

AI Is Splitting Web Search Into Three Different Realities

Originally published at <https://www.unite.ai/ai-is-splitting-web-search-into-three-different-realities/>



New research finds that Google now uses three different information systems inside its own search empire, with regular Search, AI Overviews and Gemini all favoring different sources, rankings and content.

Reductivism rules. Over the last twelve months, the ‘Let me Google that for you’ [meme](#) has been overtaken by a new ‘Let me summarize that Google search for you’ trend, wherein [AI overviews](#) in search results increasingly spare readers the trouble of clicking on search links (arguably [de-funding the source sites](#) in the process), by condensing entire search results into a few generated paragraphs.

One would think that the core knowledge surfaced, and the choice of sites from which to draw that knowledge, would be relatively similar across all three of the most popular methods of searching the internet for information: in traditional *web search*; in the *AI overviews* (AIOs) that now head most web search results; and through the [increasing use](#) of LLMs such as ChatGPT as web-oracles (with or without [external RAG calls](#)).

However, recent research from the US indicates that this is, surprisingly, far from the case; and that even within Google’s own trinity of oracles – SERPS*, AI summaries, and direct interaction with the [Gemini LLM series](#) – there appear to be significant and interesting discrepancies, for each route.

Three-Way Split

In a lucid and extensive [new paper](#), titled *How Generative AI Disrupts Search: An Empirical Study of Google Search, Gemini, and AI Overviews*, six researchers from the New Jersey Institute of Technology outline the ways that the three search methods are diverging, and offer some possible theories for these fractures in approach.

The paper states:

‘[First, we] find that for 51.5% of representative, real-user queries, AIOs are generated, and are displayed above the organic search results. Controversial questions frequently result in an AIO.

‘Second, we show that the retrieved sources are substantially different for each search engine (<0.2 average Jaccard similarity). Traditional Google search is significantly more likely to retrieve information from popular or institutional websites in government or education, while generative search engines are significantly more likely to retrieve Google-owned content.

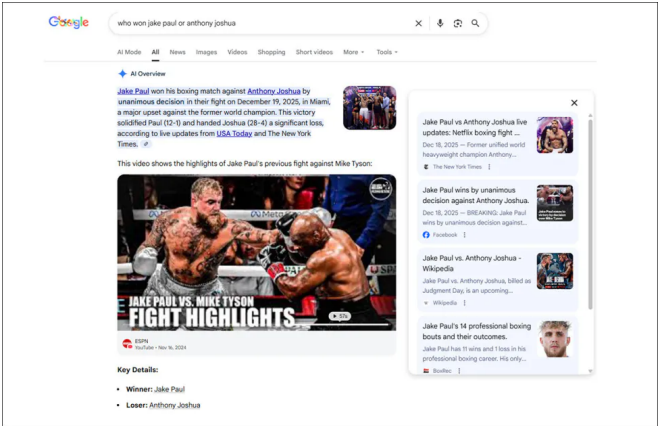
‘Third, we observe that websites that block Google’s AI crawler are significantly less likely to be retrieved by AIOs, despite having access to the content.’

Since the paper is a smorgasbord of fascinating insights, rather than conforming to the usual linear and method-driven workflow, we’ll take a closer look at these, and some other of its most surprising and illuminating insights.

The Old ‘Two-One’

One of the many interesting findings in the study indicates that Google’s AI overviews tend to be suppressed for sudden breaking news events, since the earliest and most available sources may not be the most accurate.

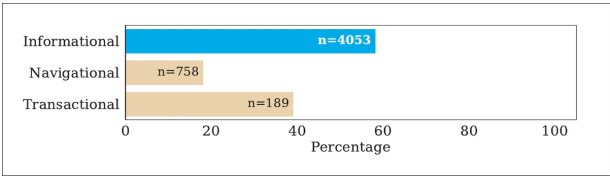
This system doesn't always work: in the example below, noted by the researchers, a Google AI overview regarding the result of a boxing match attributed the win to the wrong boxer, even though the only source stating this (incorrect) result was a satirical sports feed on Facebook:



One of the reasons that Google's AI overviews avoid time-critical summaries is that early information may be incomplete, or completely inaccurate. In this case, boxer Jake Paul actually lost the match. [Source](#)

The authors note that AIOs tend to appear when an event is at least five days old, which qualifies this as an anomaly – but nonetheless, one that the researchers were quite easily able to summon up.

AIOs were found to be more likely to be generated when the query was closed with a question-mark, and that *query intent* was a factor in whether an AIO would be presented:



Percentage of incidents where an AI search summary was produced in one of the researchers' round of tests. Here 'informational' indicates direct questions, which tend to produce AIOs more than any other type of interaction.

Additionally, the paper contends, *longer* queries tend to be more likely to produce an AI summary instead of just straight search results, though the authors do not yet proffer a theory to account for this.

