

How GPT Results Differ from Google Search: LLM-SERP Overlap Study

Large language models (LLMs) are reshaping how information is retrieved and summarized. GPT systems generate...

Large language models (LLMs) are reshaping how information is retrieved and summarized. GPT systems generate direct, conversational answers to queries, while Google Search ranks pages through authority, link structure, and topical relevance. Both deliver information, but it remains unclear whether they rely on the same sources.

SEO professionals and researchers debate whether GPT responses reflect the same informational ecosystem as Google or whether these models construct a new layer of web interpretation. The missing piece is large-scale evidence showing how closely LLM-generated citations align with search engine results.

The study analyzes 18,377 semantically matched query pairs between GPT-generated responses and Google Search Engine Results Pages (SERP). The datasets span September to October 2025, comparing URLs and domains referenced by both systems using an 82% cosine similarity threshold to identify equivalent queries.

The findings reveal that GPT results diverge from Google Search. Domain-level overlap shows partial topical alignment, while URL-level overlap remains low, confirming that GPT responses depend primarily on synthesized understanding rather than direct web retrieval.

Methodology – How Was LLM–SERP Alignment Measured?

This experiment measures how large language models (LLMs) overlap with Google Search results at both the domain and URL level. The analysis evaluates whether model-cited sources reflect the same ecosystem of websites visible in organic rankings.

This experiment matters because it defines how retrieval-augmented and reasoning-based models interact with the indexed web. Understanding alignment between LLMs and search results reveals whether AI systems reflect, filter, or redefine the modern search ecosystem.

The dataset integrates **2 primary components** listed below.

- **LLM Query Dataset.** Responses from OpenAI (GPT), Perplexity, and Gemini collected in October 2025. Each record includes the query title, platform name, timestamps, and cited URLs and domains.
- **SERP Dataset.** Keyword-level search results collected in September and October 2025. Each record contains the keyword and its associated search results in JSON format.

Each SERP record was parsed to extract **URLs** and **domains**. Domains were derived using URL parsing, and timestamps were standardized to retain only those from the September to October 2025 window. This process produced a structured dataset containing keywords, URLs, and domains prepared for embedding analysis.

The analytical steps are listed below.

1. Compute Domain Overlap (%) = $(\text{shared domains} \div \text{total unique LLM domains}) \times 100$.
2. Compute URL Overlap (%) = $(\text{shared URLs} \div \text{total unique LLM URLs}) \times 100$.
3. Aggregate overlap results by model and query intent.
4. Visualize overlap distributions with boxplots and averages with bar charts.
5. Illustrate total intersections of unique domains and URLs using Venn diagrams.

The dataset scope is defined below.

- **Total sample size.** 18,377 semantically matched query pairs.
- **Segmentation.** Models (Perplexity, GPT, Gemini) and query intent (Informational, Navigational, Transactional, Evaluation, Understanding).
- **Period.** September to October 2025.

The target variables are listed below.

- **Domain Overlap (%).** Measures general topical alignment between model citations and Google Search domains.
- **URL Overlap (%).** Measures exact page-level correspondence between model citations and ranked SERP URLs.

This framework enables both micro-level (per-query) and macro-level (platform) comparisons. The design isolates whether retrieval-driven systems mirror the authoritative sources from Google or whether reasoning-based models generate semantically consistent but citation-divergent responses.

What Is the Final Takeaway?

The analysis demonstrates that retrieval-augmented systems align most closely with the Google search ecosystem. The study shows that overlap between LLM citations and SERP results depends on model design, as retrieval models reproduce indexed sources while reasoning models generate answers from pre-trained contexts.

Perplexity achieves the highest and most consistent alignment, which confirms that live web access drives stronger correspondence with Google ranked domains. OpenAI (GPT) and Gemini display lower overlap, which shows greater reliance on internal reasoning rather than direct citation.

Domain overlap reflects topical consistency, both systems discuss the same subjects, while URL overlap measures factual fidelity and remains low across every model. This pattern reveals that GPT systems mirror what Google knows rather than what Google ranks.

