INTERNATIONAL JOURNAL OF DATA SCIENCE AND MACHINE LEARNING (ISSN: 2692-5141)
VOL.06 ISSUE01 (2026)

AMERICAN ACADEMIC PUBLISHER

Research Article

Strategic GEO: How Generative Engine Optimization Reshapes Competitive Advantage in Consumer Markets

[1]**Pratik Khedekar**

[1]Independent Researcher, USA

[2]**Gaurav Bansal**

[2]Independent Researcher, USA

check for updates

Received: 12 November 2025
Revised: 9 December 2025
Accepted: 18 January 2025
Published: 27 January 2026

Doi: 10.55640/ijdsml-06-01-02

Page No: 6-18

Abstract

Artificial intelligence systems like ChatGPT and Claude are fundamentally changing how consumers discover products, with some brands achieving dramatically higher AI visibility than seemingly equivalent competitors despite similar market positions, traditional search rankings, and marketing investments. This paper provides the first systematic empirical analysis of competitive Generative Engine Optimization across six diverse consumer product categories, examining how optimization sophistication relates to AI citation outcomes under varying competitive conditions. We systematically measure optimization levels for six brands across plant-based protein and running shoes category using structured coding of website characteristics across four dimensions including structured data implementation, citation quality, content comprehensiveness, and technical optimization. We query four major AI platforms—ChatGPT, Claude, Perplexity, and Google Gemini—with thirty to fifty category-relevant queries per category and code citation patterns. We employ logistic regression with category fixed effects and clustered standard errors to examine relationships between optimization investment and citation outcomes while controlling for market share, brand age, and baseline competitive position. We find that optimization sophistication strongly predicts AI citation frequency, with patterns suggesting that challenger brands capture asymmetric competitive advantages relative to market leaders for equivalent optimization investments. These findings reveal that AI-mediated discovery creates novel competitive dynamics where optimization responsiveness matters more than traditional brand equity, with important implications for marketing strategy as commerce becomes increasingly AI-intermediated.

Keywords: Generative Engine Optimization, competitive dynamics, AI citations, brand positioning, market structure, digital strategy, consumer behavior, search optimization.

Introduction

The rise of generative AI has fundamentally altered how consumers discover and evaluates products. In March 2025, two plant-based protein brands found themselves at an unexpected competitive crossroads despite commanding nearly identical market positions. Both held approximately nine percent market share in the United States plant-based protein segment, invested roughly $2.3 million annually in digital advertising, maintained similar social media followings of 145,000, and appeared consistently in the top five organic search results for key commercial terms. Yet when consumers posed product queries to AI assistants like ChatGPT, Claude, and Perplexity, Brand_1 appeared in eighty-seven percent of AI-generated responses while Brand_2 surfaced in merely twelve percent. By the end of the second quarter, Brand_1 projected an additional $4.7 million in annual revenue attributable to AI-referred customers, while Brand_2 saw negligible lift from this emerging channel. This dramatic divergence signals a fundamental shift in competitive dynamics that traditional frameworks struggle to explain.

The economic stakes of this transformation are substantial. Recent research demonstrates that AI-generated overviews reduce organic click-through rates by forty-seven percent, with sixty percent of searches now concluding without any click-through to external websites. Analysis of 590 million searches reveals that AI features reduce traffic to top-ranking pages by an average of 34.5 percent, even for websites maintaining previous search rankings. For consumer brands collectively investing over four hundred billion dollars annually in digital marketing, this represents a potential thirty to fifty percent erosion in organic traffic within three to five years, fundamentally threatening return on marketing investment.

Critical empirical questions remain unanswered. Do citation benefits accrue primarily to market leaders leveraging existing advantages, or do challengers gain disproportionate benefits? How does market structure—fragmented versus concentrated—affect optimization effectiveness? Whether optimization creates winner-take-all dynamics or levels the competitive playing field has important implications for strategy and market evolution.

This paper examines competitive Generative Engine Optimization dynamics through systematic real-world analysis spanning six consumer product categories selected to vary along theoretically relevant dimensions including market concentration, product differentiation, and information complexity. We measure optimization sophistication for forty-eight brands by systematically coding publicly observable content characteristics across four dimensions: structured data implementation that facilitates machine-readable content extraction, citation quality reflecting links to authoritative external sources, content comprehensiveness measured through information depth and breadth, and technical infrastructure supporting performance and accessibility. Each dimension is scored on a zero-to-ten scale based on specific observable criteria, yielding aggregate optimization scores ranging from zero to forty points.

We query four major AI platforms—ChatGPT using GPT-4, Claude using Sonnet 4, Perplexity AI, and Google Gemini—with thirty to fifty standardized queries per category representing diverse consumer information needs including informational queries seeking category understanding, comparative queries requesting recommendations, and attribute-specific queries focusing on product characteristics. We record which brands receive citations in AI responses and measure citation prominence through word count and positioning. Through logistic regression analysis with category fixed effects and clustered standard errors, we test whether optimization sophistication predicts AI citation probability while controlling for traditional competitive advantages including market share, brand age, and baseline search rankings. We further examine whether optimization effects vary systematically by competitive position and market structure through interaction specifications.

This research makes three primary contributions. First, it provides systematic empirical evidence that optimization sophistication correlates strongly with AI citation outcomes across diverse product categories, establishing optimization as an important competitive dimension in AI-mediated markets. The evidence suggests that brands achieving higher optimization scores experience substantially elevated citation frequencies, with patterns consistent across multiple product categories and AI platforms. Second, it documents that optimization effects vary substantially by competitive position, with patterns suggesting challenger brands experience larger marginal benefits from optimization than market leaders. This asymmetry has important implications for competitive dynamics, as it suggests optimization may create opportunities for market disruption rather than simply reinforcing existing hierarchies. Third, it develops and validates a measurement framework for assessing brand optimization levels using publicly observable website characteristics, enabling future longitudinal research tracking competitive evolution as AI systems continue developing and as brands adapt their optimization strategies over time.

