional studies, three of which come from vendors with a commercial interest in AI search appearing consequential. Two figures were confirmed only through secondary summaries (Table 1).
3. **Non-exhaustive search.** The review was structured but not systematic in the strict sense, and it may omit relevant studies, especially non-English and non-U.S. work.
4. **Rapid obsolescence.** Engines, retrieval policies, and interfaces change monthly. Findings from 2023 systems may not describe 2026 behaviour, and the reverse.
5. **Mixed metrics.** “Visibility” means different things across studies (share of text, mention rate, citation rate), so cross-study comparisons are directional.
6. **Contested measurement.** Google has publicly disputed the methodology behind the Pew estimates, and the extent to which click losses translate into commercial harm is not established by these studies.

Ethics: Optimisation Versus Manipulation

GEO spans a spectrum. At one end is legitimate improvement: publishing accurate, well-sourced information that helps a model and a reader. At the other is manipulation. Kumar and Lakkaraju (2024) show that text engineered for a model, rather than for a reader, can shift recommendations. Such tactics harm users, distort competition, and are likely to be penalised as engines add defences. This paper’s recommendations are confined to the first end of the spectrum. A practical test is whether a tactic would still be defensible if a reader saw exactly what the model saw.



ISSN:3048-7722

There is also a collective concern. If visibility depends on third-party coverage, brands may be tempted to purchase or fabricate coverage. Fabricated reviews and undisclosed paid placement mislead consumers and, in many jurisdictions, breach advertising and consumer-protection rules. Managers should ensure that any earned-media programme is genuine and properly disclosed.

Extensions for Future Research

1. Execute the protocol in Section 7, ideally with pre-registration and a public data release.
2. Estimate π_B by comparing responses with retrieval enabled and disabled for the same brands.
3. Link visibility to outcomes by measuring whether mention exposure changes branded search or purchase intent in an experiment.
4. Test multilingual and regional settings, including Indian languages and India-specific brand sets.
5. Model competition between brands as a game in which each brand's evidence investments change the retrieved set for all.

X. CONCLUSION AND FUTURE DIRECTIONS

This paper has argued that brand discovery is moving from a contest over position to a contest over inclusion in a composed answer, and has offered a framework in which that inclusion is decomposed into retrieval, citation, mention, and recommendation, with a parametric-memory route running alongside. The framework yields precise metrics (citation share, share of model voice, ghost-citation rate, recommendation rank) and a composite index whose weights are explicitly normative.

Synthesising eight studies, the paper finds that the evidence converges on four points: clicks fall when a summary is shown; evidence-dense content is more likely to be included than keyword-stuffed content; engines lean toward independent, third-party sources over brand-owned pages;

and being cited is not the same as being named. These regularities point to a practical reorientation. The scarce asset is no longer a page that ranks but a brand that is independently and consistently described across the web in terms a model can quote accurately.

The evidential base is nonetheless thin and largely vendor-generated, and no claim here should be read as established causal fact. The value of the eight propositions and the accompanying protocol is that they turn a fast-moving practitioner conversation into questions that can be answered. Until they are, managers should measure carefully, invest in genuine third-party credibility, and resist tactics that would not survive a reader's inspection.

A. Illustrative Funnel Parameters

Table 5 gives the hypothetical stage probabilities behind Figure 2. They are assumptions chosen to show compounding and were not estimated from data.

Table 5: Hypothetical stage probabilities and cumulative outcomes

Stage	Brand A (rich third-party)		Brand B (owned only)	
	Conditional	Cumulative	Conditional	Cumulative
Retrieved $\Pr(R)$	0.55	0.550	0.20	0.200
Cited $\Pr(C R)$	0.60	0.330	0.50	0.100
Named $\Pr(M C)$	0.50	0.165	0.30	0.030
Recommended $\Pr(P M)$	0.45	0.074	0.30	0.009

The final-stage ratio is $0.07425/0.009 \approx 8.25$. Now add the parametric route of Equation (6) with assumed $\pi_{\text{par}} = 0.10$ and $\pi_{\text{par}} = 0.03$, and take the cumulative naming probability as $\pi_{\text{ret}} (0.165$

A B

and 0.030). Then $\Pr(MA) = 1 - (0.90)(0.835) = 0.2485$ and $\Pr(MB) = 1 - (0.97)(0.97) = 0.0591$,

a ratio of about 4.2. The parametric route therefore compresses the gap at the naming stage, illustrating why a challenger with weak retrieval can still be named occasionally, and why an incumbent's memory advantage is hard to erode.