The direction of these findings remains consistent. Retrieval enables alignment, reasoning creates divergence, and overlap defines a new layer of digital visibility. SEO and AI teams need to evaluate search and LLM ecosystems together, using overlap analysis as a benchmark to measure future discoverability across AI-generated environments.

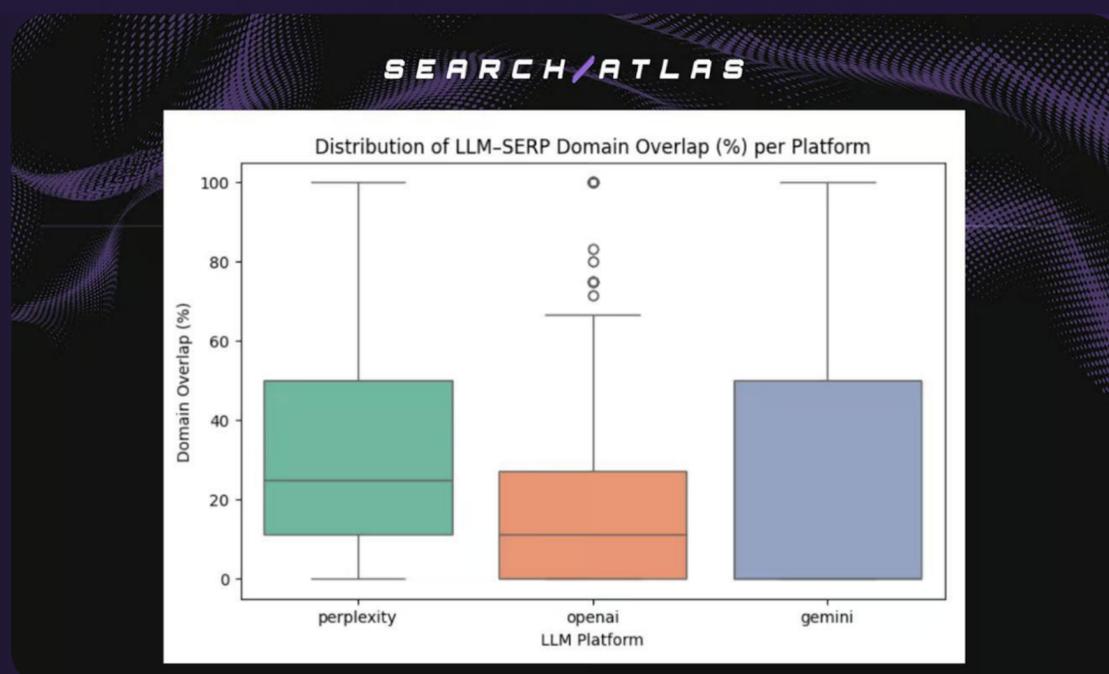
How Do LLMs Differ in Search Alignment?

I, Manick Bhan, together with the Search Atlas research team, analyzed 18,377 semantically matched LLM-SERP keyword pairs to measure how retrieval-augmented and reasoning-based models align with Google's indexed results. The breakdown to show how Perplexity, GPT, and Gemini differ in domain and URL overlap is listed below.

Domain Overlap Analysis

The domain-level analysis measures how often LLM-cited domains appear in Google SERPs. Domain overlap matters because it reveals whether a model references the same authoritative sources that appear in organic search.

The headline results are shown below.



- **Perplexity median overlap.** 25 to 30%
- **OpenAI (GPT) median overlap.** 10 to 15%
- **Gemini overlap.** Variable, ranging from near zero to strong alignment depending on topic.

Perplexity shows the highest and most stable overlap. Its live web retrieval enables citation of the same authority domains visible in Google results. OpenAI (GPT) shows lower overlap, reflecting its reliance on internal training rather than live search. Gemini displays a mixed pattern, with selective retrieval that varies by context.