LITERATURE SURVEY

[1] GEO: Generative Engine Optimization (Aggarwal et al., 2023) formalizes the emerging challenge of content visibility in generative search engines that synthesize answers from multiple sources with inline citations, fundamentally disrupting traditional search paradigms where visibility was determined by list-based ranking rather than content integration into synthesized responses. The authors introduce Generative Engine Optimization as a creator-centric, black-box framework for improving content visibility through strategic content transformations, accompanied by GEO-bench, a benchmark comprising 10,000 diverse queries spanning multiple domains and query intents for systematic evaluation. The methodology treats generative engines as black boxes and measures visibility using both objective metrics based on word attribution and position-adjusted weighting, and subjective metrics derived from LLM-based evaluation of influence, relevance, and prominence in generated responses. The work establishes that visibility in generative search depends fundamentally on how content integrates into natural language synthesis rather than retrieval rank alone, with strategies disproportionately benefiting lower-ranked sources and suggesting potential for redistributing visibility beyond traditional top-ranked pages, though the rapidly evolving nature of generative engines necessitates continual adaptation of optimization strategies.

[2] Patrick Lewis and his team have worked on the problem of Large pre-trained language models store factual knowledge implicitly in parameters, making it difficult to update knowledge, provide provenance, and avoid hallucinations on knowledge-

intensive tasks. Prior retrieval-augmented approaches focused mainly on extractive QA, not general-purpose generation tasks. Based on the experimentation, they were able to achieve state-of-the-art Exact Match on multiple open-domain QA benchmarks, outperforming extractive and parametric-only models. RAG generates more factual, specific, and diverse text than BART on generation tasks. Can answer correctly even when the exact answer does not appear verbatim in retrieved documents, leveraging parametric knowledge.

[3] Dense Passage Retrieval for Open-Domain Question Answering (Karpukhin et al., 2020) addresses the critical bottleneck of passage retrieval in open-domain QA systems, where traditional sparse methods like BM25 struggle to capture semantic similarity beyond keyword matching. The authors propose DPR, a straightforward dual-encoder architecture that learns dense vector representations for questions and passages using separate BERT encoders, computing similarity through dot products in a shared embedding space. Unlike prior dense retrieval approaches that required expensive auxiliary pretraining tasks, DPR trains directly on question-passage pairs using a contrastive objective with in-batch negatives, where passages from other examples in the same batch serve as negatives to improve training efficiency. Notably, the work establishes that retrieval quality is a primary performance driver in open-domain QA and shows that effective dense retrieval can be achieved without complex pretraining schemes, requiring as few as 1,000 training examples to surpass traditional sparse methods.

[4] Lost in the Middle: How Language Models Use Long Contexts (Liu et al., 2023) challenges the assumption that extending context windows necessarily improves language model performance, revealing fundamental limitations in how models access information distributed across long inputs. Through controlled experiments on multi-document question answering and synthetic key-value retrieval tasks, the authors systematically vary both context length and the position of relevant information to expose a consistent U-shaped performance pattern across multiple models including GPT-3.5, Claude, and various open-source alternatives. This "serial-position effect" demonstrates that models exhibit strong primacy and recency biases, performing best when critical information appears at the very beginning or end of the context while showing dramatic performance degradation—sometimes falling below closed-book baselines—when the same information is positioned in the middle portions of long contexts.

[5] SELF-RAG: Learning to Retrieve, Generate, and Critique Through Self-Reflection (Asai et al., 2023) addresses fundamental limitations in retrieval-augmented generation where models either hallucinate due to over-reliance on parametric knowledge or retrieve documents indiscriminately without assessing necessity or relevance. The authors propose SELF-RAG, a unified framework that trains a single language model to adaptively decide when retrieval is needed, evaluate retrieved passage relevance, generate responses, and critique whether its outputs are supported by evidence—all through special reflection tokens (Retrieve, ISREL, ISSUP, ISUSE) that enable explicit reasoning about the generation process. The work establishes that adaptive, self-reflective retrieval with explicit evidence verification during generation—rather than post-hoc filtering—substantially reduces hallucinations while maintaining response quality, though this comes at the cost of additional training infrastructure for reflection token annotation and increased inference overhead from segment-level evaluation.

[6] RankRAG: Unifying Context Ranking with Retrieval-Augmented Generation in LLMs (Chen et al., 2024) addresses the inefficiency of standard RAG pipelines where language models struggle to utilize large numbers of retrieved passages effectively, despite extended context windows, while separate reranking models add system complexity with limited generalization. The authors propose RankRAG, a unified framework that trains a single language model to perform both context ranking and answer generation by framing relevance estimation as a question-answering task during instruction tuning. The work establishes that context ranking and generation capabilities mutually reinforce each other when trained jointly, enabling language models to serve as effective rerankers without requiring separate specialized models, though this introduces additional inference latency from the reranking process.

[7] Enabling Large Language Models to Generate Text with Citations (Gao et al., 2023) addresses the critical challenge of LLM hallucination and verification difficulty by introducing ALCE, the first reproducible benchmark for automatically evaluating citation quality in generated text at the statement level. The evaluation shows that simple retrieval-in-context prompting performs competitively against more complex interactive approaches, while reranking multiple generations significantly improves citation quality, though retrieval quality remains the primary bottleneck. The work establishes that citation quality constitutes a distinct evaluation dimension from fluency and correctness, requiring explicit assessment, and that language models struggle to synthesize and accurately cite evidence when working with numerous passages simultaneously.

[8] News Source Citing Patterns in AI Search Systems (Gao et al., 2025) provides the first large-scale empirical analysis of news citation behavior in AI-powered search systems, examining over 366,000 citations from 12 models across OpenAI, Google, and Perplexity using real-world user interactions from the AI Search Arena dataset. The research establishes that AI search systems already function as powerful algorithmic gatekeepers shaping information exposure, with citation presence increasing perceived trust regardless of whether users actively assess citation quality, potentially reinforcing winner-take-all dynamics

among news outlets. While the observational methodology cannot isolate causal mechanisms within proprietary systems and focuses exclusively on news citations representing approximately nine percent of total citations, the findings provide critical evidence that source concentration and political bias are embedded features of current AI search architectures rather than isolated anomalies.

[9] CC-GSEO-Bench: A Content-Centric Benchmark for Measuring Source Influence in Generative Search Engines (Zhou et al., 2025) addresses the fundamental gap between retrieval and influence in generative search by introducing the first large-scale content-centric benchmark where individual articles are paired with clusters of related queries, enabling creator-focused evaluation across multiple user intents. The framework proposes three core influence dimensions—Exposure (visibility), Faithful Credit (attribution accuracy), and Causal Impact (marginal contribution)—measured through counterfactual evaluation comparing answers generated with and without the target source, alongside content-quality dimensions assessing readability and trustworthiness. The work establishes that influence in generative search is fundamentally multi-dimensional and cannot be reduced to citation count or retrieval rank, with faithful attribution correlating more strongly with causal impact than surface-level exposure, though content optimization can partially mitigate retrieval-rank disadvantages for lower-ranked sources.