B. Audit Protocol Details

Table 6: Proposed coding scheme for each engine response

Variable	Definition	Type
cited	A link to the brand's registered domain appears among the answer's sources	Binary
named	The brand name or a recognised alias appears in the answer text	Binary
recommended	The brand is offered as a suggested option or ranked list entry	Binary
position	Order of first appearance among named brands	Integer
sentiment	Coder rating of the tone of the mention	Ordinal (1–5)
source type	Earned, owned, social, or reference source for each citation	Categorical
run id	Repeat index for the same prompt, engine, and day	Integer
engine version	Model or interface label captured at run time	String

Prompt intents. Informational (“what is. . .”), comparative (“X versus Y”), transactional (“where to buy. . .”), and recommendation-seeking (“best. . . for. . .”). Each category receives an equal number of prompts per intent, phrased in at least two paraphrases.

Session controls. Logged-out sessions, fixed location and language, cleared history, prompts issued in randomised order, and all runs date-stamped.



ISSN:3048-7722

Pre-registration. Hypotheses P1–P8, the model in Equation (7), the exclusion rules, and the stopping rule should be lodged in a public registry before any data are collected.

C.Source Register with Locations

Table 7: Where each source was located

Source	Location
Aggarwal et al. (2024)	arXiv:2311.09735; doi.org/10.1145/3637528.3671900
Chen et al. (2025)	arXiv:2509.08919
Liu et al. (2023)	arXiv:2304.09848; ACL Anthology 2023.findings-emnlp.467
Agarwal and Sen (2026)	papers.ssrn.com, abstract id 6513059
Chapekis and Lieb (2025)	pewresearch.org/short-reads/2025/07/22/
Linehan (2025)	ahrefs.com/blog/ai-overview-brand-correlation/
Loktionova (2026)	semrush.com/blog/the-ghost-citations-study/
PPC Land (2026)	ppc.land
Kumar and Lakkaraju (2024)	arXiv:2404.07981

REFERENCES

1. Aaker, D. A. (1991). *Managing Brand Equity: Capitalizing on the Value of a Brand Name*. Free Press, New York, NY.
2. Agarwal, S. and Sen, A. (2026). The impact of Google AI Overviews on publisher traffic and user experience: Evidence from a field experiment. SSRN Working Paper, abstract id 6513059 (posted April 3, 2026; revised June 17, 2026).
3. Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K., and Deshpande, A. (2024). GEO: Generative engine optimization. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 5–16. doi:10.1145/3637528.3671900.
4. Chapekis, A. and Lieb, A. (2025). Google users are less likely to click on links when an AI summary appears in the results. Pew Research Center, Short Reads, July 22, 2025.
5. Chen, M., Wang, X., Chen, K., and Koudas, N. (2025). Generative engine optimization: How to dominate AI search. arXiv preprint arXiv:2509.08919.
6. Hauser, J. R. and Wernerfelt, B. (1990). An evaluation cost model of consideration sets. *Journal of Consumer Research*, 16(4):393–408.
7. Keller, K. L. (1993). Conceptualizing, measuring, and managing customer-based brand equity.
8. *Journal of Marketing*, 57(1):1–22.
9. Kumar, A. and Lakkaraju, H. (2024). Manipulating large language models to increase product visibility. arXiv preprint arXiv:2404.07981.
10. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Ku"ttler, H., Lewis, M., Yih, W.-t., Rockta"schel, T., Riedel, S., and Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.
11. Linehan, L. (2025). An analysis of AI Overview brand visibility factors (75K brands studied).
12. Ahrefs Blog, May 26, 2025.
13. Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., and Liang, P. (2024). Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12:157–173.
14. Liu, N. F., Zhang, T., and Liang, P. (2023). Evaluating verifiability in generative search engines.
15. In *Findings of the Association for Computational Linguistics: EMNLP 2023*.
16. Loktionova, M. (2026). Why 62% of AI citations don't lead to brand mentions. Semrush Blog, June 9, 2026.
17. Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2):311–329.
18. Petty, R. E. and Cacioppo, J. T. (1986). *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. Springer, New York, NY.
19. PPC Land (2026). Semrush: 36 brands win AI visibility everywhere, 1,200 vanish on one.
20. Trade-press report on the Semrush AI Visibility Index 2026, June 2026.
21. Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374.