Perplexity Says Cloudflare Is Blocking Legitimate AI Assistants via @sejournal, @martinibuster

Perplexity published a response to Cloudflare's claims that it disrespects robots.txt and engages in stealth crawling. Perplexity argues that Cloudflare is mischaracterizing AI Assistants as web crawlers, saying that they should not be subject to the same restrictions since they are user-initiated assistants.

Perplexity AI Assistants Fetch On Demand

According to Perplexity, its system does not store or index content ahead of time. Instead, it fetches webpages only in response to specific user questions. For example, when a user asks for recent restaurant reviews, the assistant retrieves and summarizes relevant content on demand. This, the company says, contrasts with how traditional crawlers operate, systematically indexing vast portions of the web without regard to immediate user intent.

Perplexity compared this on-demand fetching to Google's user-triggered fetches. Although that is not an apples-to-apples comparison because Google's user-triggered fetches are in the service of reading text aloud or site verification, it's still an example of user-triggered fetching that bypasses robots.txt restrictions.

In the same way, Perplexity argues that its AI operates as an extension of a user's request, not as an autonomous bot crawling indiscriminately. The company states that it does not retain or use the fetched content for training its models.

Criticizes Cloudflare's Infrastructure

Perplexity also criticized Cloudflare's infrastructure for failing to distinguish between malicious scraping and legitimate, user-initiated traffic, suggesting that Cloudflare's approach to bot management risks overblocking services that are acting responsibly. Perplexity argues that a platform's inability to differentiate between helpful AI assistants and harmful bots causes misclassification of legitimate web traffic.

Perplexity makes a strong case for the claim that Cloudflare is blocking legitimate bot traffic and says that Cloudflare's decision to block its traffic was based on a misunderstanding of how its technology works.

Read Perplexity's response:

Agents or Bots? Making Sense of AI on the Open Web

https://www.searchenginejournal.com/perplexity-says-cloudflare-is-blocking-legitimate-ai-assistants/552927/

Cloudflare Delists And Blocks

Perplexity From Crawling Websites via @sejournal, @martinibuster

Cloudflare announced that they delisted Perplexity's crawler as a verified bot and are now actively blocking Perplexity and all of its stealth bots from crawling websites. Cloudflare acted in response to multiple user complaints against Perplexity related to violations of robots.txt protocols, and a subsequent investigation revealed that Perplexity was using aggressive rogue bot tactics to force its crawlers onto websites.

Cloudflare Verified Bots Program

Cloudflare has a system called Verified Bots that whitelists bots in their system, allowing them to crawl the websites that are protected by Cloudflare. Verified bots must conform to specific policies, such as obeying the robots.txt protocols, in order to maintain their privileged status within Cloudflare's system.

Perplexity was found to be violating Cloudflare's requirements that bots abide by the robots.txt protocol and refrain from using IP addresses that are not declared as belonging to the crawling service.

Cloudflare Accuses Perplexity Of Using Stealth Crawling

Cloudflare observed various activities indicative of highly

aggressive crawling, with the intent of circumventing the robots.txt protocol.

Stealth Crawling Behavior: Rotating IP Addresses

Perplexity circumvents blocks by using rotating IP addresses, changing ASNs, and impersonating browsers like Chrome.

Perplexity has a list of official IP addresses that crawl from a specific ASN (Autonomous System Number). These IP addresses help identify legitimate crawlers from Perplexity.

An ASN is part of the Internet networking system that provides a unique identifying number for a group of IP addresses. For example, users who access the Internet via an ISP do so with a specific IP address that belongs to an ASN assigned to that ISP.

When blocked, Perplexity attempted to evade the restriction by switching to different IP addresses that are not listed as official Perplexity IPs, including entirely different ones that belonged to a different ASN.

Stealth Crawling Behavior: Spoofed User Agent

The other sneaky behavior that Cloudflare identified was that Perplexity changed its user agent in order to circumvent attempts to block its crawler via robots.txt.

For example, Perplexity's bots are identified with the following user agents:

- PerplexityBot
- Perplexity-User

Cloudflare observed that Perplexity responded to user agent blocks by using a different user agent that posed as a person

crawling with Chrome 124 on a Mac system. That's a practice called spoofing, where a rogue crawler identifies itself as a legitimate browser.