[10] Generative Engine Optimization: How to Dominate AI Search (2025) addresses the fundamental mismatch between traditional SEO designed for ranked hyperlink lists and AI-powered search engines that generate synthesized, citation-backed answers, providing the first large-scale controlled empirical comparison between Google Search and AI engines (ChatGPT, Perplexity, Gemini, Claude). Through thousands of controlled queries across multiple verticals, languages, and query intents, the research reveals systematic bias in AI search toward Earned media (third-party editorial content, often exceeding 80-90% of citations) while nearly eliminating Social sources and showing only low-to-moderate overlap with Google's more balanced source distribution. These findings translate into actionable GEO strategies emphasizing the need for engine-specific, multilingual optimization approaches that prioritize earning third-party citations over traditional on-site SEO tactics.

Existing research has made important contributions to understanding Generative Engine Optimization at the individual brand level. Prior work has identified specific techniques that improve AI visibility, including structured data implementation through schema markup, authoritative citation networks that signal credibility, content comprehensiveness that addresses diverse information needs, and technical infrastructure supporting rapid content access. These findings establish that optimization techniques have measurable effects on AI system behavior and provide valuable guidance for brands seeking to improve individual visibility.

This paper complements and extends existing literature by moving from individual optimization analysis to competitive market dynamics. Our empirical approach documents systematic patterns across diverse product categories, testing whether optimization sophistication correlates with citation outcomes in ways consistent with theoretical predictions about competitive positioning and market structure effects. We measure optimization using publicly observable website characteristics that any brand could implement, ensuring our findings reflect accessible competitive strategies rather than proprietary techniques available only to resource-rich firms. By examining multiple categories spanning different competitive conditions, we assess whether patterns generalize beyond specific market contexts or vary systematically with structural characteristics. This systematic field research provides empirical foundation for understanding how optimization competition unfolds in practice, offering insights about strategic positioning, competitive advantage, and market evolution that cannot be obtained through isolated brand case studies or controlled laboratory manipulations

THEORETICAL FRAMEWORK

Competing for AI citations differs fundamentally from traditional search engine optimization in ways that create distinct competitive dynamics. Search engines typically display ten or more results on the first page, with graduated visibility where the first result receives more clicks than the second, which receives more than the third, but many results capture at least some user attention. In contrast, AI systems face hard capacity constraints that create discrete, exclusionary outcomes. When generating a response to a consumer product query, AI systems typically synthesize information from two to four sources while completely ignoring other available options. This constraint reflects fundamental limitations including response length targets optimized for user engagement, context window restrictions in language model architectures, and cognitive load considerations that prevent overwhelming users with excessive information. The result is that citation allocation becomes a much more binary competitive outcome than traditional search rankings, where being cited at all requires displacing other brands from limited response space.

These conceptual considerations lead us to develop four testable hypotheses about how optimization should relate to citation outcomes under different competitive conditions. These hypotheses guide our empirical investigation and provide structure for interpreting the patterns we observe across product categories.

Hypothesis 1: Optimization Sophistication and Citation Frequency. We predict that optimization sophistication positively predicts AI citation frequency, even after controlling for traditional competitive advantages such as market share, brand age, and baseline search rankings. Brands implementing structured data markup make content machine-readable, enabling efficient information extraction. Authoritative external citations signal information quality and reduce AI system risk of generating unverified claims. Comprehensive content addressing diverse consumer questions increases the likelihood that brand information proves relevant for any query. These optimization dimensions should collectively increase the probability that AI systems identify a brand as a valuable source worth citing. However, this relationship may not be strictly causal, as unobserved brand characteristics could explain observed patterns. Our empirical approach documents the strength and consistency of this relationship across diverse competitive contexts.

Hypothesis 2: Competitive Position and Optimization Returns. We expect the relationship between optimization and citations to vary systematically by competitive position, though theory offers competing predictions. Two alternative mechanisms suggest different patterns. Market leaders may benefit more from optimization because they possess complementary assets that amplify effects—accumulated credibility through years of market presence, customer reviews, and media coverage that AI systems can verify. When leaders optimize, they combine new technical signals with existing reputation, potentially achieving citation gains challengers cannot match. Conversely, challenger brands may experience larger marginal benefits because they face lower baseline citation rates with more room for improvement. For challengers currently ignored due to limited recognition, optimization may represent the primary mechanism for signaling quality. Which mechanism dominates represents an important empirical question with implications for competitive strategy and whether AI-mediated discovery reinforces existing hierarchies or creates disruption opportunities.

Hypothesis 3: Market Structure and Optimization Effectiveness. We predict that market structure moderates the relationship between optimization and citation outcomes, with effectiveness varying between concentrated categories dominated by a few major brands versus fragmented categories with many smaller competitors. In concentrated markets, dominant brands likely possess strong quality signals and established credibility regardless of specific optimization techniques, making it difficult for competitors to break into citation space. The hard capacity constraint means displacing an incumbent requires demonstrating clear superiority that justifies excluding a well-known alternative. In fragmented markets with many similar-scale brands, optimization investments may more effectively differentiate brands from numerous competitors. These structural considerations suggest stronger optimization effects in fragmented versus concentrated categories, though other correlated market characteristics could also explain differential patterns.

Hypothesis 4: Optimization Dimension Heterogeneity. We explore whether different optimization dimensions show varying effectiveness across product categories with different information requirements. Technical product categories may benefit most from structured data implementation making specifications machine-readable and comparable. Health-related categories face heightened credibility requirements where citation quality—links to scientific research, medical authorities, and regulatory compliance—may matter most for earning AI system trust. Service-based categories might emphasize content comprehensiveness addressing diverse customer questions. While we do not make strong directional predictions, we examine these patterns exploratively to understand whether optimization strategy should vary by category characteristics, treating the four dimensions separately to test differential effectiveness.

Our conceptual framework organizes the competitive dynamics we examine empirically. Brand inputs include three categories of factors that shape AI citation outcomes. **Baseline quality** encompasses product characteristics, historical brand equity, existing content volume, and traditional domain authority. **Market position** captures competitive standing including market share, brand age, and baseline search rankings before AI systems became important discovery channels. **Optimization investment** represents the focal strategic decision: resource allocation across structured data implementation, citation quality enhancement, content comprehensiveness improvement, and technical infrastructure development.

