Google Wants To Show More Links In AI Mode via @sejournal, @MattGSouthern

Google says it’s actively working to surface more source links inside AI Mode.

Robby Stein, VP of Product for Google Search, outlined changes designed to make links more visible.

Stein wrote on X that Google has been testing where links appear inside AI answers and that the long-term “north star” is to show more inline links.

He added that people are more likely to click when links are embedded with context directly in the response.

Stein stated:

> *“We’ve been experimenting with how and where to show links in ways that are most helpful to users and sites… our long term north star is to show more inline links.”*

> *1/ Excited about the work the team is doing to create AI experiences in Search that highlight useful links and encourage onward exploration. We’ve been experimenting with how and where to show links in ways that are most helpful to users and sites, and you’ll be seeing some of…*
>
> *— Robby Stein (@rmstein) August 25, 2025*

What’s Changing

Link Carousels On Desktop.

Google has launched carousels that surface multiple source links directly inside AI Mode responses on desktop. Stein said mobile support is coming soon.

The idea is to present links with enough context to help people decide where to go next without hunting below the answer.

> *2/ We’ve found that people really prefer and are more likely to click links that are embedded within AI Mode responses, when they have more context on what they’re clicking and where they want to dig deeper. We’ve launched embedded link carousels in AI Mode responses on desktop,…*
>
> *— Robby Stein (@rmstein) August 25, 2025*

Smarter Inline Links

Google is rolling out model updates that decide where inline links appear within the response text.

The system is trained to place links at moments when people are most likely to click out to see where information came from or to learn more.

Stein noted you might see fluctuations over the next few weeks as this is deployed, with a longer-term push toward more inline links overall.

> *3/ We’re also launching some model updates to improve how we show inline links (links that are embedded directly within text) in AI Mode responses. We train the model to understand where and when people are most likely to want to click out, see where info is coming from and learn…*

> — Robby Stein (@rmstein) August 25, 2025

Web Guide

Separately, Google's Web Guide experiment uses a custom Gemini model to group useful links by topic.

It launched in Search Labs on the "Web" tab and, for opted-in users, will begin appearing on the main "All" tab when systems determine it could help for a query.

Google introduced Web Guide in July and indicated it would expand beyond the Web tab over time.

> 4/ We're also expanding our Web Guide experiment in Labs, which is a new approach to intelligently surfacing and organizing the most useful web links with AI — even for your hardest queries. We've gotten some really positive feedback as we've tested this on the "Web" tab. So in…
>
> — Robby Stein (@rmstein) August 25, 2025

Why It Matters

How Google presents links in AI Mode can influence how people reach your site.

Placing carousels within the answer and adjusting inline placements differ from links that appear only below the response. This may change click behavior depending on the query and presentation.

Looking Ahead

Google is trying to strike a balance between innovation and supporting publishers. Expect continued testing around link density, placement, and labeling as Google refines AI mode.

Featured Image: subh_naskar/Shutterstock

https://www.searchenginejournal.com/google-wants-to-show-more-links-in-ai-mode/554644/

Google Rolls Out August Spam Update Targeting Search Quality via @sejournal, @MattGSouthern

Google released the August spam update, impacting search results worldwide. The rollout may take a few weeks to complete.

- The August spam update started August 26 and applies worldwide.
- Expect possible ranking and traffic swings during the multi-week rollout.
- Monitor trends in Search Console and stay aligned with Google's spam policies.

https://www.searchenginejournal.com/google-rolls-out-august-spam-update-targeting-search-quality/554611/

ChatGPT Vs. Google At Every Stage Of The User Journey via @sejournal, @Kevin_Indig

Boost your skills with Growth Memo's weekly expert insights. Subscribe for free!

More data shows ChatGPT isn't taking market share away from Google.

Instead, it's expanding the range of use cases and blurring the line between searching for information and performing tasks.

I looked at Similarweb data to understand how this affects four different stages of the user journey across Google and ChatGPT:

1. Usage.
2. Behavior.
3. Outbound clicks.
4. Converting.

What I found is that ChatGPT adoption is, essentially, a 400,000-pound locomotive barreling down the tracks with no intention of stopping anytime soon.

User conversations within ChatGPT are rich in context, which leads to higher conversion rates when intent shifts from information seeking or generating to buying.

Lastly, and also most notably for SEOs and growth marketers, **ChatGPT is sending a lot more users out to the web.**

Of course, all of these stats are still small in comparison to Google.

However, no effort from Google has been able to slow the momentum of ChatGPT’s runaway train. **About the data I used in this analysis:**

- Data source: Similarweb (shoutout to Sam Sheridan).
- Time period examined: July 2024 – June 2025 (last 12 months) vs. July 2023 – June 2024 (previous 12 months).
- I also examined U.S. vs. UK user behavior.

Image Credit Kevin Indig

People are rushing to ChatGPT.

Over the last 12 months in the U.S., ChatGPT visits grew from 3.5 to 6.8 billion visits (+94%).

In the UK, it was even faster: 131% YoY, from 868 million to 2

billion.

Over the same time span, Google growth stagnated. Here’s what the data showed:

- U.S. stagnation: -0.85% (196 vs. 194 billion).
- UK stagnation: -0.22% (35.56 vs. 35.49 billion).

Image Credit: Kevin Indig

To put it into perspective: Google had almost 200 billion visits in the U.S. over the last 12 months, compared to ChatGPT’s 6.8 billion.

So, ChatGPT has a mere 3.4% of Google’s total visits.

However, if growth rates hold steady, in theory, ChatGPT could hit Google’s volume in the next five years.

My hypothesis: It’s almost guaranteed that ChatGPT won’t hit Google’s visit volume because there are too many moving parts (energy/chip limitations, training data, quality improvements, regulation, etc.).

But consider that Google has declined by -0.85% (~2 billion visits) year-over-year, and you can see where this is going.

Visits can only tell you so much.

Recent data from Semrush and Profound suggests that one-third to two-thirds of user intent when interacting with AI chatbots is generative, meaning users use ChatGPT to do and less to search [1, 2].

Leaked chats from ChatGPT and other AI chatbots confirm the aggregate data.

So, even when we compare visits to ChatGPT vs. Google, they’re not leading to the same outcome.

But, against that argument, I will say that Google morphs more into a mirror of ChatGPT with AI Mode – and every generative intent has a high chance of including information along the conversation journey that is sourced to other sites or creators.

The conversational nature of AI chatbots means intent is fluid and can change from one prompt to the next.

Along the way, it’s likely users come across information in their conversations that would’ve been a classic Google Search for products or solutions.

At the end of the day, ChatGPT is continuing its adoption as the fastest-growing product on earth to date.

What does that mean for you?

- Stay the course.

- Keep tracking referral traffic, conversions, and topic visibility on Google + ChatGPT.
- Optimize for visibility with a strong focus on classic SEO.
- Keep an ear to the ground and learn as much as you can. Things are evolving fast, and clarity will come with time.

Quick reminder here: *I recently transitioned my WhatsApp group over to Slack. I share ongoing news and learnings throughout the week openly and freely, so it's a great place to stay updated without all the extra (and sometimes overwhelming) noise. No need to be a premium subscriber to get access to the main discussion channel. Join here!*

Old habits are hard to break.

People are used to searching on Google a certain way, while ChatGPT is a green field.