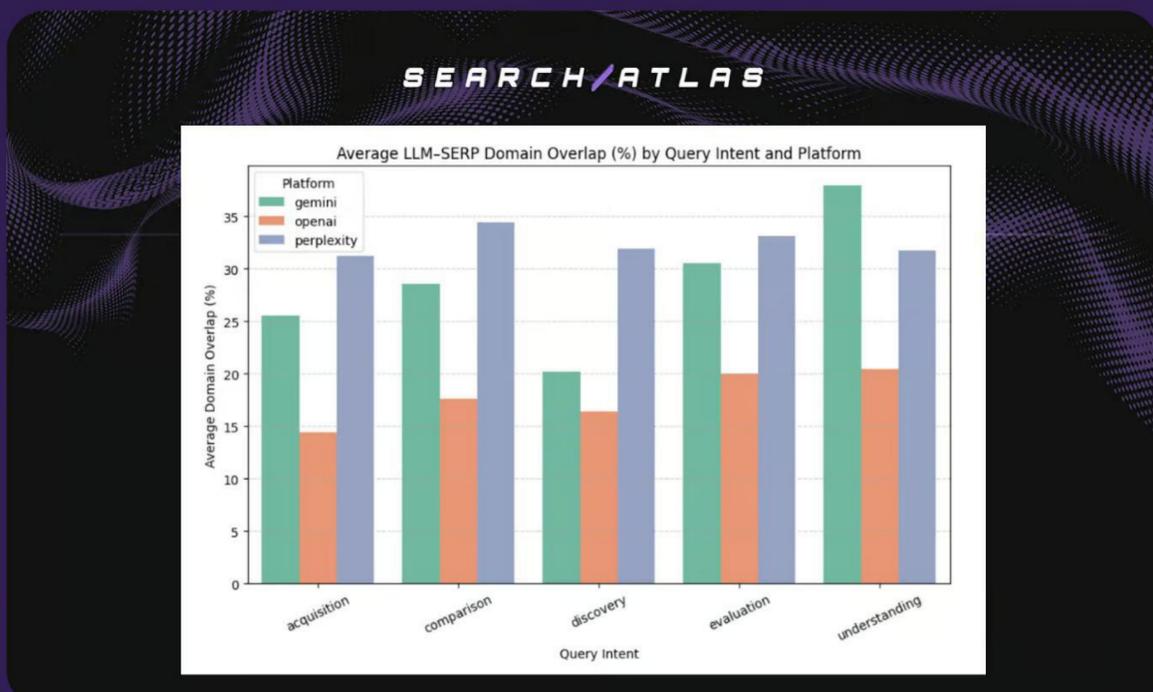
Average Domain Overlap by Query Intent

Average domain overlap across 5 intent categories reveals how model design influences search consistency. The examples of queries for each intent category are shown below.

SEARCH/ATLAS

intent	title
acquisition	how to choose a criminal defense attorney
acquisition	commercial solar financing options Texas
comparison	compare motorcycle accident lawyers in Illinois
comparison	comparison of social media management tools for businesses
discovery	Preferred Remodeling home improvement financing options
discovery	HR outsourcing for small businesses
evaluation	dryer vent cleaning cost Las Vegas
evaluation	Fast Layne Plumbing & HVAC review
understanding	components of an effective roofing maintenance plan
understanding	Wholesale Peptide distributors and partners

The headline results are shown below.