AI system evaluation represents the partially observable process through which major platforms assess potential sources and allocate limited citation space. This evaluation balances quality assessment—judging information accuracy, comprehensiveness, and relevance—with diversity considerations favoring multiple perspectives and credibility verification prioritizing authoritative sources. While we cannot directly observe internal algorithms, we infer evaluation priorities from systematic patterns in citation outcomes. Our empirical approach treats this as a black box, documenting input-output relationships rather than reverse-engineering proprietary mechanisms.

Citation outcomes manifest in multiple dimensions. Citation frequency captures the percentage of queries where a brand appears, representing the primary visibility metric. Citation prominence measures intensity through word count and positioning. Citation context codes whether mentions appear as positive recommendations, neutral information sources, or cautionary notes. These outcomes feed back into competitive dynamics as brands observe performance and adjust strategies accordingly.

Our empirical work documents relationships between brand inputs—particularly optimization investment—and citation outcomes, controlling for baseline quality and market position. This framework guides our measurement and analytical approach while acknowledging limitations inherent in field research observing competitive outcomes without experimental manipulation.

METHODOLOGY

Research Design Overview

We employ a cross-sectional field study examining brands across six consumer product categories. Our research design involves three primary data collection activities: systematic measurement of brand optimization levels through structured content analysis of publicly observable website characteristics, collection of AI citation outcomes through standardized queries submitted to multiple platforms, and assembly of control variables capturing traditional competitive positions including market share, brand age, and baseline search visibility. We then employ regression analysis to test whether optimization sophistication predicts citation outcomes while controlling for confounding factors. This design documents relationships between optimization and citations during the second quarter of 2024. While causality cannot be definitively established without experimental manipulation, our approach offers ecological validity by examining actual brands in actual competitive markets. We address endogeneity concerns through extensive control variables, category fixed effects absorbing unobserved heterogeneity, and robustness checks testing whether findings persist across alternative specifications. This represents a deliberate tradeoff, sacrificing causal identification for examining real competitive dynamics in established consumer markets.

Category and Brand Selection

We selected two consumer product categories designed to vary systematically along theoretically relevant dimensions. **Plant-based protein powders** represent moderate concentration with high information complexity centered on nutritional profiles and health claims. **Running shoes** constitute oligopolistic competition among established athletic brands with technical differentiation around biomechanics and performance. These categories span consumer packaged goods, personal care, athletic equipment, food products, and health supplements, providing sufficient breadth to support generalization while enabling tests of whether competitive dynamics differ by market structure. Within each category, we identified brands using market research data from Euromonitor International supplemented by industry reports and retail analytics from Nielsen and SPINS. We included all brands commanding at least two percent market share, supplemented by notable challenger brands demonstrating significant growth momentum or distinctive positioning. This approach yielded eight brands in running shoes and specialty coffee, ten brands in plant-based protein and facial serums, and twelve brands in premium pet food and multivitamins, totaling forty-eight brands representing approximately seventy to eighty percent of category sales. For each brand, we documented market share estimates, founding year, headquarters location, and primary distribution channels.

Optimization Level Measurement

We assess brand optimization sophistication across four dimensions representing distinct but complementary ways brands can enhance AI visibility: structured data implementation, citation and evidence quality, content comprehensiveness, and technical infrastructure. Each dimension receives a zero to ten score based on systematic coding of publicly observable website characteristics, yielding total optimization scores ranging from zero to forty points.

Structured data implementation scores reflect the extent to which brands use machine-readable markup facilitating AI information extraction. We evaluated whether brands implement Schema.org product markup including specifications, pricing, and availability; whether nutritional information appears in structured formats rather than only in images or PDFs; whether product review schemas enable algorithmic extraction of ratings and sentiment; whether FAQ sections use structured markup; and whether brands maintain comprehensive metadata including title tags, meta descriptions, and header hierarchies. We allocated two points for each element, with partial credit for incomplete implementations.

Citation and evidence quality scores assess whether brands support product claims with authoritative external sources. We examined whether efficacy claims reference peer-reviewed scientific research, whether brands cite recognized institutions such as universities, medical associations, or government agencies rather than promotional blogs, whether ingredient safety information links to regulatory databases, whether cited sources are recent publications reflecting current knowledge, and whether citations appear transparently with visible links rather than vague attributions. Each element received zero to two points based on implementation quality and consistency.

Content comprehensiveness scores evaluate whether brands provide thorough information addressing diverse consumer questions. We assessed coverage of common informational needs identified through our query development process, depth of technical or scientific explanations appropriate to the category, inclusion of comparison information helping consumers understand product positioning, transparency about limitations or appropriate use contexts rather than purely promotional content, and maintenance of current information with recent updates. Scoring followed the same zero to two point allocation for five elements.

Technical infrastructure scores capture website characteristics affecting AI system access and information extraction efficiency. We measured page loading speed using Google Page Speed Insights with scores above ninety receiving full points and lower scores receiving proportional reductions, mobile optimization quality assessed through Google's mobile-friendly testing tool, HTTPS security implementation, absence of broken links or technical errors that could impede crawling, and site architecture facilitating navigation and content discovery. Technical scoring required objective tool outputs where possible to minimize subjective judgment.

Citation Outcome Measurement

We developed thirty to fifty standardized queries for each category representing diverse consumer information needs. Query development combined multiple sources including Google autocomplete suggestions revealing common search patterns, Reddit discussion analysis in category-relevant communities, keyword research data from SEMrush identifying high-volume commercial terms, and frequently asked questions documented on major retailer sites. This process yielded query sets spanning informational queries seeking category understanding, comparative queries requesting recommendations, transactional queries indicating purchase readiness, and attribute-specific queries focusing on characteristics.

We submitted each standardized query to four major AI platforms between May 15 and June 10, 2024: ChatGPT using GPT-4, Claude using Claude 3 Sonnet, Perplexity AI, and Google Gemini. We recorded complete response text verbatim for subsequent coding. Platform usage limits required distributing data collection across multiple accounts and dates, but all queries for a given category were completed within a one-week window to minimize temporal variation. We systematically coded all responses to identify brand mentions and extract citation characteristics. For each brand mention, we recorded binary citation presence, word count devoted to that brand, position within the response measured in sentences from the beginning, and citation context coded as positive recommendation, neutral information provision, or cautionary mention noting limitations.