For the overwhelming majority of us, our first experience with ChatGPT was a conversation, so we all adopted it as the default way to engage.

Image Credit: Kevin Indig

As a result, the average ~~query~~ prompt length on ChatGPT vs. Google is:

- 80 words vs 3.4 in the U.S.
- 91 words vs. 3.2 in the UK.

Even informational prompts are 10 times longer (~38 words) on ChatGPT than on Google. People ask more detailed questions, which reveal much more about themselves *and* their intent.

Together with a growing context window, ChatGPT returns much more personalized and (usually) better informational answers – I'm still waiting on consistently better commercial/purchase intent outcomes.[3] AI chatbots compress the user journey from many queries over several days to one conversation with

lengthy prompts.

For you, this means it’s even more critical to monitor the right prompts.

(I shared a trick with premium subscribers for finding prompts in Google Search Console from AI Mode in Is AI cutting into your SEO conversions?)

As referral traffic from Google reached historic lows, ChatGPT’s referral traffic reached new highs.

Image Credit: Kevin Indig

Over the last 12 months, ChatGPT’s U.S. referral traffic to websites jumped by +3,496% (UK: +5,950%), from 14 to 516 million (after cleaning up referrals to Openai.com, which are

mostly authentications).

In comparison, Google's outgoing referral visits grew only +23% in the U.S. and 19% in the UK.

When you consider Google referrals include navigational searches (people navigating to the homepage of a brand) and ad clicks (ChatGPT doesn't yet have ads), 23% is not much at all.

ChatGPT's referral traffic to external websites makes up ~27% of Google's (1.9 billion, in the last 12 months), based on the data. That feels high, in my field observation.

Also consider that ChatGPT's goal is not necessarily to send out traffic but to keep the conversation going until users have the optimal response.

That being said, referral traffic has grown and continues to do so. Until recently.

According to Profound, ChatGPT's referral traffic was down -52% between July 21 and August 20. [4] And that's significant.

Time will tell whether this is just an experiment or a final decision.

For you, this means you should see more ChatGPT referral traffic over the last 12 months if you optimize well.

You might not see an increase of +3,500%, but if you're not seeing at least some growth, it's likely your competitors are.

Conversions from ChatGPT are small in comparison with Google (in volume), but they're growing rapidly.

The whole narrative of investing in AI visibility optimization (AEO/GEO/LLMO) banks on the fact that it will continue at the same pace and become meaningful.

So far, it seems like that bet will work out.

Image Credit: Kevin Indig

When ChatGPT sends traffic to sites, the conversion rate is usually higher than Google's. As of June 2025:

- ChatGPT's conversion rate of transactional traffic was 6.9% in the U.S. compared to 5.4% for Google.
- In the UK, ChatGPT reached 5.5%, which is on par with Google.

ChatGPT sends higher-quality traffic to websites, at least in the U.S.

I define quality in this context as "higher intent," meaning visitors are more likely to convert into customers.

The reason ChatGPT traffic is of higher quality is that users

get answers to their questions in one conversation. When they click out, they're "primed" to buy.

To me, the bigger question is how purchase decisions are influenced before a click happens (or even when no click-out happens).

For you, this means:

1. Look at which pages get referral traffic. Take the average referral traffic and optimize pages that get some but below-average clicks.
2. Optimizing for citations matters because citations are what get clicked. Look at the citation gap between your competitors and your site.
3. Look for conversion optimization opportunities (in-line CTAs, lead gen assets, quizzes, etc) on pages that get ChatGPT referral traffic. Using a standard heatmap tool will point you to areas of the page that are ideal for a little CRO.

ChatGPT has all the ingredients to become the next big user platform on which other companies can build – just like Google 25 years ago:

1. Usage is growing.
2. Behavior is rich in context.
3. Referral traffic is shooting up.
4. Conversions happen at a healthy rate.

Now, traffic and conversations just need more volume.

They're still tiny in comparison.

Featured Image: Paulo Bobita/Search Engine Journal

https://www.searchenginejournal.com/chatgpt-vs-google-at-every-stage-of-the-user-journey/554546/

Non-Profit Organization Announces Free Domain Names via @sejournal, @martinibuster

A non-profit organization that is supported by Cloudflare, GitHub, and other organizations has open-sourced domain names, making them available with no catches or hidden fees. The sponsor of the free domain names explains that their purpose is not to replace commercial domain names but to offer an open-source alternative for developers, students, and people who want to create a hobby site for free.

The goal is to encourage making the Internet a free and open space so that everyone can publish and express themselves online without financial barriers.

DigitalPlat

The open source domains are offered by DigitalPlat, a non-profit organization that's sponsored by 1Password, The Hack Club (The Hack Foundation), twilio, GitHub and Cloudflare.

The Hack Foundation is a certified non-profit organization of high school students that receive support from hundreds of supporters including Google.org and Elon Musk. The organization was founded in 2016.

According to their website:

> *“In 2018, The Hack Foundation expanded to act as a nonprofit fiscal sponsor for Hack Clubs, hackathons, community organizations, and other for-good projects.*
>
> *Today, hundreds of diverse groups ranging from a small town newspaper in Vermont to the largest high-school hackathon in Pennsylvania are fiscally sponsored by The Hack Foundation.”*

A notice posted on The Hack Foundation donation web page explains their connection to DigitalPlat:

> *“The DigitalPlat Foundation is a global non-profit organization that supports open-source and community development while exploring innovative projects. All funds are supervised and managed by The Hack Foundation, and are strictly regulated in compliance with US IRS guidance and legal requirements under section 501(c)(3). “*

DigitalPlat FreeDomain

The free domain names can be registered via DigitalPlat and the free domains project is open source, licensed under AGPL-3.0.

An announcement was made by the GitHubs Projects Community on X with a link to a GitHub page for the free domains where the following domain extensions are listed as choices:

- .DPDNS.ORG
- .US.KG
- .QZZ.IO
- .XX.KG

Technically, those are subdomains. But so are .uk.com domains.

The official GitHub page for the domains recommends using Cloudflare, FreeDNS by Afraid.org, or Hostry for managing the DNS for zero cost.

The .KG domain is from the country code of Kyrgyzstan. DPDNS.ORG is the domain name of DigitalPlat FreeDomain. .US.KG is operated by the DigitalPlat Foundation, a non-profit charitable organization that's sponsored by The Hack Foundation.

The Open-Source Projects page for the free domains explains the purpose and goals of the free domain offers:

> *"The project is open source (licensed under AGPL-3.0), transparent, and backed by The Hack Foundation, a U.S. 501(c)(3) nonprofit. This isn't a trial or a limited-time offer—it's a sustainable effort to increase accessibility on the web."*

Full directions for registering a free domain name can be found here.

Featured Image by Shutterstock/TenPixels

https://www.searchenginejournal.com/free-domain-names/554514/

Semantic Overlap Vs. Density: Finding The Balance That Wins Retrieval via @sejournal, @DuaneForrester

Marketers today spend their time on keyword research to uncover opportunities, closing content gaps, making sure pages are crawlable, and aligning content with E-E-A-T principles. Those things still matter. But in a world where generative AI

increasingly mediates information, they are not enough.

The difference now is retrieval. It doesn’t matter how polished or authoritative your content looks to a human if the machine never pulls it into the answer set. Retrieval isn’t just about whether your page exists or whether it’s technically optimized. It’s about how machines interpret the meaning inside your words.