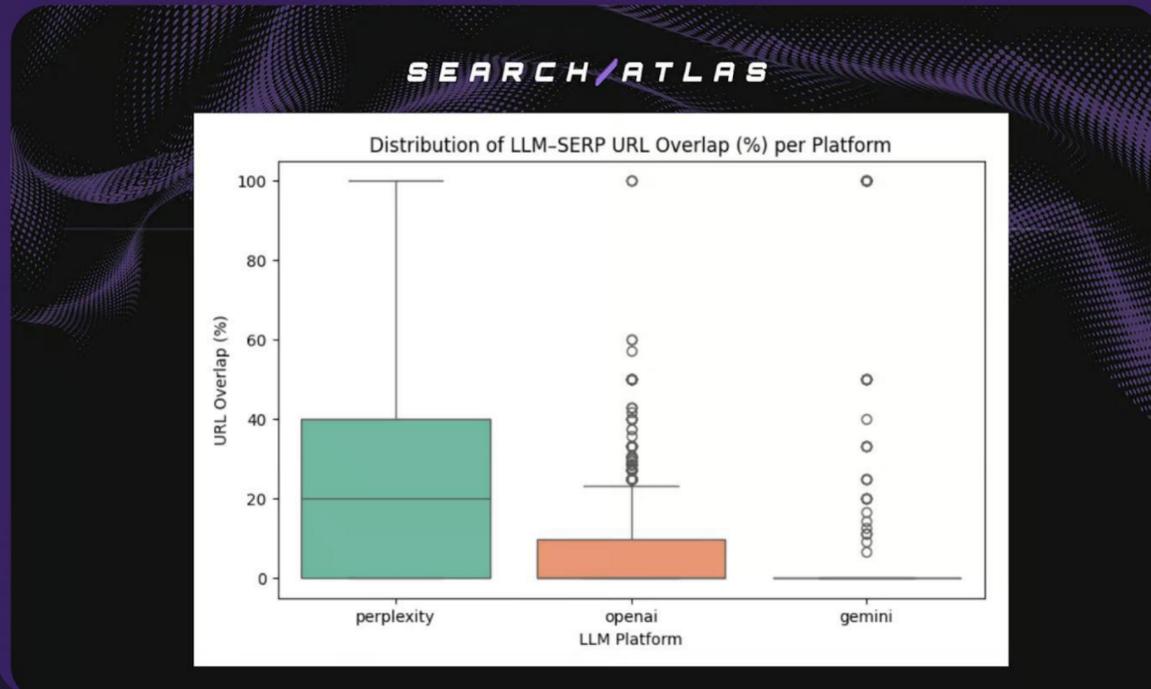


- **Perplexity average overlap.** 30 to 35% across all intents
- **OpenAI (GPT) average overlap.** Below 15% across all intents
- **Gemini average overlap.** Strongest for Understanding queries

Perplexity maintains consistent domain alignment across intents, confirming that live retrieval sustains search-level consistency. Gemini performs best on Understanding queries, where longer explanations increase source diversity. OpenAI remains lowest, as conceptual synthesis replaces direct citation.

URL Overlap

URL overlap measures exact page matches between LLM citations and URLs listed in SERP. URL overlap matters because it represents direct retrieval fidelity rather than general topic alignment. The headline results are shown below.



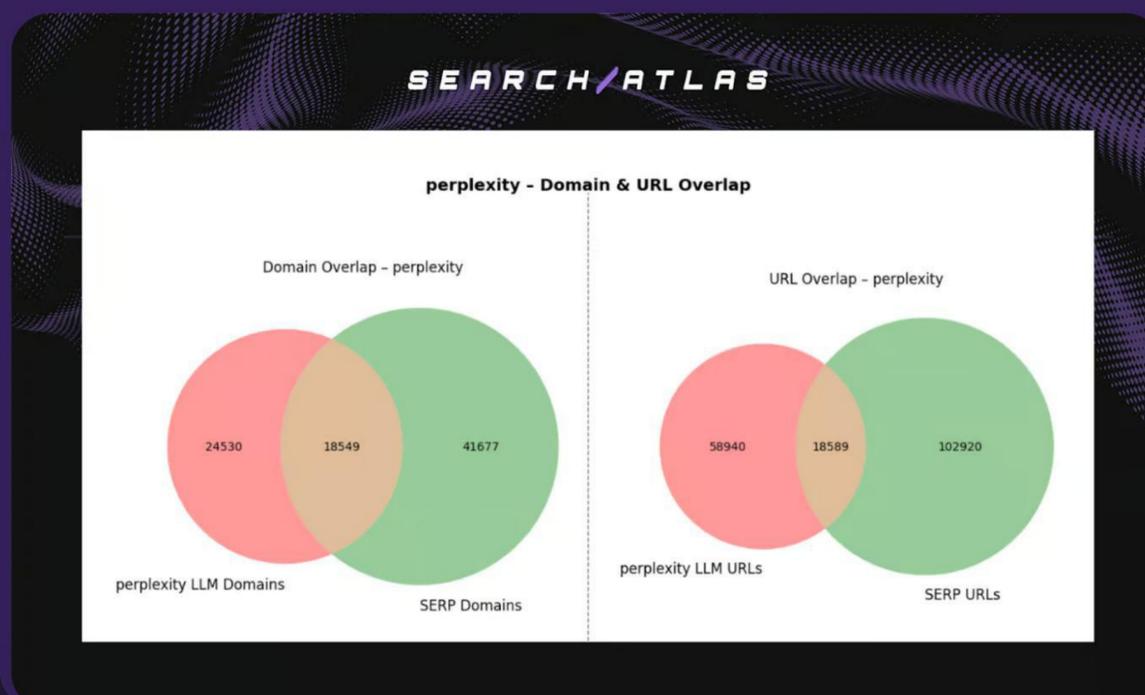
Perplexity again leads in URL alignment, reflecting its continuous access to live search data. OpenAI (GPT) and Gemini show lower overlap because both generate from internal knowledge and rarely reproduce identical URLs.