A Kingdom Divided

Perhaps the most surprising overarching result from the new work is the relatively small crossover in results quality/type between Google's (aforementioned) three search platforms.

The paper repeatedly shows that regular Google Search, AI overviews and Gemini (LLM) retrieve strikingly different sources for the same query, with overlap scores low enough to imply three *competing retrieval logics* inside one company, whereas users might assume that Google has one authoritative index, and one ranking philosophy:

Query Subset	Jaccard			RBO		
	AIO	GEM	SERP	AIO	GEM	SERP
	SERP	SERP		SERP	SERP	
ORCAS	0.17	0.12	0.20	0.24	0.17	0.26
Amazon Retail	0.08	0.06	0.08	0.12	0.10	0.10
Retail-Comp	0.11	0.08	0.08	0.15	0.10	0.10
Retail-Q	0.12	0.10	0.10	0.16	0.14	0.13
Debate	0.24	0.10	0.11	0.27	0.12	0.14
ELI5	0.21	0.11	0.14	0.25	0.15	0.18
Localized	0.14	0.07	0.16	0.17	0.10	0.21
NQ	0.19	0.13	0.19	0.24	0.17	0.26
NQ Keywords	0.19	0.13	0.20	0.24	0.16	0.26
Total	0.18	0.11	0.16	0.23	0.15	0.21

Even inside Google's own ecosystem, the overlap between traditional Search, AI Overviews and Gemini proved surprisingly small, with the same query often producing substantially different source lists depending on which Google system handled the request. In this comparison, we see how closely the three systems matched each other across thousands of search queries, from shopping and debate topics to local searches and general knowledge questions, with lower scores indicating less agreement between the sources selected.

Regarding this section of their analysis, the authors state[†]:

[The table above] presents the average similarity between the list of sources returned by the AIO, Gemini, and traditional SERP for each query in the benchmark dataset.

'The main takeaway is that regardless of query subset and which pair of search engines is compared, the retrieved lists are dissimilar, despite all three being developed by Google.'

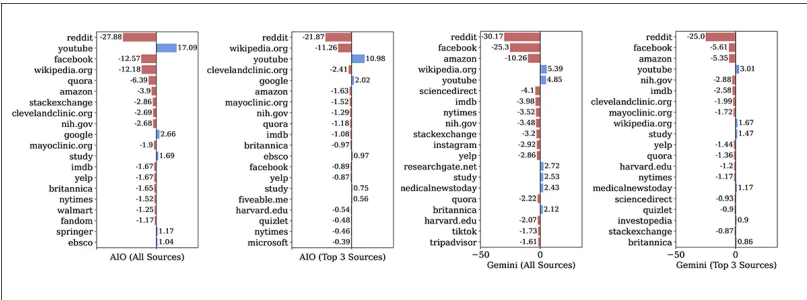
The researchers further state that no search method tested proved to have a [rank-biased overlap](#) (RBO) above 0.27, which is a very low score. They note further that Amazon Retail and localized queries (i.e., ‘shops near me’) had the lowest similarity among the search methods.

They attribute the low agreement to fundamental ‘inconsistency between search engines’, noting that neither randomness nor any other obvious factor can be made to account for this dis-syncoption.

One intuitive explanation, arguably, is that training data points are assigned rank in a very different way to the methods that Google has been developing for PageRank and its successors over the last two decades. Further, on the off-chance that Google Search’s algorithm [has a secret agenda](#), that kind of interference, or ‘gaming’, is much harder to consistently implement in [diffusion-based](#) AIs such as Gemini (even via filtering, system prompts, and the [various other methods](#) of corraling that are imposed on commercial models).

Self-Service..?

Certain websites, or categories of website, appear to have been affected by the advent of AI summaries and the encroachment of LLM-based search into the traditional search space – both adversely and positively, depending on the case:



Compared to traditional Google Search, AI Overviews and Gemini both reduced citations from many major websites, while increasing visibility for a smaller number proved to be one of the biggest beneficiaries across both systems, while Reddit, Wikipedia, Facebook and many institutional sources appeared less frequently in

The authors note that some unexpected preferences emerge among the three methods, during testing:

- ‘We have three main takeaways from [the graphs above]. First, large and well-known websites are the most affected (both positively and negatively). This is intuitive as large websites have the reputation and diversity in content to be relevant to many different queries.
- ‘Second, the overwhelming majority of these websites receive fewer overall, and fewer top three, citations with generative search engines (indicated by red bars and negative numbers in [graphs above]). This suggests that generative search tends to source information from more niche sources than traditional search engines.
- ‘Third, Google’s AIOs favor Google websites (i.e., google.com and youtube.com domains).
- ‘Gemini also favors YouTube in comparison to traditional Google Search, but the absolute difference is smaller.’

Any ‘Blockers’..?

The study found also that publishers which block [Google’s AI web crawler](#) – the automated web-bot that scrapes data from your site unless you tell it not to with a [robots.txt](#) file – tend to not appear in AI summaries.