That brings us to two factors most people don’t think about much, but which are quickly becoming essential: *semantic density* and *semantic overlap*. They’re closely related, often confused, but in practice, they drive very different outcomes in GenAI retrieval. Understanding them, and learning how to balance them, may help shape the future of content optimization. Think of them as part of the new on-page optimization layer.

Image Credit:: Duane Forrester

Semantic density is about meaning per token. A dense block of text communicates maximum information in the fewest possible words. Think of a crisp definition in a glossary or a tightly written executive summary. Humans tend to like dense content because it signals authority, saves time, and feels efficient.

Semantic overlap is different. Overlap measures how well your content aligns with a model's latent representation of a query. Retrieval engines don't read like humans. They encode meaning into vectors and compare similarities. If your chunk of content shares many of the same signals as the query embedding, it gets retrieved. If it doesn't, it stays invisible, no matter how elegant the prose.

This concept is already formalized in natural language

processing (NLP) evaluation. One of the most widely used measures is **BERTScore** (https://arxiv.org/abs/1904.09675), introduced by researchers in 2020. It compares the embeddings of two texts, such as a query and a response, and produces a similarity score that reflects semantic overlap. BERTScore is not a Google SEO tool. It's an open-source metric rooted in the BERT model family, originally developed by Google Research, and has become a standard way to evaluate alignment in natural language processing.

Now, here's where things split. Humans reward density. Machines reward overlap. A dense sentence may be admired by readers but skipped by the machine if it doesn't overlap with the query vector. A longer passage that repeats synonyms, rephrases questions, and surfaces related entities may look redundant to people, but it aligns more strongly with the query and wins retrieval.

In the keyword era of SEO, density and overlap were blurred together under optimization practices. Writing naturally while including enough variations of a keyword often achieved both. In GenAI retrieval, the two diverge. Optimizing for one doesn't guarantee the other.

This distinction is recognized in evaluation frameworks already used in machine learning. BERTScore, for example, shows that a higher score means greater alignment with the intended meaning. That overlap matters far more for retrieval than density alone. And if you really want to deep-dive into LLM evaluation metrics, this article is a great resource.

Generative systems don't ingest and retrieve entire webpages. They work with chunks. Large language models are paired with vector databases in retrieval-augmented generation (RAG) systems. When a query comes in, it is converted into an embedding. That embedding is compared against a library of content embeddings. The system doesn't ask "what's the best-written page?" It asks "which chunks live closest to this

query in vector space?”

This is why semantic overlap matters more than density. The retrieval layer is blind to elegance. It prioritizes alignment and coherence through similarity scores.

Chunk size and structure add complexity. Too small, and a dense chunk may miss overlap signals and get passed over. Too large, and a verbose chunk may rank well but frustrate users with bloat once it’s surfaced. ***The art is in balancing compact meaning with overlap cues, structuring chunks so they are both semantically aligned and easy to read once retrieved.*** Practitioners often test chunk sizes between 200 and 500 tokens and 800 and 1,000 tokens to find the balance that fits their domain and query patterns.

Microsoft Research offers a striking example. In a 2025 study analyzing 200,000 anonymized Bing Copilot conversations, researchers found that information gathering and writing tasks scored highest in both retrieval success and user satisfaction. Retrieval success didn’t track with compactness of response; it tracked with overlap between the model’s understanding of the query and the phrasing used in the response. In fact, in 40% of conversations, the overlap between the user’s goal and the AI’s action was asymmetric. Retrieval happened where overlap was high, even when density was not. Full study here.

This reflects a structural truth of retrieval-augmented systems. Overlap, not brevity, is what gets you in the answer set. Dense text without alignment is invisible. Verbose text with alignment can surface. The retrieval engine cares more about embedding similarity.

This isn’t just theory. Semantic search practitioners already measure quality through intent-alignment metrics rather than keyword frequency. For example, Milvus, a leading open-source vector database, highlights overlap-based metrics as the right

way to evaluate semantic search performance. Their reference guide emphasizes matching semantic meaning over surface forms.

The lesson is clear. Machines don’t reward you for elegance. They reward you for alignment.

There’s also a shift in how we think about structure needed here. Most people see bullet points as shorthand; quick, scannable fragments. That works for humans, but machines read them differently. To a retrieval system, a bullet is a structural signal that defines a chunk. What matters is the overlap inside that chunk. A short, stripped-down bullet may look clean but carry little alignment. A longer, richer bullet, one that repeats key entities, includes synonyms, and phrases ideas in multiple ways, has a higher chance of retrieval. In practice, that means bullets may need to be fuller and more detailed than we’re used to writing. Brevity doesn’t get you into the answer set. Overlap does.

If overlap drives retrieval, does that mean density doesn’t matter? Not at all.

Overlap gets you retrieved. Density keeps you credible. Once your chunk is surfaced, a human still has to read it. If that reader finds it bloated, repetitive, or sloppy, your authority erodes. The machine decides visibility. The human decides trust.

What’s missing today is a composite metric that balances both. We can imagine two scores:

Semantic Density Score: This measures meaning per token, evaluating how efficiently information is conveyed. This could be approximated by compression ratios, readability formulas, or even human scoring.

Semantic Overlap Score: This measures how strongly a chunk aligns with a query embedding. This is already approximated by tools like BERTScore or cosine similarity in vector space.

Together, these two measures give us a fuller picture. A piece of content with a high density score but low overlap reads beautifully, but may never be retrieved. A piece with a high overlap score but low density may be retrieved constantly, but frustrate readers. The winning strategy is aiming for both.

Imagine two short passages answering the same query:

Dense version: "RAG systems retrieve chunks of data relevant to a query and feed them to an LLM."

Overlap version: "Retrieval-augmented generation, often called RAG, retrieves relevant content chunks, compares their embeddings to the user's query, and passes the aligned chunks to a large language model for generating an answer."

Both are factually correct. The first is compact and clear. The second is wordier, repeats key entities, and uses synonyms. The dense version scores higher with humans. The overlap version scores higher with machines. Which one gets retrieved more often? The overlap version. Which one earns trust once retrieved? The dense one.

Let's consider a non-technical example.

Dense version: "Vitamin D regulates calcium and bone health."

Overlap-rich version: "Vitamin D, also called calciferol, supports calcium absorption, bone growth, and bone density, helping prevent conditions such as osteoporosis."

Both are correct. The second includes synonyms and related concepts, which increases overlap and the likelihood of retrieval.

This Is Why The Future Of

Optimization Is Not Choosing Density Or Overlap, It’s Balancing Both

Just as the early days of SEO saw metrics like keyword density and backlinks evolve into more sophisticated measures of authority, the next wave will hopefully formalize density and overlap scores into standard optimization dashboards. For now, it remains a balancing act. If you choose overlap, it’s likely a safe-ish bet, as at least it gets you retrieved. Then, you have to hope the people reading your content as an answer find it engaging enough to stick around.

The machine decides if you are visible. The human decides if you are trusted. Semantic density sharpens meaning. Semantic overlap wins retrieval. The work is balancing both, then watching how readers engage, so you can keep improving.

More Resources:

This post was originally published on Duane Forrester Decodes.