Which LLM Model Aligns Best with Google Search?

The 3 leading models (Perplexity, OpenAI, and Gemini) were analyzed to determine which system most closely aligns with Google Search results. The comparisons measure domain overlap, URL overlap, and retrieval behavior to reveal how each model references or diverges from the web sources that define organic visibility.

The breakdown to show which model achieves the highest domain and URL overlap is outlined below.

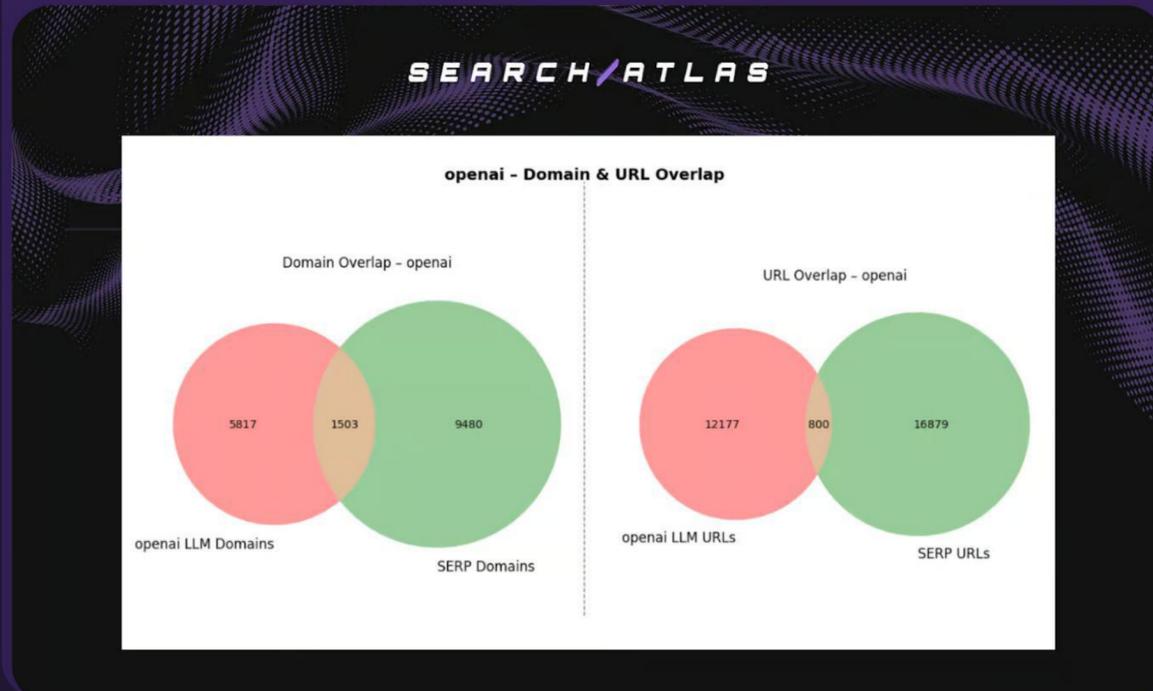
Perplexity: Search-Aligned Retrieval



Perplexity achieves the highest and most consistent overlap with Google Search. It shares **43% of domains** and **24% of URLs** found in SERP results. This confirms that live-web retrieval enables Perplexity to mirror Google's authoritative sources. Its design integrates real-time data, which allows it to reference current articles, businesses, and domains visible in organic results.