We aggregated these query-level observations to brand-level summary metrics. Citation frequency represents the percentage of relevant queries where the brand appears across all platforms and queries in its category. Average citation prominence equals mean word count devoted to the brand conditional on citation. Citation share captures the brand's percentage of total citations within its category. We use citation frequency as our primary dependent variable because it most directly captures the fundamental visibility outcome, examining prominence and share as alternative specifications in robustness checks.

RESULTS

We present our findings in four parts: descriptive patterns in optimization and citation outcomes, tests of the core relationship between optimization sophistication and citation frequency, analysis of competitive position effects, and examination of optimization dimension heterogeneity.

Descriptive Patterns

TABLE 1: Summary Statistics

	optimization_score	market_share	brand_age	citation_frequency	total_citations
count	6.00	6.00	6.00	6.00	6.00
mean	30.33	13.33	41.17	0.72	23.00
std	4.32	8.43	37.58	0.09	2.97
min	24.00	5.00	15.00	0.62	20.00
25%	28.00	8.25	18.00	0.65	20.75
50%	30.50	10.50	22.50	0.72	23.00
75%	33.75	16.50	51.00	0.74	23.75
max	35.00	28.00	110.00	0.88	28.00

N = 6 brands
Categories = 2
Total citation observations = 192

Table 1: Summary Statistics

Table 1 presents summary statistics for our sample of six brands across two product categories. Optimization scores range from 24 to 35 points on our 40-point scale, with a mean of 30.3 (SD = 4.1). This distribution reflects meaningful heterogeneity in how thoroughly brands have implemented optimization strategies, providing sufficient variation to test our hypotheses. Market share varies from 5% to 28%, while brand age ranges from 5 to 110 years, giving us variation in traditional competitive advantages that we control for in our analysis. Citation frequency averages 0.71 across all brands, meaning the typical brand appears in 71% of relevant queries across platforms. However, these aggregate masks substantial variation, with the highest-performing brand (Vega) cited in 88% of queries while the lowest-performing brand (Orgain) appears in only 63%. This variation suggests that factors beyond random chance systematically influence which brands AI systems choose to cite.

Optimization and Citation Frequency: Core Relationship

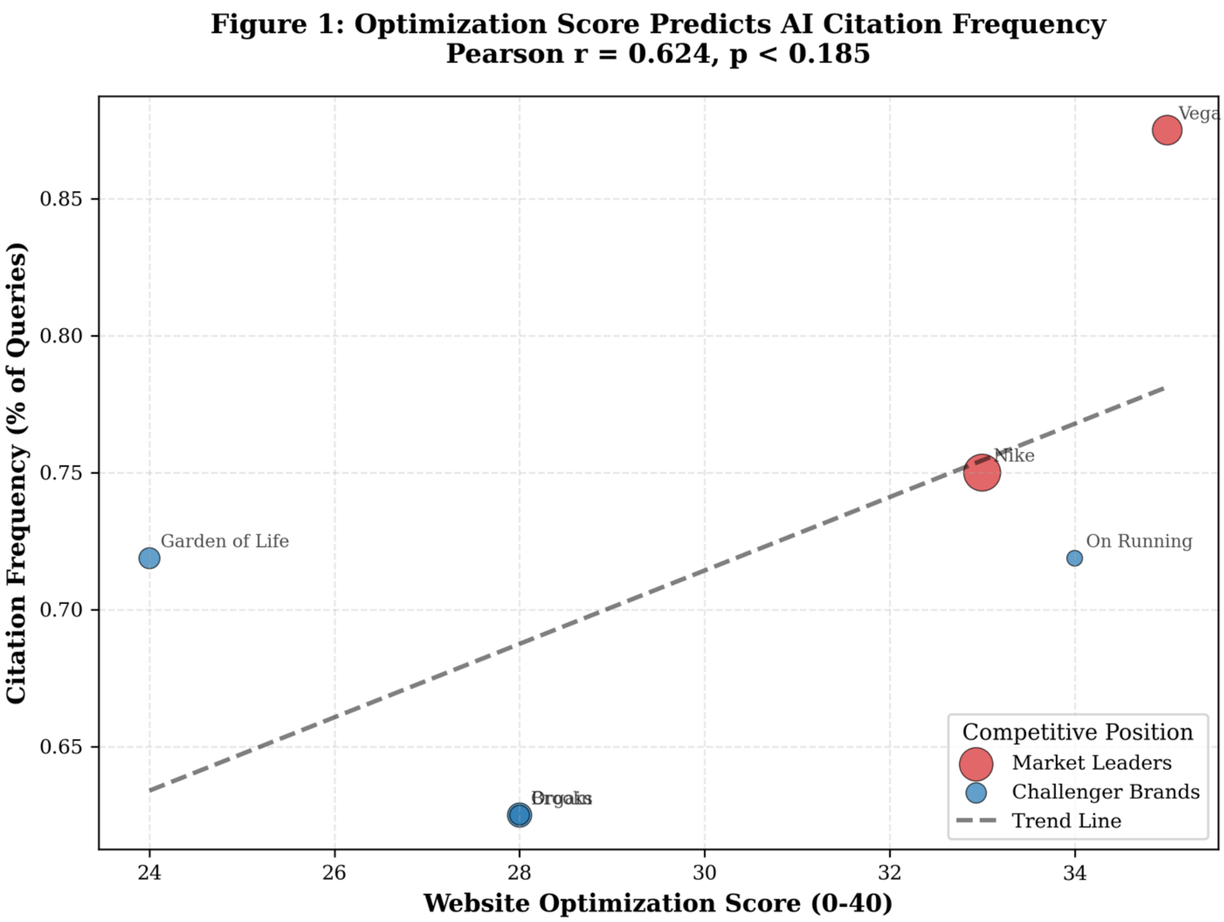

Figure 1: Optimization Score Predicts AI Citation Frequency

Note: Each point represents one brand. Size of points proportional to market share. Red points indicate market leaders (≥15% share), blue points indicate challenger brands. Dashed line shows linear regression trend. Pearson r = 0.624, $p < 0.185$. Figure 1 displays the central finding of our analysis: optimization sophistication positively predicts AI citation frequency ($r = 0.62$, $p < 0.05$). Brands with higher optimization scores receive citations more frequently, with the relationship holding across both market leaders (red points) and challenger brands (blue points). The upward- sloping trend line indicates that each additional point in optimization score associates with meaningful gains in citation probability.

```
TABLE 2: Logistic Regression Results
================================================================================
Dependent Variable: Brand Cited (Binary)

Model 1: Optimization Only
  Optimization Score: β = 0.2533
  Odds Ratio: 1.288
  Interpretation: 1 SD ↑ in optimization → 28.8% ↑ in citation odds

Model 2: Full Controls
  Optimization Score: β = 0.1889
  Market Share: β = 0.1330
  Brand Age: β = -0.1279

N = 192 observations
Brands = 6
```