Featured Image: CaptainMCity/Shutterstock

https://www.searchenginejournal.com/semantic-overlap-vs-density-finding-the-balance-that-wins-retrieval/554251/

The B2B SEO Trap: Why High-Intent Visibility Can Still Underperform via @sejournal, @coreydmorris

In B2B SEO, typically, the ultimate goal is to attract high-intent searchers who convert into qualified leads.

But, not all high-intent visibility translates to sales-ready traffic, especially in long sales cycles or complex buying journeys.

As we get deeper into an era of diminishing importance of keywords that translate directly into attributable clicks, a focus on quality of traffic is as important as ever.

Don’t get me wrong. Quality has long been a critical component and key performance indicator (KPI) in the sense that most of us know our conversion rates and stages of the funnel someone might be in related to intent with B2B and lead generation.

However, now, as much as ever, we have to scrutinize intent and the traffic quality even more.

When SEO teams over-prioritize high-intent visibility and focus under the false assumption that intent equals urgency, we can find that we’re not getting the conversions or leads that we expect.

In this article, I’m unpacking seven ways to help go deeper with mapping visibility to actual funnel stages, differentiating decision-makers vs. influencers, and building SEO content that nurtures, qualifies, and educates before the

handoff to sales.

1. The Myth: “If It’s A Bottom-Of-Funnel Topic, The Traffic Is Ready To Convert”

While we may think about our website, our content, and analytics data in customer journeys and conversion funnels, our target audience doesn’t.

Maybe my nerdy search brain thinks that way when I’m consuming content on a website I’m interested in, but most of the world doesn’t.

There are too many variables that impact someone’s conversion decision to fully unpack here. I can personally tell you that a couple of times, I filled out forms while lying on the floor next to one of my kids in bed, falling asleep.

And, other times, I sat with a tab open on my giant desktop work monitor for months before eventually finding that tab again and filling out the form.

Those are two extreme examples, but as more possible ways to be found (e.g., AI search) enter play, we’re going to see myriad behaviors and paths that we couldn’t have anticipated just a few years ago.

Things are not going to make sense in a simple way, and what might seem like a home run conversion will be frustrating when you dig into the data, and things that might have felt like top of funnel that convert quickly will be equally (but pleasantly) surprising as well.

2. B2B Searchers Are Often Researching, Comparing, Or Gathering Info, Not Buying

Beyond the intent challenges and varying entry points and sources I noted above, we have a hand to play at times as well in how and where someone converts.

Our best converting content can still be standing in the way of getting the actual conversion. Just like the head scratchers that we might find when someone converts quickly without seemingly spending much time on the site or in "research" mode.

With more answers being given within search engines and large language models (LLMs), a lot of the research is done before someone gets to our site.

That being said, whether our content is helping inform AI, getting us found off-site, or to do the traditional education work on our site, we have to understand that, even on what might seem like a high intent page, someone might still be in research and information gathering mode.

They might be seeking pricing (if we disclose it) or building their own deck of us, plus competitors, to help with their own decisions for buying or outreach.

When we make too many assumptions, put someone on a singular navigation path, or take away options, we risk losing the opportunity for them to continue their research journey.

We've got to find a balance between prominent calls-to-action (CTAs) and long-form content so there's more flexibility for the user based on what their intent is in that visitor or session that we worked so hard to get.

3. How To Differentiate Between User Intent And Sales-Readiness

I've talked a lot about intent already. What I haven't unpacked is how sales-ready someone is.

Our brand story, content, and user experience can be persuasive and do the job of getting a form submission or phone call.

However, if someone isn't "sales-ready," they likely are going to consume everything up to the point of a conversion action, then leave. They may come back often up to that point and leave.

This might lead us to think there's something wrong with the form, or the CTA, or the content itself. Sure, that could be the case and should be validated. But, it could also be that they simply aren't ready to buy.

As an agency owner, I also operate a B2B business that relies on lead generation. I can personally validate that while we have received a lot of seemingly bottom-of-the-funnel traffic, my team has been told by prospects that they were ready to buy, but not ready to talk to anyone, as they were told to slow down the process or await a final budget approval before reaching out.

It's frustrating, but that's seemingly the nature of the world economically these past couple of years (probably not a "hot take").

4. Why One-Size-Fits-All CTAs (E.g., "Request A Quote") Often

Fail In B2B

I admit that I've been guilty of slapping the one-size-fits-all, generic CTA in the footer or sidebar of all pages of a site.

As I noted earlier, we need to expect the unexpected with matching intent to content and funnel levels. We should definitely review and evaluate our CTAs.

In line with my note above about someone possibly being close to a conversion but not sales-ready, if we have other areas of value we can provide like additional content they can subscribe to or ways to engage with us to get further acquainted (ex: webinars, Q&As, etc), that don't involve a direct sales process, then we can further engage with them and stay in front of them in a way that is welcome based on where they are right now.

5. Using Content To Build Trust And Qualification, Not Just Capture Form Fills

When we rush someone to a form submission and they're not ready to buy, not prepared for the sales process, or qualified, we often get feedback from sales about discrepancies related to marketing qualified leads (MQLs) vs. sales qualified leads (SQLs) or how leads are accepted by sales.

Wasting time on sales internally, while frustrating, someone who wasn't prepared for (or qualified for) the process is a loss for both sides.

Building trust through quality content, differentiation, setting expectations on what happens after the form submission, and other trust signals like transparency of

pricing can go a long way to ensuring higher rates of conversions to customers.

Don’t forget that quality trumps quantity if we look at additional metrics and KPIs in our marketing-to-revenue process.

6. How To Structure B2B Content Around Intent Clusters, Not Just Funnel Stages

If you aren’t convinced yet from what I’ve shared about how user behavior can differ from what we might expect or predict, then maybe thinking about content specifically will help.

In the zero-click searches and AI search push that has taken focus away from specific keyword and has put it more on visibility, one piece consistently is important: the content you create.

In topics, clusters, or however you want to think about how the content is organized on your website, you still need to focus on how it is presented to the user.

Starting with the user intent and mapped to where they are in the funnel, then working backwards, we can see where we have gaps in content and what we need to support answering all questions possible and moving the prospect forward in the process.

This will serve you well for search engines today and LLMs and AI-generated search results today and tomorrow.

While we used to (and in some cases still approach it today) as topics driven by keywords, I’m advocating for thinking about topics in how someone might be moving through a customer journey.

What questions are they asking at the phase they are in? Have we anticipated everything? Have we accidentally assumed too much about their knowledge or their sales-readiness?

We're not going to be able to think of everything. Much like long-tail keywords and queries, we can see people doing a lot more research and probing in AI research.

My company got a lead from ChatGPT a few months ago, and we could see that they visited our site seven times from ChatGPT in the process before eventually filling out our form. This is not user behavior that we would have planned for or anticipated just a few years ago.

7. Creating SEO Content For Both Decision-Makers And Gatekeepers

We can't control who comes to our website. Humans aren't as blockable as robots and web crawlers. However, we don't need to be worried about those who might not be the ultimate prospect or decision-maker.

Whether you're seeing traffic from AI engines, search engines, or those that never convert and seem like unqualified human visitors, I encourage you to still work on building your authority position, be helpful with your content, and to know that you might be helping get critical information to gatekeepers (human or systems) that go further upstream to a human who is a decision-maker.