Overlap patterns show that Perplexity behaves like a search-continuous model, maintaining strong alignment across all query intents. Its retrieval system acts as a bridge between generative and search-indexed content, reflecting both freshness and topical precision.

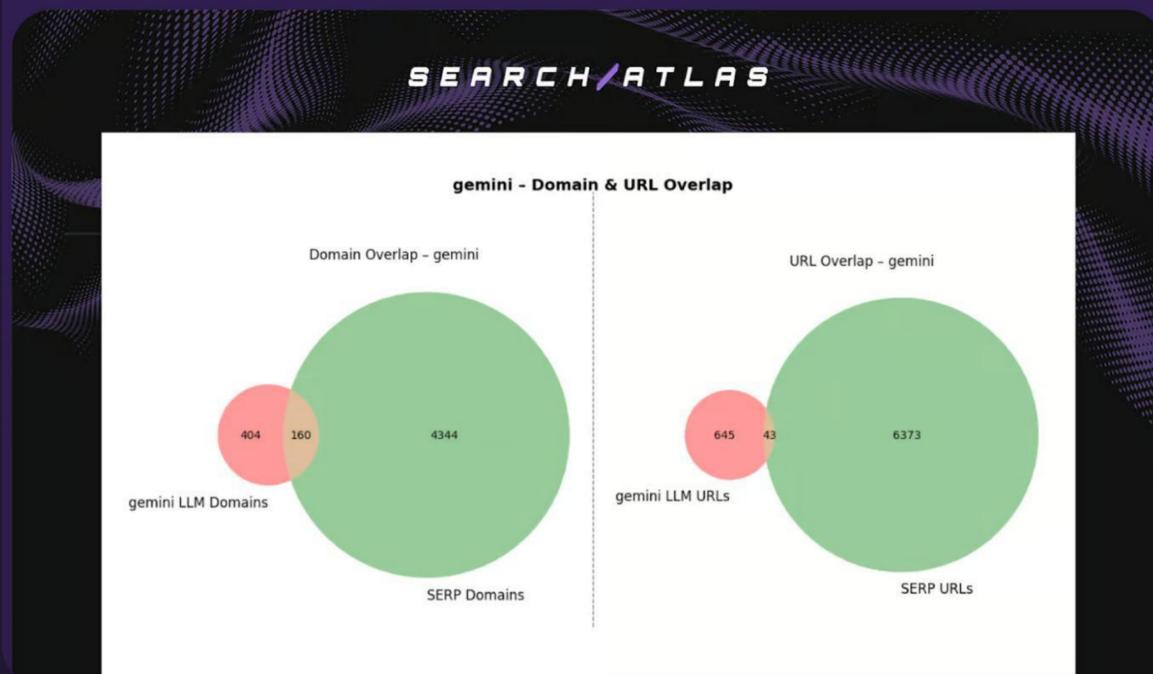
OpenAI (GPT): Reasoning Over Retrieval



GPT shows **21% domain overlap** and **7% URL overlap** with Google Search. These results demonstrate that GPT relies on conceptual reasoning and pre-trained knowledge rather than direct retrieval. Its responses reflect synthesized understanding rather than literal web citation.

GPT aligns with Google thematically but diverges at the source level, producing answers that paraphrase, summarize, or generalize existing web knowledge. This pattern confirms that GPT represents a reasoning layer on top of the search ecosystem rather than a reflection of it.