According to Cloudflare, Perplexity used the following stealth user agent:

> *"Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/124.0.0.0 Safari/537.36"*

Cloudflare Delists Perplexity

Cloudflare announced that Perplexity is delisted as a verified bot and that they will be blocked:

> *"The Internet as we have known it for the past three decades is rapidly changing, but one thing remains constant: it is built on trust. There are clear preferences that crawlers should be transparent, serve a clear purpose, perform a specific activity, and, most importantly, follow website directives and preferences. Based on Perplexity's observed behavior, which is incompatible with those preferences, we have de-listed them as a verified bot and added heuristics to our managed rules that block this stealth crawling."*

Takeaways

- **Violation Of Cloudflare's Verified Bots Policy**
  Perplexity violated Cloudflare's Verified Bots policy, which grants crawling access to trusted bots that follow common-sense rules like honoring the robots.txt protocol.
- **Perplexity Used Stealth Crawling Tactics**
  Perplexity used undeclared IP addresses from different ASNs and spoofed user agents to crawl content after

being blocked from accessing it.

- **User Agent Spoofing**
  Perplexity disguised its bot as a human user by posing as Chrome on a Mac operating system in attempts to bypass filters that block known crawlers.
- **Cloudflare's Response**
  Cloudflare delisted Perplexity as a Verified Bot and implemented new blocking rules to prevent the stealth crawling.
- **SEO Implications**
  Cloudflare users who want Perplexity to crawl their sites may wish to check if Cloudflare is blocking the Perplexity crawlers, and, if so, enable crawling via their Cloudflare dashboard.

Cloudflare delisted Perplexity as a Verified Bot after discovering that it repeatedly violated the Verified Bots policies by disobeying robots.txt. To evade detection, Perplexity also rotated IPs, changed ASNs, and spoofed its user agent to appear as a human browser. Cloudflare's decision to block the bot is a strong response to aggressive bot behavior on the part of Perplexity.

https://www.searchenginejournal.com/cloudflare-delists-and-blocks-perplexity-from-crawling-websites/552899/

How AI Search Should Be Shaping Your CEO's & CMO's Strategy [Webinar] via

@sejournal, @theshelleywalsh

AI is rapidly changing the rules of SEO. From generative ranking to vector search, the new rules are not only technical but also reshaping how business leaders make decisions.

Join **Dan Taylor** on **August 14, 2025**, for an exclusive **SEJ Webinar** tailored for C-suite executives and senior leaders. In this session, you'll gain essential insights to understand and communicate SEO performance in the age of AI.

Here's what you'll learn:

AI Search Is Impacting Everything. Are You Ready?

AI search is already here, and it's impacting everything from SEO KPIs to customer journeys. This webinar will give you the tools to lead your teams through the shift with confidence and precision.

Register now for a business-first perspective on AI search innovation. If you can't attend live, don't worry. Sign up anyway, and we'll send you the full recording.

https://www.searchenginejournal.com/ai-search-shaping-your-strategy/551490/

Which SEO Jobs AI Will Reshape & Which Might Disappear via @sejournal, @DuaneForrester

You've probably seen the headlines like: "AI will kill SEO," "AI will replace marketing roles," or the latest panic: "Is your digital marketing job safe?"

Well, maybe not those exact headlines, but you get the idea, and I'm sure you have seen something similar.

Let's clear something up: AI is not making SEO irrelevant. It's making certain tasks obsolete. And yes, some jobs built entirely around those tasks are at risk.

A recent Microsoft study analyzed over 200,000 Bing Copilot interactions to measure task overlap between human job functions and AI-generated outputs. Their findings are eye-opening:

- **Translators and Interpreters:** 98% overlap with AI tasks.
- **Writers and Authors:** 88% overlap.
- **Public Relations Specialists:** 79% overlap.

SEO as a field wasn't directly named in the study, but many roles common within SEO map tightly to these job categories.

If you write, edit, report, research, or publish content as part of your daily work, this isn't a hypothetical shift. It's already happening.

(Source: Microsoft AI Job Impact – Business Insider – follow through this link to reach the download location for the original PDF of the study. BI summarizes the information, but links to MSFT, which in turn links to the source for the PDF.)

What’s Actually Changing

AI isn’t replacing SEO. It’s changing what “search engine optimization” means, and where and how value is measured.

In traditional SEO, the focus was clear:

- Rank high.
- Earn the click.
- Optimize the page for humans and crawlers.

That still matters. But, in AI-powered search systems, the sequence is different:

1. **Content is chunked** behind the scenes, paragraphs, lists, and answers are sliced and stored in vector form.
2. **Prompts trigger retrieval**, the LLM pulls relevant chunks, often based on embeddings, not just keywords. (So, concepts and relationships, not keywords per se.)
3. **Only a few chunks make it into the answer**. Everything else is invisible, no matter how high it once ranked.

This new paradigm shifts the rules of engagement. Instead of asking, “Where do I rank?” the better question is, “Was my content even retrieved?” That makes this a binary system, not a sliding scale.

In this new world of retrieval, the direct answer to the question, “Where do I rank?” could be “ChatGPT,” “Perplexity,” “Claude,” or “CoPilot,” instead of a numbered position.