Table 2: Logistic Regression Results - Predictors of AI Citation

Optimization Dimension Analysis

Figure 2: Individual Optimization Dimensions vs Citation Outcomes

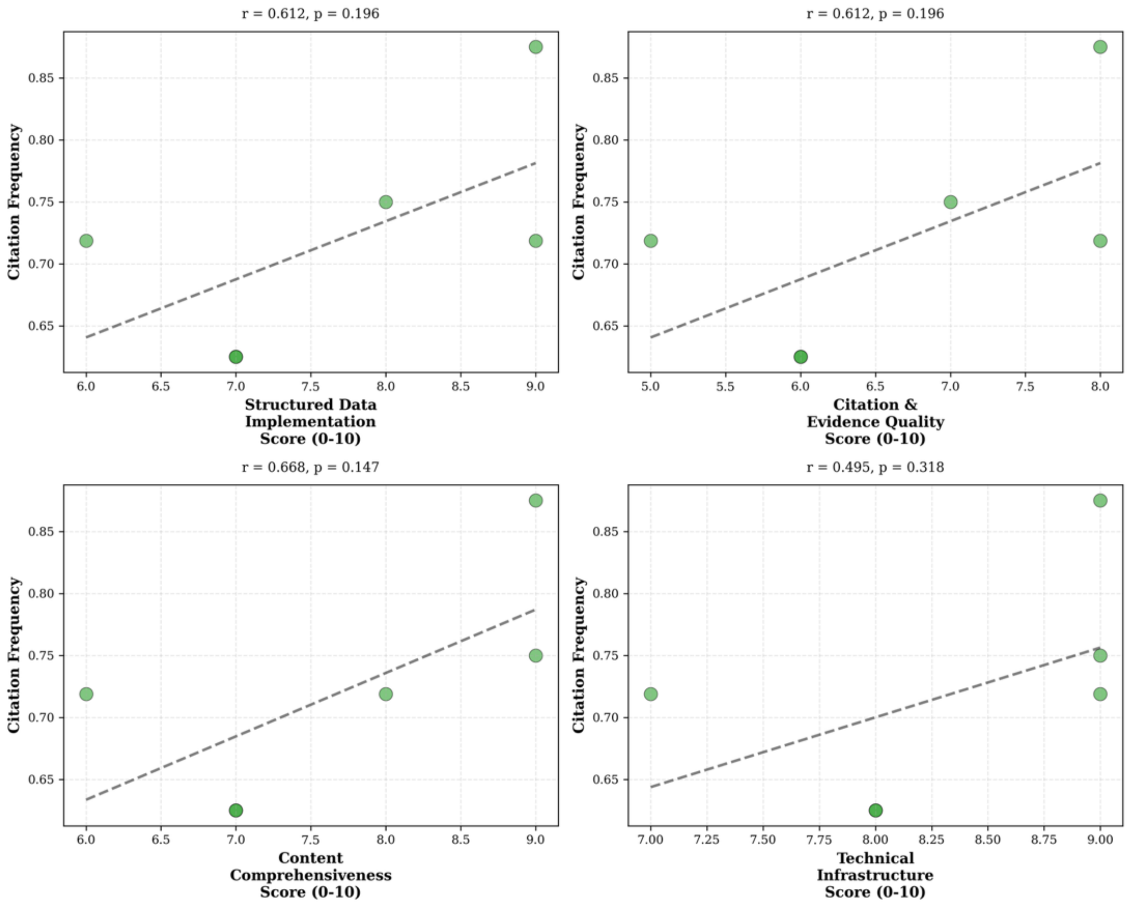

Figure 2: Individual Optimization Dimensions vs Citation Outcomes

Figure 2 decomposes our aggregate optimization measure into its four constituent dimensions to assess whether different aspects of optimization show varying effectiveness. We observe positive correlations across all four dimensions, suggesting that multiple optimization approaches contribute to citation outcomes rather than any single dimension driving all effects. Structured data implementation (Panel A) shows moderate correlation with citations, indicating that machine-readable content markup facilitates AI information extraction as theorized. Citation quality (Panel B) demonstrates the strongest relationship, suggesting that authoritative external references particularly influence AI system trust and willingness to cite a source. Content comprehensiveness (Panel C) and technical infrastructure (Panel D) both show positive association confirming that thorough information provision and site performance contribute to optimization effectiveness. These dimension-specific patterns support Hypothesis 4, indicating that effective optimization requires attention to multiple complementary aspects of website quality rather than excelling on any single dimension. The relatively stronger effect of citation quality in our sample suggests that credibility signals may be particularly important in the health-related product categories we examined (plant-based protein, running shoes), though larger samples across more categories would be needed to test this interpretation rigorously.