This may seem like an obviously self-inflicted wound, but in fact Google [has publicly stated](#) that content from platforms that block AI crawlers *will not be prevented from appearing in AI summaries*; rather, the publishers will simply not have their data scraped, curated into a collection, and run through the next round of AI training for Gemini and other Google AI projects.

However, that was not the conclusion that the new paper’s researchers came to, finding instead that popular AI-banning publishers were very infrequently cited by Gemini, either in the LLM or stripped-down and more agile search-results incarnation. The ‘effectively banned’ publishers were reported by the paper to be NYTimes, CNN, BBC, ScienceDirect, Reuters, Wiley, Nature, ESPN, Business Insider, CNBC, NPR, WIRED, USA Today, NBC News, Genius, National Geographic, The Conversation, U.S. News & World Report, Scientific American, Consumer Reports, and STAT.

User-agent: bycespider	Disallow: /
User-agent: CCBot	Disallow: /
User-agent: cohere-ai	Disallow: /
User-agent: DiffBot	Disallow: /
User-agent: FacebookBot	Disallow: /
User-agent: FriendlyCrawler	Disallow: /
User-agent: GPTBot	Disallow: /
User-agent: ImagesiftBot	Disallow: /
User-agent: meta-externalag	Disallow: /
User-agent: Meta-ExternalAg	Disallow: /
User-agent: PhindBot	Disallow: /
User-agent: AI2Bot-DeepResearchEval	Disallow: /
User-agent: KlaviyoAIBot	Disallow: /
User-agent: LinerBot	Disallow: /
User-agent: Poggio-Citations	Disallow: /
User-agent: Devin	Disallow: /
User-agent: bigsur.ai	Disallow: /
User-agent: ChatGPT Agent	Disallow: /

Some of the robots.txt AI-scraping bans effected by the publishers listed above. But has it led to a wider censure by Google?

The authors state:

'In our analyses of the most affected domains, we found that 21 popular [publishers] (which are retrieved for at least 20 unique queries by both Google Search and AIOs) were never cited by Gemini.'

'Several popular social media (Facebook, Instagram, Tiktok) and review websites (IMDb, Yelp, Tripadvisor) also received zero citations from Gemini. Upon further investigation, we found that all of these websites block the Google-Extended bot in their robots.txt files.'

If this finding turns out to be verified elsewhere and persistent, one could speculate that these companies are potentially being pressured by Google into capitulation and cooperation with its AI operations through partial de-listing. At a glance, the results seem punitive – but then, the findings of the new work are more indicative of chaos than premeditation; therefore the only reasonable comment one can settle on is that these results look superficially 'spiteful', whatever is really causing them.

Conclusion

Opinion This is a clear-headed [zip-bomb](#) of a paper, whose mere ten primary pages unfurl into an almost overwhelming cascade of additional findings. Since we have had time to cover only a small section of these, I recommend the source PDF even to the casual reader (a rare event).

Though a 'yellow' disposition could cast many negative interpretations on the authors' discoveries, the work is perhaps best treated as indicative of a global tech leader attempting to obtain and retain a global lead in AI-based search, using highly-contrasting platforms which developed in very different circumstances and across very different eras.

While three search methods are examined in the paper, the real contention is between traditional search engine results, ranked by proprietary methods, and the sharply-contrasting distribution-based selection methods that dominate data curation and AI training.

AI Like It's 1999

[Before the advent of Google](#), it was possible to 'game' search results through sheer volume, and in this way, one could often achieve front-page SERPS placement with minimal (frequently automated) effort. This 'numbers game' was effectively ended around 2002 by Google's more sophisticated and secretive search ranking algorithm. But since the stakes were significant, high-volume and low-quality content never went away in any meaningful sense.

Therefore, by the time hyperscale collections such as [Common Crawl](#) set the foundations of the modern AI revolution, data-prominence was destined to be dominated by the extent to which automated processes could *filter and rank incoming data quality*, and (far less likely), the extent to which money was available to pay people to rank that data.

There was a [lot of bad or low-quality data](#) in those huge and indiscriminate collections; data that may not have included nudity or swearing or racist tropes, or any of the other things that are relatively easy to filter out of training datasets – but which was nonetheless self-serving and voluminous, just like results from internet search circa 1999-2001.

Because those data induction processes [are still not great](#), it is very hard even for Google to get AI to act in a business-like manner, since Gemini's PageRank-style decisions are dictated not by Google's policy engineers, but by an imperfect understanding of how hyperscale data transforms into data distributions and latent embeddings during the training of an AI model.

* Search Engine Results Pages.

† Authors' emphasis/es, not mine. However, I have substituted bold for italic, since italic emphasis does not work well in quotes that are already primarily italic.

First published Wednesday, May 13, 2026

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.

Portfolio site: [martinanderson.ai](#)

Contact: martin@martinanderson.ai



Discover More