Google Antitrust Case: AI Overviews Use FastSearch, Not Links via @sejournal, @martinibuster

A sharp-eyed search marketer discovered the reason why Google's AI Overviews showed spammy web pages. The recent Memorandum Opinion in the Google antitrust case featured a passage that offers a clue as to why that happened and speculates how it reflects Google's move away from links as a prominent ranking factor.

Ryan Jones, founder of SERPrecon (LinkedIn profile), called attention to a passage in the recent Memorandum Opinion that shows how Google grounds its Gemini models.

Grounding Generative AI Answers

The passage occurs in a section about grounding answers with search data. Ordinarily, it's fair to assume that links play a role in ranking the web pages that an AI model retrieves from a search query to an internal search engine. So when someone asks Google's AI Overviews a question, the system queries Google Search and then creates a summary from those search results.

But apparently, that's not how it works at Google. Google has a separate algorithm that retrieves fewer web documents and does so at a faster rate.

The passage reads:

> *"To ground its Gemini models, Google uses a proprietary technology called FastSearch. Rem. Tr. at 3509:23–3511:4 (Reid). FastSearch is based on RankEmbed signals—a set of search ranking signals—and generates abbreviated, ranked web results that a model can use to produce a grounded response. Id. FastSearch delivers results more quickly than Search because it retrieves fewer documents, but the resulting quality is lower than Search's fully ranked web results."*

Ryan Jones shared these insights:

> *"This is interesting and confirms both what many of us thought and what we were seeing in early tests. What does it mean? It means for grounding Google doesn't use the same search algorithm. They need it to be faster but they also don't care about as many signals. They just need text that backs up what they're saying.*
>
> *…There's probably a bunch of spam and quality signals that don't get computed for fastsearch either. That would explain how/why in early versions we saw some spammy sites and even penalized sites showing up in AI overviews."*

He goes on to share his opinion that links aren't playing a role here because the grounding uses semantic relevance.

Related: Google Confirms Links Are Not That Important

What Is FastSearch?

Elsewhere the Memorandum shares that FastSearch generates limited search results:

> *"FastSearch is a technology that rapidly generates limited organic search results for certain use cases, such as grounding of LLMs, and is derived primarily from the RankEmbed model."*

Now the question is, what's the RankEmbed model?

The Memorandum explains that RankEmbed is a deep-learning model. In simple terms, a deep-learning model identifies patterns in massive datasets and can, for example, identify semantic meanings and relationships. It does not understand anything in the same way that a human does; it is essentially identifying patterns and correlations.

The Memorandum has a passage that explains:

> *"At the other end of the spectrum are innovative deep-learning models, which are machine-learning models that discern complex patterns in large datasets. …(Allan)*
>
> *…Google has developed various "top-level" signals that are inputs to producing the final score for a web page. Id. at 2793:5–2794:9 (Allan) (discussing RDXD-20.018). Among Google's top-level signals are those measuring a web page's quality and popularity. Id.; RDX0041 at -001.*
>
> *Signals developed through deep-learning models, like RankEmbed, also are among Google's top-level signals."*

User-Side Data

RankEmbed uses "user-side" data. The Memorandum, in a section about the kind of data Google should provide to competitors, describes RankEmbed (which FastSearch is based on) in this manner:

> *"User-side Data used to train, build, or operate the RankEmbed model(s); "*

Elsewhere it shares:

> *"RankEmbed and its later iteration RankEmbedBERT are ranking models that rely on two main sources of data: _____% of 70*

> *days of search logs plus scores generated by human raters and used by Google to measure the quality of organic search results."*

Then:

> *"The RankEmbed model itself is an AI-based, deep-learning system that has strong natural-language understanding. This allows the model to more efficiently identify the best documents to retrieve, even if a query lacks certain terms. PXR0171 at -086 ("Embedding based retrieval is effective at semantic matching of docs and queries");*
>
> *…RankEmbed is trained on 1/100th of the data used to train earlier ranking models yet provides higher quality search results.*
>
> *…RankEmbed particularly helped Google improve its answers to long-tail queries.*
>
> *…Among the underlying training data is information about the query, including the salient terms that Google has derived from the query, and the resultant web pages.*
>
> *…The data underlying RankEmbed models is a combination of click-and-query data and scoring of web pages by human raters.*
>
> *…RankEmbedBERT needs to be retrained to reflect fresh data…"*

A New Perspective On AI Search

Is it true that links do not play a role in selecting web pages for AI Overviews? Google's FastSearch prioritizes speed. Ryan Jones theorizes that it could mean Google uses multiple indexes, with one specific to FastSearch made up of sites that tend to get visits. That may be a reflection of the RankEmbed part of FastSearch, which is said to be a combination of

“click-and-query data” and human rater data.

Regarding human rater data, with billions or trillions of pages in an index, it would be impossible for raters to manually rate more than a tiny fraction. So it follows that the human rater data is used to provide quality-labeled examples for training. Labeled data are examples that a model is trained on so that the patterns inherent to identifying a high-quality page or low-quality page can become more apparent.

See also: Google Shares Valuable SEO Takeaway About Quality Raters Guidelines

Featured Image by Shutterstock/Cookie Studio

https://www.searchenginejournal.com/google-antitrust-case-ai-overviews-use-fastsearch-not-links/555220/