In some ways, this isn’t as big a shift as some folks would have you believe. After all, as the old joke asks, “Where do you hide a dead body?” To which the correct answer is “…on Page 2 of Google’s results!”

Morbid humor aside, the implication is no one goes there, so there’s no value, and while that sentiment actually drops a lot of the real, nuanced details that actual click through

rate data shows us (like the top of page 2 results actually has better CTRs than the bottom of page 1 typically), it does serve up a meta point: If you're not in the first few results on a traditional SERP, the drop off of CTRs is precipitous.

So, it could be argued that with most "answers" today in generative AI systems being comprised of a very limited set of references, that today's AI-based systems offer a new display path for consumers, but ultimately, those consumers will only be interacting with the same number of results they historically engaged with.

I mean, if we only ever really clicked on the top 3 results (generalizing here), and the rest were surplus to needs, then cutting an AI-sourced answer down to some words with only 1, 2 or 3 cited results amounts to a similar situation in terms of raw numbers of choice for consumers … 1, 2 or 3 clickable options.

Regardless, it does mark a shift in terms of work items and workflows, and here's how that shift shows up across some core SEO tasks. Obviously, there could be many more, but these examples help set the stage:

- **Keyword research becomes embedding relevance and semantic overlap.** It's not about the exact phrase match in a gen AI result. It's about aligning your language with the concepts AI understands. It's about the concept of query fan-out (not new, by the way, but very important now).
- **Meta tag and title optimization become chunked headers and contextual anchor phrases.** AI looks for cues inside content to determine chunk focus.
- **Backlink building becomes trust signal embedding and source transparency.** Instead of counting links, AI asks: Does this source feel credible and citable?
- **Traffic analytics becomes retrieval testing and AI response monitoring.** The question isn't just how many

visits you got, it’s whether your content shows up at all in AI-generated responses.

What this means for teams:

- Your title tag isn’t just a headline; it’s a semantic hook for AI retrieval.
- Content format matters more: bullets, tables, lists, and schema win because they’re easier to cite.
- You need to test with prompts to see if your content is actually getting surfaced.

None of this invalidates traditional SEO. But, the visibility layer is moving. If you’re not optimizing for retrieval, you’re missing the first filter, and ranking doesn’t matter if you’re never in the response set.

The SEO Job Risk Spectrum

Microsoft’s study didn’t target SEO directly, but it mapped 20+ job types by their overlap with current AI tasks. I used those official categories to extrapolate risk within SEO job functions.

Image Credit: Duane Forrester

High Risk — Immediate Change Needed

SEO Content Writers

Mapped to: Writers & Authors (88% task overlap in the study: 88% of these tasks an AI can do today).

Why: These roles often involve creating repeatable, factual content, precisely the kind of output AI handles well today (to a degree, anyway). Think meta descriptions, product overviews, and FAQ pages.

The writing isn’t disappearing, but humans aren’t always required for first drafts anymore. Final drafts, yes, but

first? No. And I’m not debating how factual the content is that an AI produces.

We all know the pitfalls, but I’ll say this: If your boss is telling you your job is going away, and your argument is “but AIs hallucinate,” think about whether that’s going to change the outcome of that meeting.

Link Builders/Outreach Specialists

Mapped to: Public Relations Specialists (79% overlap).

Why: Cold outreach and templated link negotiation can now be automated.

AI can scan for unlinked mentions, generate outreach messages, and monitor link placement outcomes, cutting into the core responsibilities of these roles.

Moderate Risk — Upskill To Stay Relevant

SEO Analysts

Mapped to: Market Research Analysts (~65% overlap).

Why: Data gathering and trend reporting are susceptible to automation. But, analysts who move into interpreting retrieval patterns, building AI visibility reports, or designing retrieval experiments can thrive.

Admittedly, SEO is a bit more specialized, but bottom or top of this stack, the risk remains moderate. This one, however, is heavily dependent on your actual job tasks.

Technical SEOs

Mapped to: Web Developers (not perfect, but as close as the study got).

Why: Less overlap with generative AI, but still pressured to evolve. Embedding hygiene, chunk structuring, and schema

precision are now foundational.

The most valuable technical SEOs are becoming AI optimization architects. Not leaving their traditional work behind, but adopting new workflows.

Content Strategists/Editors

Mapped to: Editors & Technical Writers.

Why: Editing for humans and tone alone is out. Editing for retrievability is in. Strategists now must prioritize chunking, citation density, and clarity of topic anchors, not just user readability.

Or, at least, now consider that LLM bots are de facto users as well.

Lower Risk – Expanded Value And Influence

SEO Managers/Leads

Mapped to: Marketing Managers.