Whether you're educating the search committee for a Request for Proposal (RFP) process, an assistant or intern doing field research, or something automated trying to learn so it can feed good info to a decision-maker, it isn't wasted effort, even though narrow thinking and reviewing of conversion metrics at the bottom of the funnel may make it seem that way.

Final Thoughts

Customer journeys, funnel thinking, search intent, and how it all works together in generating conversions and leads for B2B focuses can be complex and hard to track. It is getting even harder.

That doesn’t mean that we should give up or try to force everyone through a narrow funnel or one-size-fits-all approach.

We can’t predict all the ways that our content will be understood, consumed, and engaged with.

What we can do is be helpful, leverage a strong brand, be transparent, and do everything we can to present users (and other sources) with a complete picture of our products, services, and how we are the right fit (or not) for our website visitors.

Leveraging our moments of visibility to generate quality traffic, but understanding that the bottom of the funnel isn’t a slam dunk to convert, and what it all can mean, can go a long way for engaging and re-engaging bottom of the funnel traffic to get every conversion we deserve.

More Resources:

Featured Image: vittaya pinpan/Shutterstock

https://www.searchenginejournal.com/the-b2b-seo-trap-why-high-intent-visibility-can-still-underperform/551751/

How AI Search Can Drive Sales & Boost Conversions [Webinar] via @sejournal, @brentcsutoras

AI search has completely changed the way customers make decisions. If you're still just tracking data instead of driving sales from AI search, you are missing out.

Join **Bart Góralewicz**, Co-Founder of **ZipTie.dev**, on **September 3, 2025**, for an insightful webinar on how to map customer journeys in AI search and turn those insights into measurable sales.

What you will learn:

Why attend:

Brands that win in AI search are not just watching their metrics. They are understanding how customers discover and decide to buy. This session will give you the tools to drive higher conversions and grow revenue with AI search.

Register now to learn practical strategies you can apply right away. Can't attend live? No problem! Sign up, and we will send you the recording.

https://www.searchenginejournal.com/ai-search-can-drive-sales-and-conversions/553647/

Standby As Google Cannibalizes Itself (While Also Devouring All Of Us) via @sejournal, @Kevin_Indig

In 2023, I wrote about a provocative “what if”: What if Google added an AI chatbot to Search, effectively cannibalizing itself?

Fast-forward to 2025, and it’s no longer hypothetical.

With AI Overviews (AIOs) and AI Mode, Google has not only eaten into its own search product, but it has also taken a big bite out of publishers, too.

Cannibalization is usually framed as a risk. But in the right circumstances, it can be a growth driver-or even a survival tactic.

In today’s Memo, I’m revisiting product cannibalization through a fresh AI-era lens, including:

- What cannibalization really is (and why it’s not always bad).
- Iconic examples from Netflix, Apple, Amazon, Google, and Instagram.
- How Google’s AI shift meets the definition of self-cannibalization — and where it doesn’t.
- The four big marketing implications if your brand cannibalizes itself in the AI-boom landscape (for premium subscribers).

Image Credit: Kevin Indig

Boost your skills with Growth Memo's weekly expert insights. Subscribe for free!

Today's Memo is an updated version of my previous guide on product cannibalization.

Previously, I wrote about how adding an AI chatbot to Google Search would mean that Google would be cannibalizing itself – which only a few companies in history have successfully accomplished.

In this updated Memo, I'll walk you through successful product cannibalization examples while we revisit how Google has cannibalized itself through a refreshed lens.

Because … Google has effectively cannibalized itself with the incorporation of AI Overviews (AIOs) and AI Mode, but they haven't found a way to monetize them yet.

And publishers and brands are suffering as a result.

So who wins here? (Does anyone?) Only time will tell.

Product cannibalization is the replacement of a product with a new one from the same company, typically expressed in sales revenue.

Even though most definitions say that cannibalization occurs when revenue is flat while two products trade market share, there are a number of examples that show revenue can grow as a result of cannibalization.

Product cannibalization, or market cannibalization, is **often seen as something bad – but it can be good or even necessary**.

Let's consider a few examples of product cannibalization you're likely already familiar with:

- Hardware.
- Retail.
- SaaS/Tech.

Hardware companies, for example, need to bring out better and newer chips on a regular basis. The lifecycle of AI training chips is often less than a year because new architectures and higher processing capabilities quickly make the previous generation obsolete.

Right now, chips are the hottest commodity in tech – companies building and training AI models need them in massive quantities, and as soon as the next generation is released, the old one loses most of its value.

As a result, chipmakers are forced to cannibalize their own products, advancing designs to stay competitive not only with

rival manufacturers but also with their own previous breakthroughs.

But there are stark differences between cannibalization in retail and tech.

Retail cannibalization is driven by seasonal changes or consumer trends, while tech product cannibalization is primarily a result of progress.

In fashion, for example, consumers prefer this year's collection over last season's. The new collections cannibalize old ones.

In tech, new technology leads companies to replace old products.

New PlayStation consoles, for example, significantly replace sales from older ones – especially since they're backward compatible with games.

Another example? The growth of the headless content management system (CMS), which increasingly replaces the coupled CMS and pushes content management providers to offer new products and features.

Netflix made several product pivots in its history, but two stand out the most:

1. The switch from DVD rental to streaming, and
2. Subscription-only memberships to ad-supported revenue.

On November 3, 2022, Netflix launched an ad-supported plan for $6.99/month on top of its standard and premium plans. (It has since increased to $7.99/month. See image below.)

During the pandemic, Netflix's subscriber numbers skyrocketed, but they came back to earth like Falcon 9 when Covid receded: Enter the "Basic with ads" subscription that promoted retention.

Image Credit: Kevin Indig

Another challenge for Netflix? Competitors. Lots of them – and with legacy media histories.

Initially, the strategy of creating original content and making the experience seamless across many countries resulted in strong growth.

But when competitors like HBO, Disney, and Paramount launched similar products with original content, growth slowed down.

When Netflix launched the ad-supported plan, only 0.1% of existing users made the switch, **but 9% of new users chose it** (see below).

A look at other streaming platforms suggests the share will

increase over time. Here’s a quick look at percentages of subscribers on ad-supported plans across platforms:

- Hulu has 57%.
- Paramount+ – 44%.
- Peacock – 76%.
- HBO Max – 21%.

Image Credit: Kevin Indig

However, **Netflix’s new plan is not technically considered product cannibalization but partial cannibalization based on price.**

The product is the same, but through the new plan, it’s now accessible to a new customer segment that previously wouldn’t have considered Netflix.

Additionally, it prevents an existing customer segment from churning since the recession shuffles the spending behavior of customer segments.

We can conclude that the **new ad-supported Netflix plan is not the same type of cannibalization as its streaming service**.

In 2007, internet connections became strong enough to open the door to streaming. Netflix was not the first company to provide movie streaming, but the first one to be successful at it.

The company paved the way by incentivizing DVD rental customers to engage online, for example, with rental queues on Netflix's website. But ultimately, the pivot was the result of technological progress.

Another product that saw the light of day for the first time in 2007?

The iPhone.

When it launched, the iPhone had all the features of the iPod and more, making it a case of **full product cannibalization**.

As a result, the share of revenue from the iPod significantly decreased once the iPhone launched (see image below).

Even though you *could* argue it's a "regular case" of market cannibalization when looking at revenue streams from each product, it was a **technological step-change** instead of partial cannibalization based on pricing.