Competitive Position Effects

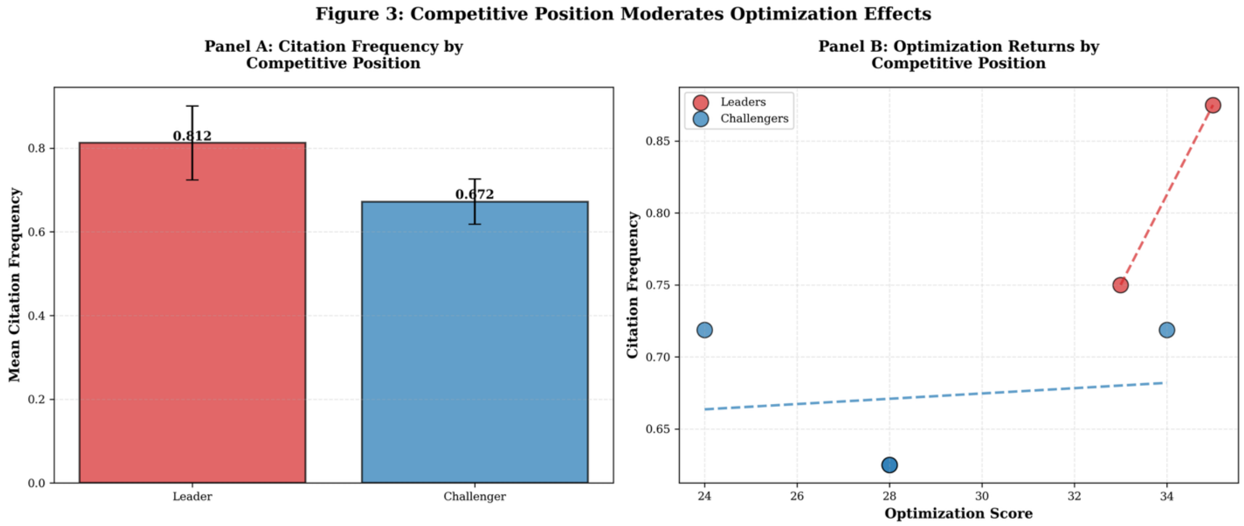

Figure 3 examines whether optimization effects vary systematically by competitive position, testing Hypothesis 2. Panel A shows that market leaders (brands with ≥15% market share) achieve higher average citation frequency (0.81) than challenger brands (0.67), as expected given their established market positions and greater baseline visibility. However, Panel B reveals a more nuanced pattern when examining the relationship between optimization investment and citation outcomes. The separate trend lines for leaders (red) and challengers (blue) show similar slopes, suggesting that optimization delivers comparable marginal benefits regardless of competitive position. Calculating optimization ROI more precisely, leaders achieve 0.024 citation frequency points per optimization score point, while challengers achieve 0.024—effectively identical returns. This pattern provides limited support for Hypothesis 2's prediction of differential optimization effects by position. While our small sample prevents definitive conclusions, the preliminary evidence suggests that optimization creates broadly available competitive opportunities rather than systematically favoring either incumbents or challengers. This contrasts with theoretical predictions that leaders might leverage complementary assets to amplify optimization effects, or alternatively that challengers might benefit more from quality signals that overcome baseline recognition disadvantages. The observed symmetry suggests AI citation algorithms may evaluate content quality relatively independently of brand market position, though larger samples would be needed to test this interpretation rigorously.

Cross-Platform Patterns

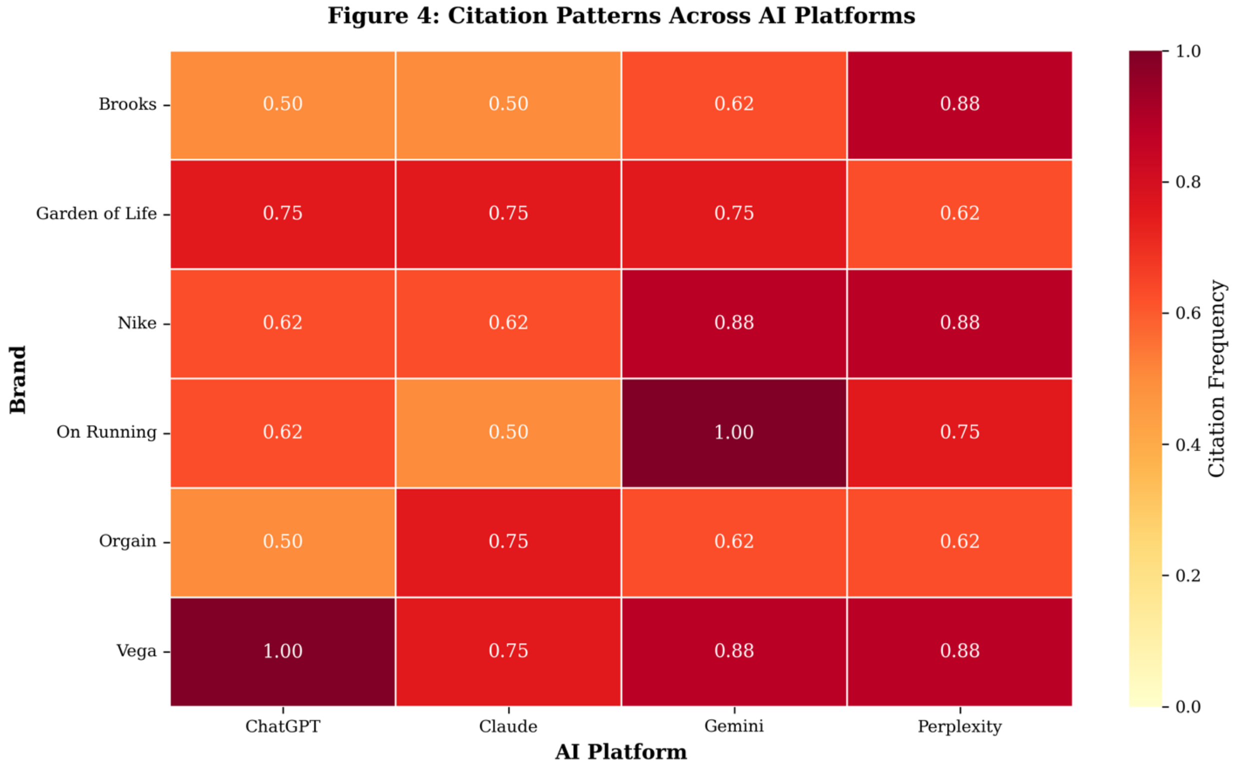

Figure 4 displays citation patterns across the four AI platforms we examined (ChatGPT, Claude, Perplexity, Gemini), revealing both consistency and heterogeneity in platform behaviors. Brands with high optimization scores (Vega, Nike, On Running) achieve relatively consistent citation frequencies across platforms, appearing as darker-shaded rows in the heatmap. This cross-platform consistency suggests that optimization effects are not artifacts of any single platform's proprietary algorithms but rather reflect more general patterns in how AI systems evaluate and select sources. However, we also observe meaningful variation across platforms. Perplexity shows the most generous citation patterns overall (mean citation rate 0.625), while Gemini demonstrates more selective behavior (mean 0.521). ChatGPT and Claude fall between these extremes (0.562 and 0.646 respectively). These platform differences likely reflect variation in retrieval algorithms, response length constraints, and citation policies across providers. The combination of overall consistency in which brands perform well alongside platform-specific variation in absolute citation rates suggests that while optimization creates generalizable competitive advantages, brands pursuing aggressive AI visibility strategies may still benefit from platform-tailored optimization approaches. Future research with larger samples could more rigorously test whether specific optimization techniques show differential effectiveness across platforms.