Gemini: Selective Precision



Gemini records **28% domain overlap** and **6% URL overlap** with Google Search. Despite being a Google-developed model, Gemini references a smaller, more curated set of sources. Its retrieval design favors selectivity and factual precision over citation volume.

The overlap data indicates that Gemini filters aggressively, citing only high-confidence or context-specific domains. While its domain overlap appears moderate, the small absolute count of shared URLs reveals that Gemini emphasizes accuracy and source verification over breadth.

Which Metrics Best Explain LLM–SERP Alignment?

The 2 core metrics analyzed are **Domain Overlap (%)** and **URL Overlap (%)**. Both metrics capture different layers of alignment between large language models and Google Search. Domain overlap measures how often LLMs cite the same websites found in SERP results. URL overlap measures how often they reference the exact pages ranked by Google.

Each metric contributes differently. **Domain Overlap** confirms that models share similar topic coverage and conceptual understanding with Google Search. **URL Overlap** reveals whether models retrieve identical web sources rather than summarizing from internal knowledge. The gap between the 2 shows where generative reasoning replaces direct citation.

Domain Overlap emerges as the stronger predictor of LLM–SERP alignment. It maintains stable performance across all query intents and platforms, especially in retrieval-based systems such as Perplexity. **URL Overlap** fluctuates widely and remains low across all models, confirming that literal page matching is rare even when topic similarity is high.

Together, these metrics prove that semantic understanding defines modern AI alignment more than link duplication. Retrieval systems achieve the closest parity with Google Search, while reasoning models replicate meaning instead of matching sources. **Domain overlap therefore represents the clearest indicator of how AI models interpret and mirror the web.**

What Should SEO and AI Teams Do with These Findings?

SEO and AI teams need to treat LLM Visibility as a core performance metric. Visibility inside AI-generated answers reflects brand exposure that occurs before a traditional Google search. Tracking this metric reveals how often a brand appears, how it is represented, and how consistently it dominates within AI ecosystems.

Teams need to align content for **semantic precision**. Pages with clear topical focus, factual grounding, and structured data achieve higher citation rates across both SERP and LLM responses. Semantic clarity improves the likelihood that large language models identify and reference the same sources that rank well in Google.

Benchmarking across retrieval platforms is essential. Comparing Perplexity, GPT, and Gemini citation patterns monthly reveals shifts in authority and emerging gaps between AI visibility and organic rankings. Retrieval-augmented systems reflect live search behavior, while reasoning-based systems surface conceptual authority. Monitoring both dimensions provides a complete visibility profile.

Integrating LLM Visibility dashboards inside Search Atlas enables unified tracking of AI and search performance. Access the data through [LLM Visibility feature](#) to analyze brand presence, sentiment, and citation overlap across ecosystems. The alignment of SERP and LLM signals defines the next frontier of SEO measurement.

What Are the Study's Limitations?

Every model has limitations. The limitations of this study are listed below.

- **Query Intent Coverage.** Some query types were unevenly represented, which have influenced overlap distributions across models.
- **Semantic Similarity Threshold.** The 0.82 similarity score ensures strong linguistic resemblance but does not guarantee identical user intent.
- **Temporal Scope.** The two-month period (September to October 2025) offers a focused yet limited snapshot of search and model behavior.
- **Model Design Differences.** Retrieval-enabled and non-retrieval systems differ structurally, which affects comparability in overlap measurement.

Despite these limits, the analysis confirms that retrieval-augmented systems achieve stronger alignment with Google Search, while reasoning-based systems prioritize semantic synthesis. The study establishes a clear baseline for understanding how LLMs interpret and mirror the web. Future analyses need to extend the timeframe, refine similarity thresholds, and explore longitudinal trends in LLM–Search correspondence.



Manick Bhan

Manick Bhan is a 3x INC 5000 founder and CTO of Search Atlas which is an AI SEO automation platform used by thousands of brands and agencies and awarded Best SEO Platform by the Global Search Awards, Shortlisted by Capterra, Front Runners by Software Advice, Category Leaders by GetApp, and best tool for customer satisfaction and usability by Gartner.