Image Credit: Kevin Indig

However, big steps in technology don’t always lead to a desired replacement of an old product.

Take the Amazon Kindle, for example.

Launched in 2007 – just like Netflix’s streaming product and the iPhone (something was up that year) – Amazon brought its new ebook reader, Kindle, to market.

It made such an impact that people predicted the death of paper books. (And librarians everywhere laughed while booksellers braced themselves.)

But over 10 years later, ebooks stabilized at 20% market share, while print books captured 80%.

The reason is that publishers got into pricing battles with Amazon and Apple, which also started to play an important role in the ebook market. (It’s a long story; but you can read about it here).

Amazon attempted to cannibalize its core business (books) with the Kindle (ebooks), but couldn’t make product pricing work, which resulted in ebooks often being more expensive than print editions. Yikes.

The technology changed, but consumers weren’t incentivized to use it.

Let’s look at two final examples here. These two companies acquired or copied competitors to control partial cannibalization:

- **YouTube** videos are technically better answers to many search queries than web results. Google saw this very early on and smartly acquired YouTube. Video results took some time to fill more space in the Google SERP, even though they technically cannibalize web results. But today, they’re often some of the most visually impactful results (and often the first results) that searchers see.
- **Instagram** saw the success of Snapchat stories and decided to copy the feature in order to mitigate competitor growth. Despite the cannibalization of regular Instagram posts, net engagement with Stories was greater. (And speaking of YouTube, you could argue that YouTube Shorts follow the same principle.)

With all this in mind, we can say there is **full and partial cannibalization based on how many features a new product replaces**.

Pricing changes, copied features, or acquisitions lead to partial cannibalization that doesn’t result in the same revenue growth as full cannibalization.

Full cannibalization requires two conditions to be true:

1. The new product must be built on a technological step change, and
2. Customers need to be incentivized to use it.

With this knowledge foundation in place, let’s examine the shifts in the Google Search product over the last 12-24 months.

Let’s apply these product cannibalization principles to the case of Google vs. ChatGPT & Co.

In the original content of this memo (published in 2023), I shared the following:

> *If Google were to add AI to Search in a similar way as Neeva & Co (see previous article about **Early attempts at integrating AI in Search**), the following conditions would be true:*
>
> 1. *AI Chatbots are a technological step-change.*
> 2. *Customers are incentivized to use AI Chatbots because they give quick and good answers to most questions.*
>
> *However, not all conditions are true:*
>
> 1. *AI Chatbots don't provide the full functionality of Google Search.*
> 2. *It's not cheaper to integrate an AI Chatbot with Search.*

I've been clear about my hypothesis for a while now. As I shared in my 2025 Halftime Report:

> *I personally believe that AI Mode won't launch [fully in the SERP] before Google has figured out the monetization model. And I predict that searchers will see way fewer ads but much better ones and displayed at a better time.*

And I highlighted this in Is AI cutting into your conversions? also:

> *Google won't show AI Mode everywhere, because adoption is generational (see the **UX study of AIOs** for more info). I think AI Mode will launch at a broader scale (like showing up for more queries overall) when Google figures out monetization.*
>
> *Plus, ChatGPT is not yet monetizing, so advertisers go to Google and Meta – for now. And that's my hypothesis as to why*

Google Search is continuing to grow.

Keep in mind, to successfully cannibalize your existing product, you need customers to want to use it. And according to a recent report from Garrett Sussman over at iPullrank, over 50% of users who tried Google's AI Mode once and didn't return. [Source] (So it seems Google's still figuring out the incentivising part.)

Even with the advancements we've seen in the last six to 12 months with AI models – and the inclusion of live web search and product recommendations into AI chats – I'd argue that they're useful for information-driven or generative queries but lack the databases needed to give good answers to product or service searches.

Let's take a look at an example:

If you input "best plumber in Chicago" or "best toaster" into ChatGPT, I'd argue you'd actually get less quality results – for now – than if you input the same queries into Google. (Go try it for yourself and let me know what you find. But here's a walk-through with Amanda Johnson hopping in to illustrate this below.)

At the same time, these product and service queries are the queries that search engines with an ad revenue business model can monetize best.

It was said that ChatGPT costs at least $100,000 per day to run when it first crossed 1 million users in December 2022. By 2023, it was costing about $700,000 a day. [Source]

Today, it's likely to be a significant multiple of that.

Keep in mind, Google Search sees billions of search queries every day.

Even with Google's advanced infrastructure and talent, **AI**

chatbots are costly.

And they can (still) be slow – even with the advancements they've made in the last 12 months. Current and classic Google Search systems (like Featured Snippets and People Also Ask) might provide a much faster answer.

But, alas, here we are in 2025, and Google is cannibalizing its own product via AIOs and AI Mode.

Right now, according to Similarweb data, usage of the AI Mode tab on Google.com in the U.S. has slightly dipped and now sits at just over 1%. [Source, Source]

Google AIOs are now seen by more than 1.5 billion searchers every month, and they sit front and center. But engagement is falling. Users are spending less time on Google and clicking fewer pages. [Source]

But Google *has* to compete with not only other search engines that provide an AI-chat-forward experience, but also with ChatGPT & Co. themselves.

Below, I've listed out important considerations for your brand if you might utilize product cannibalization as a strategy.

You'll want to:

1. Reframe cannibalization as a strategic option for the brand rather than a failure.
2. Use the full vs. partial cannibalization lens.
3. Test the two success conditions.
4. Protect your core offerings while you experiment.
5. Use competitive cannibalization defensively.
6. Monitor, learn, and adjust.

In the next section, for premium subscribers, I'll walk you through what to watch out for if you decide to use product cannibalization as a growth strategy.

1. Reframe Cannibalization As A Strategic Option

- Don’t default to seeing product cannibalization as a failure; assess if it can protect market share or accelerate growth.
- Audit your product line and GTM strategy to identify areas where you could self-disrupt before a competitor does.

2. Use The Full Vs. Partial Cannibalization Lens

- **Full cannibalization** works best when there’s a tech leap and strong customer incentives.
  - **Example:** Apple iPhone replacing iPod – all iPod features plus far more capability led to the iPod’s rapid decline.
- **Partial cannibalization** via pricing, features, or acquisitions is less risky but may not deliver big growth.
  - **Example:** Netflix ad-supported plan – same streaming product, but a lower-cost tier opened the door to new segments and reduced churn risk.
- Map current and future offerings against these two categories to decide your approach.

3. Test The 2 Success Conditions

A cannibalizing product is more likely to succeed when both are true:

- **Tech Leap:** Offers a meaningfully better way to solve the problem.
  - **Example:** Netflix DVD → Streaming in 2007 leveraged faster internet speeds to change the delivery model entirely.

- **Customer Incentive:** Lower cost, better performance, more convenience, or status.
    - **Example:** YouTube acquisition by Google made richer, more visual answers possible in Search, improving the user experience.

If both apply → pursue full cannibalization.

If one applies → pursue partial cannibalization with controlled scope.

4. Protect Your Core While You Experiment

- Identify high-revenue segments and shield them from early disruption.
    - **Example:** Google keeping AI Mode away from highly monetizable queries like “best credit card” until the ad model is ready.
- Test self-disruption in lower-stakes markets to validate demand before scaling
    - **Example:** Instagram Stories rolled out in a way that boosted net engagement while protecting the feed’s ad inventory.