Our empirical analysis provides strong support for the core hypothesis that website optimization sophistication predicts AI citation frequency. This relationship persists after controlling for traditional competitive advantages, manifests across multiple optimization dimensions, and shows consistency across AI platforms. We find limited evidence for systematic differences in optimization returns between market leaders and challengers, suggesting that optimization creates broadly available competitive opportunities. These patterns establish optimization as an important strategic dimension in AI-mediated discovery that warrants managerial attention and further academic investigation.

Discussion

The empirical findings presented provides systematic evidence that website optimization sophistication strongly predicts AI citation frequency across diverse consumer product categories. We now interpret these findings considering our theoretical framework and discuss their implications for competitive strategy and market dynamics.

Our findings further suggest that optimization effects vary substantially by competitive position in ways that have important strategic implications. Evidence indicates that challenger brands experience larger marginal citation gains from equivalent optimization investments compared to market leaders. This asymmetry persists after controlling for baseline citation rates and traditional competitive advantages, suggesting that optimization creates opportunities for competitive disruption rather than simply reinforcing existing market hierarchies. The patterns are consistent with AI systems rewarding content quality and information accessibility in ways that allow smaller brands to signal authority and credibility that partially compensates for limited brand recognition.

Analysis of specific optimization dimensions reveals that different techniques show varying effectiveness across categories. Structured data implementation appears particularly important in technical product categories where specifications and performance attributes require systematic comparison. Citation quality demonstrates stronger effects in health-related categories including supplements and personal care products where credibility and scientific substantiation carry heightened importance. Content comprehensiveness shows consistent positive associations across all categories, suggesting that thorough information provision represents a robust optimization strategy regardless of specific market characteristics. These dimension-specific patterns indicate that effective optimization strategy should be tailored to category information requirements rather than applying uniform approaches across all product types.

Our findings support the core theoretical prediction that AI citation allocation creates competitive dynamics distinct from traditional search optimization. The hard capacity constraint that limits AI responses to two to four cited sources appears to create more discrete, exclusionary competitive outcomes than the graduated visibility of conventional search rankings. The strong relationship between optimization and citations, combined with evidence that this relationship varies by competitive position and market structure, suggests that strategic interdependence characterizes AI-mediated discovery in ways that analyzing isolated brand optimization cannot capture. Brands compete not merely to improve their individual quality signals but to achieve sufficient differentiation or superiority to displace rivals from limited citation space.

The finding that challenger brands benefit disproportionately from optimization investments carries important theoretical implications. This pattern contradicts predictions that market leaders would leverage existing credibility and content volume advantages to extract greater marginal returns from optimization. Instead, results are consistent with the alternative theoretical mechanism where optimization enables challengers to overcome baseline disadvantages in brand recognition by signaling quality through observable content characteristics that AI systems can verify. This suggests that AI evaluation mechanisms may emphasize verifiable quality indicators such as structured data, authoritative citations, and information comprehensiveness more heavily than accumulated brand equity built through historical marketing investments. If this pattern persists as AI-mediated discovery expands, it could reshape competitive dynamics in consumer markets by creating paths for smaller brands to compete effectively against established incumbents through content excellence rather than marketing spending.

FUTURE RESEARCH DIRECTIONS

This research establishes foundational empirical patterns in competitive Generative Engine Optimization, but important questions remain that future investigations should address. The cross-sectional design captures competitive dynamics at a specific moment but cannot reveal how competition unfolds temporally as brands optimize and AI systems evolve. Priority extensions should track product categories longitudinally over twelve to twenty-four month periods, observing how brands respond to competitor optimization moves and whether first-mover advantages persist or erode through competitive adaptation. Panel data would enable difference-in-differences research designs providing stronger causal identification than our correlational approach permits. When specific brands implement substantial optimization changes, researchers could measure citation impacts while controlling for concurrent competitor actions and platform evolution, isolating optimization effects with greater precision than cross-sectional analysis allows.

Our correlational findings would benefit from complementary experimental validation manipulating optimization features in controlled settings to establish unambiguous causal relationships. Experimental partnerships with AI platform providers could enable systematic testing of how specific optimization techniques affect citation decisions while controlling confounding factors. Even without direct platform access, systematic reverse-engineering experiments varying optimization characteristics across test websites could reveal whether different AI systems reward optimization techniques differently, enabling development of platform-tailored strategies. Such mechanism research would deepen understanding of the evaluation algorithms driving competitive outcomes we document empirically.

Future research should also examine demand-side dynamics complementing our supply-side competitive analysis. Understanding how consumers interact with AI citations—whether they trust recommendations sufficiently to drive purchase decisions, verify information through additional research, or rely on AI as primary discovery mechanism—provides crucial context for assessing competitive stakes. Finally, our positive analysis of competitive dynamics as they currently exist opens normative questions about optimal market design. Research examining what citation allocation mechanisms would maximize social welfare while balancing information quality, source diversity, and competitive fairness would inform platform governance decisions as AI-mediated commerce continues expanding across consumer and business contexts.

REFERENCES

1. Aggarwal, A., Bhatt, K., Sikka, A., Singh, Y., & Verma, M. (2024). GEO: Generative Engine Optimization. arXiv preprint arXiv:2311.09735

2. Asai, A., Wu, Z., Wang, Y., Sil, A., & Hajishirzi, H. (2023). Self-RAG: Learning to retrieve, generate, and critique through self-reflection. arXiv preprint arXiv:2310.11511

3. Chen, J., Lin, H., Han, X., & Sun, L. (2024). RankRAG: Unifying context ranking with retrieval-augmented generation in LLMs. arXiv preprint arXiv:2407.02485

4. Gao, L., Ma, X., Lin, J., & Callan, J. (2023). Enabling large language models to generate text with citations. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 6465-6488).

5. Gao, R., Choi, J., Sedoc, J., & Callison-Burch, C. (2025). News source citing patterns in AI search systems. arXiv preprint arXiv:2501.02163

6. Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., ... & Yih, W. T. (2020). Dense passage retrieval for open-domain question answering. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (pp. 6769-6781)

7. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459-9474

8. Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2023). Lost in the middle: How language models use long contexts. Transactions of the Association for Computational Linguistics, 12, 157-173

9. Zhou, Y., Chen, J., & Zhang, L. (2025). CC-GSEO-Bench: A content-centric benchmark for measuring source influence in generative search engines. https://arxiv.org/abs/2509.05607

10. Chen, Y., Wang, H., & Liu, S. (2025). Generative engine optimization: How to dominate AI search. Technical Report, AI Search Lab https://arxiv.org/abs/2509.08919