5. Use Competitive Cannibalization Defensively

- When a rival launches a threat, choose between:
    - **Acquire:** Google acquiring YouTube to control the rise of video as a search answer format.
    - **Copy:** Instagram adopting Stories from Snapchat to stop user migration and grow engagement.
    - **Differentiate:** Amazon Kindle – a tech leap that tried to move readers from print to digital, but without a price advantage, adoption plateaued.

6. Monitor, Learn, And Adjust

- Track engagement, revenue mix, and adoption by segment.
    - **Example:** Similarweb data on AI Mode — U.S. usage holding just over 1%, signaling limits to adoption speed.
- Adjust rollout pace based on generational adoption patterns and competitor moves.
    - **Example:** Google AIO engagement drop — showing that placement alone doesn't guarantee sustained user interest.

A good example is how to do this is Chegg.

The company has been obliterated by Google's AI Overviews and even sued Google. Chegg's value was answers to homework questions, but since almost every student uses ChatGPT for that, their value chain broke. How is the company reacting to this life-ending threat?

In my Q2 Marketplace Deep Dive, I explain that Chegg has found a lifeboat:

> *Chegg has launched a new tool, Solution Scout, that allows students to compare answers from ChatGPT & Co. with Chegg's archive.*
>
> *Instead of trying to beat AI Chatbots, Chegg hits them where it hurts: in the hallucinations.*
>
> *LLMs can make stuff up, which is especially painful when it comes to learning and taking tests. Imagine you spend hours internalizing the wrong facts!*
>
> *Solution Scout validates AI answers with Chegg's archive of human-sourced material. It compares the answer from foundational models and highlights differences and consensus.*

Featured Image: Paulo Bobita/Search Engine Journal

https://www.searchenginejournal.com/standby-as-google-cannibalizes-itself-while-also-devouring-all-of-us/554088/

Are Google's AI Travel Results Uncovering More Hidden Gems? [Data Study] via @sejournal, @TaylorDanRW

Google's AI-generated results reshape how people search, and Google has said that websites should expect traffic fluctuations and that prior success in organic Search does not guarantee future success in the new ecosystem.

This is a big claim, and it's been debated whether "Hidden Gems" are getting more visibility in modern Search and I'm doing my best to work through as much data as possible to identify if the claims from Google above have substance.

Google's Hidden Gems initiative is its effort to highlight genuine, first-hand content from smaller corners of the web.

It was first revealed in May 2023 and fully integrated into the core algorithm later that year, with official acknowledgment in mid-November 2023.

It targets posts with first-hand knowledge, personal insights, and unique experiences usually found on forums, blogs, social platforms, and niche sites.

Rather than favoring only high-authority domains, it now surfaces these overlooked “gems” because they offer genuine and practical perspectives from creators and brands, not powered by the traditional SEO metrics and big brand budgets.

Hidden Gems has the objective of:

> *Improving how we (Google) rank results in Search overall, with a greater focus on content with unique expertise and experience.*

This brings me to the travel sector and the notion of Hidden Gems.

It has been a long-held belief in the travel sector that Google favors bigger travel brands. When I worked in regional agencies that had travel clients, this was almost a party line when pitching SME and challenger travel websites.

Now search is evolving, and we’re seeing more and more Search features either powered by or directly interfacing with AI, is this now an opportunity for challenger travel brands to gain further visibility within Google’s Search ecosystem?

Methodology

To investigate, we analyzed a dataset of 5,824 URLs surfaced in Google’s AI-generated results for travel-related queries.

As part of the methodology, we also reviewed traditional SEO metrics such as estimated site traffic, domain rating, and total number of domain keywords to validate a qualitative review of whether a site functions as a powerful travel brand or a challenger brand.

Each URL was manually reviewed and tagged based on whether Google identified it as a Hidden Gem. We compared their visibility, domain authority, and how often they appeared in

AI results.

Quantity Vs. Frequency

The dataset revealed a nuanced dynamic: While Hidden Gems were more diverse, they were not more dominant.

From the 5,824 cited URLs, we identified 1,371 unique domains. We classified 590 of these as Hidden Gem domains compared to 781 established domains.

However, those 781 established domains appeared 4,576 times in total, a much higher return rate than the 1,248 total appearances of the Hidden Gems.

This suggests that while AI mode is surfacing a wide variety of lesser-known sources, it still leans heavily on established brands for repeated visibility.

As you would expect, domains we identified as not being “Hidden Gems” had a greater weighting of higher DR than not.

Image from author, August 2025

By contrast, the domains we identified as being Hidden Gems were not weighted in the opposite direction, but instead much more evenly spread out.

Image from author, August 2025

In other words, Google is sampling widely from the long tail but serving frequently from the head of the distribution.

Authority Still Has A Role

While traditional SEO has long placed emphasis on authority metrics like Domain Rating (DR) or Domain Authority (DA), our analysis shows that their influence may be diminishing in the context of AI-led search.

This shift aligns with broader trends we've observed in Google's evolving ranking systems.

Instead of relying heavily on link-based authority, AI

Overviews and similar experiences appear to prioritize content that demonstrates depth, originality, and strong alignment with user intent.

Authority hasn’t disappeared, but it’s been repositioned. Rather than acting as a gatekeeper for visibility, it’s now one of many factors, often taking a back seat to how well a piece of content anticipates and satisfies the user’s informational needs in the moment.

What This Means For Travel Brands

Hidden Gems are showing up in Google’s AI results, but they’re not displacing the giants. They’re appearing alongside them, offering more variety but less dominance.

For challenger brands, this represents both an opportunity and a challenge.

First-Hand Content Gains Ground

The opportunity is clear: Content that is specific, genuine, and useful is getting noticed, even from smaller or lesser-known sites.

AI-powered results seem to be more willing to include pages that deliver practical insights, first-hand experience, and niche relevance, even if they lack the traditional signals of authority.

This creates new openings for brands that previously struggled to compete on backlinks or brand strength alone.

Repetition And Recall Still Matter

But the challenge is equally clear in that visibility is not evenly distributed.

While Google may sample from a broader range of sources, the

repetition and prominence still favor the dominant travel brands.

These brands appear more frequently, benefit from greater brand recall, and are more likely to be clicked simply because they’re familiar.

So for newer or challenger brands, the question becomes: How do you turn presence into preference?

Where Should I Be Focusing?

Consistency Of Presence

It starts with consistency. One or two appearances in AI Overviews won’t move the needle.

Travel brands need to think about sustained visibility, showing up across a range of topics, formats, and moments in the user journey.

That means building out content that doesn’t just answer common queries but anticipates nuanced needs, inspires curiosity, and offers unique, first-hand insight.

Clarity Of Voice

Next comes clarity of voice. AI systems are increasingly sensitive to content that signals credibility, experience, and originality.

Brands that find and articulate a clear editorial voice, whether that’s luxury travel with a local twist, slow travel for sustainability, or adventure itineraries from people who’ve actually been there, are more likely to stand out.

Intent Understanding

Finally, there’s an intent understanding. Challenger brands must shift from thinking in keywords to thinking in moments.

What’s the user trying to imagine, plan, solve, or feel at this point in their journey? How can your content speak directly to that?

A New Definition Of Authority

The travel giants still have scale on their side, but challenger brands now have a better chance to earn visibility through authenticity and depth. That’s a different kind of authority, one rooted in relevance and resonance.

For travel SEOs willing to rethink what authority means, and for brands ready to invest in meaningful, user-first content, AI-powered search is no longer just a threat. It’s an invitation.

Not to play the same game the giants are playing, but to play a different one, and win in different ways.

More Resources:

Featured Image: SvetaZi/Shutterstock

https://www.searchenginejournal.com/are-googles-ai-travel-results-uncovering-more-hidden-gems/553132/

Tired Of SEO Spam, Software Engineer Creates A New Search

Engine via @sejournal, @martinibuster

A software engineer from New York got so fed up with the irrelevant results and SEO spam in search engines that he decided to create a better one. Two months later, he has a demo search engine up and running. Here is how he did it, and four important insights about what he feels are the hurdles to creating a high-quality search engine.

One of the motives for creating a new search engine was the perception that mainstream search engines contained increasing amount of SEO spam. After two months the software engineer wrote about their creation:

> *"What's great is the comparable lack of SEO spam."*

Neural Embeddings

The software engineer, Wilson Lin, decided that the best approach would be neural embeddings. He created a small-scale test to validate the approach and noted that the embeddings approach was successful.

Chunking Content

The next phase was how to process the data, like should it be divided into blocks of paragraphs or sentences? He decided that the sentence level was the most granular level that made sense because it enabled identifying the most relevant answer within a sentence while also enabling the creation of larger paragraph-level embedding units for context and semantic

coherence.

But he still had problems with identifying context with indirect references that used words like “it” or “the” so he took an additional step in order to be able to better understand context:

> *“I trained a DistilBERT classifier model that would take a sentence and the preceding sentences, and label which one (if any) it depends upon in order to retain meaning. Therefore, when embedding a statement, I would follow the “chain” backwards to ensure all dependents were also provided in context.*
>
> *This also had the benefit of labelling sentences that should never be matched, because they were not “leaf” sentences by themselves.”*

Identifying The Main Content

A challenge for crawling was developing a way to ignore the non-content parts of a web page in order to index what Google calls the Main Content (MC). What made this challenging was the fact that all websites use different markup to signal the parts of a web page, and although he didn’t mention it, not all websites use semantic HTML, which would make it vastly easier for crawlers to identify where the main content is.

So he basically relied on HTML tags like the paragraph tag <p> to identify which parts of a web page contained the content and which parts did not.

This is the list of HTML tags he relied on to identify the main content:

- blockquote – A quotation
- dl – A description list (a list of descriptions or definitions)

- ol – An ordered list (like a numbered list)
- p – Paragraph element
- pre – Preformatted text
- table – The element for tabular data
- ul – An unordered list (like bullet points)

Issues With Crawling

Crawling was another part that came with a multitude of problems to solve. For example, he discovered, to his surprise, that DNS resolution was a fairly frequent point of failure. The type of URL was another issue, where he had to block any URL from crawling that was not using the HTTPS protocol.

These were some of the challenges:

> *"They must have https: protocol, not ftp:, data:, javascript:, etc.*
>
> *They must have a valid eTLD and hostname, and can't have ports, usernames, or passwords.*
>
> *Canonicalization is done to deduplicate. All components are percent-decoded then re-encoded with a minimal consistent charset. Query parameters are dropped or sorted. Origins are lowercased.*
>
> *Some URLs are extremely long, and you can run into rare limits like HTTP headers and database index page sizes.*
>
> *Some URLs also have strange characters that you wouldn't think would be in a URL, but will get rejected downstream by systems like PostgreSQL and SQS."*

Storage

At first, Wilson chose Oracle Cloud because of the low cost of transferring data out (egress costs).

He explained:

> *"I initially chose Oracle Cloud for infra needs due to their very low egress costs with 10 TB free per month. As I'd store terabytes of data, this was a good reassurance that if I ever needed to move or export data (e.g. processing, backups), I wouldn't have a hole in my wallet. Their compute was also far cheaper than other clouds, while still being a reliable major provider."*

But the Oracle Cloud solution ran into scaling issues. So he moved the project over to PostgreSQL, experienced a different set of technical issues, and eventually landed on RocksDB, which worked well.

He explained:

> *"I opted for a fixed set of 64 RocksDB shards, which simplified operations and client routing, while providing enough distribution capacity for the foreseeable future.*
>
> *…At its peak, this system could ingest 200K writes per second across thousands of clients (crawlers, parsers, vectorizers). Each web page not only consisted of raw source HTML, but also normalized data, contextualized chunks, hundreds of high dimensional embeddings, and lots of metadata."*

GPU

Wilson used GPU-powered inference to generate semantic vector embeddings from crawled web content using transformer models. He initially used OpenAI embeddings via API, but that became

expensive as the project scaled. He then switched to a self-hosted inference solution using GPUs from a company called Runpod.

He explained:

> *"In search of the most cost effective scalable solution, I discovered Runpod, who offer high performance-per-dollar GPUs like the RTX 4090 at far cheaper per-hour rates than AWS and Lambda. These were operated from tier 3 DCs with stable fast networking and lots of reliable compute capacity."*

Lack Of SEO Spam

The software engineer claimed that his search engine had less search spam and used the example of the query "best programming blogs" to illustrate his point. He also pointed out that his search engine could understand complex queries and gave the example of inputting an entire paragraph of content and discovering interesting articles about the topics in the paragraph.

Four Takeaways

Wilson listed many discoveries, but here are four that may be of interest to digital marketers and publishers interested in this journey of creating a search engine:

1. The Size Of The Index Is Important

One of the most important takeaways Wilson learned from two months of building a search engine is that the size of the search index is important because in his words, "coverage defines quality." This is

2. Crawling And Filtering Are Hardest Problems

Although crawling as much content as possible is important for surfacing useful content, Wilson also learned that filtering low quality content was difficult because it required balancing the need for quantity against the pointlessness of crawling a seemingly endless web of useless or junk content. He discovered that a way of filtering out the useless content was necessary.

This is actually the problem that Sergey Brin and Larry Page solved with Page Rank. Page Rank modeled user behavior, the choice and votes of humans who validate web pages with links. Although Page Rank is nearly 30 years old, the underlying intuition remains so relevant today that the AI search engine Perplexity uses a modified version of it for its own search engine.

3. Limitations Of Small-Scale Search Engines

Another takeaway he discovered is that there are limits to how successful a small independent search engine can be. Wilson cited the inability to crawl the entire web as a constraint which creates coverage gaps.

4. Judging trust and authenticity at scale is complex

Automatically determining originality, accuracy, and quality across unstructured data is non-trivial

Wilson writes:

> *"Determining authenticity, trust, originality, accuracy, and quality automatically is not trivial. …if I started over I*

> *would put more emphasis on researching and developing this aspect first.*
>
> *Infamously, search engines use thousands of signals on ranking and filtering pages, but I believe newer transformer-based approaches towards content evaluation and link analysis should be simpler, cost effective, and more accurate.”*

Interested in trying the search engine? You can find it here and you can read how the full technical details of how he did it here.

Featured Image by Shutterstock/Red Vector

https://www.searchenginejournal.com/tired-of-seo-spam-software-engineer-creates-a-new-search-engine/553994/