Why: Managers who understand both traditional and AI SEO have more leverage than ever. They're responsible for team alignment, training decisions, and tool adoption.

This is a growth role, if guided by data, not gut instinct. Testing is life here.

CMOs/Strategy Executives

Mapped to: Marketing Executives.

Why: Strategic thinking isn't automatable. AI can suggest, but it can't set priorities across brand, trust, and investment.

Executives who understand how AI affects visibility will steer their companies more effectively, especially in content-heavy verticals.

Tactical Response By Role Type

Every job category on the risk curve deserves practical action.

Now, let’s look at how people in SEO roles can pivot, strengthen, or evolve, based on clear, verifiable capabilities.

High-Risk Roles: SEO Content Writers, Editors, Link Builders

- Shift from traditional copywriting to creating structured, retrieval-friendly content.
- Focus on chunk-based writing: short Q&A blocks, bullet-based explanations, and schema-rich snippets.
- Learn AI prompt testing: Use platforms like ChatGPT or Google Gemini to query key topics and see if your content is surfaced without requiring a click.
- Use gen AI visibility tools verified to support AI search tracking:
    - Profound tracks your brand’s appearance in AI search results across platforms like ChatGPT, Perplexity, and Google Overviews. You can see where you’re cited and which topics AI engines associate with you.
    - SERPRecon offers AI-powered content outlines and helps reverse-engineer AI overview logic to show what keywords and phrasing matter most. So, use a tool like this, then take the output as the basis for your query fan-out work.
- Reinvent your role:
    - Write in chunks that AI can cite.
    - Embed trust signals (clear sourcing, authoritativeness).
    - Collaborate with data teams on embedding accuracy and chunk performance.

Moderate-Risk Roles: SEO Analysts, Technical SEOs, Content Strategists

- Expand traditional ranking reports with retrievability diagnostics:
    - Use prompt simulations that probe content retrieval in real-time across AI engines.
    - Audit embedding and semantic alignment at the paragraph or chunk level.
- Employ tools like those mentioned to analyze AI Overviews and generate content improvement outlines.
- Monitor AI visibility gaps through new dashboards:
    - Track citation share versus competitors.
    - Identify topic clusters where your domain is cited less.
- Understand structured data and schema:
    - Use markup to clearly define entities, relationships, and context for AI systems.
    - Prioritize formats like FAQPage, HowTo, and Product schema, where applicable. These are easier for LLMs and AI Overviews to cite.
    - Align semantic clarity within chunks to schema-defined roles (e.g., question/answer pairs, step lists) to improve retrievability and surface relevance.
- Join or lead internal “AI-SEO Workshops”:
    - Teach teams how to test content visibility in ChatGPT, Perplexity, or Google Overviews.
    - Share experiments in prompt engineering, chunk format outcomes, and schema effectiveness.

Lower-Risk Roles: SEO Managers, Digital Leads, CMOs

- Sponsor retraining initiatives for semantic and vector-led SEO practices.

- Revise hiring briefs and job descriptions to include skills like embedding knowledge, prompt testing, schema fluency, and chunk analysis.
- Implement AI-visibility dashboards using dedicated tools:
    - Benchmark brand presence across search engines and generative platforms.
    - Use insights to guide future content and authority decisions.
- Keep traditional SEO strong alongside AI tactics:
    - Technical optimization, speed, quality of content, etc., still matter.
    - Hybrid success requires both sides working in sync.
- Set internal AI literacy standards:
    - Offer training on retrieval engineering, LLM behavior, and chunk visibility.
    - Ensure everyone understands AI's core behaviors, what it cites, and what it ignores.

Reframing The Opportunity

This isn't a "get out now" scenario for these jobs. It's a "rebuild your toolkit" moment.

High overlap doesn't mean you're obsolete. It means the old version of your job won't hold value without adaptation. And what gets automated away often wasn't the best part of the job anyway.

AI isn't replacing SEO, it's distilling it. What's left is:

- Strategy that aligns with machine logic and user needs.
- Content structure that supports fast retrieval, not just ranking.
- Authority based on more, deeper, sometimes implied, trust signals, not just age or backlinks. Like E-E-A-T++.

Think of it this way: AI strips away the boilerplate. What's left is your real contribution. Your judgment. Your design. Your clarity.

New opportunity lanes are forming right now:

- Writers who evolve into retrievability engineers.
- Editors who become semantic format strategists.
- Technical SEOs who own chunk structuring and indexing hygiene.
- Analysts who specialize in AI visibility benchmarking.

These aren't job titles (yet), but the work is happening. If you're in a role that touches content, structure, trust, or performance, now is the time to sharpen your relevance, not to fear automation.

Final Word

The fundamentals still matter. Technical SEO, content quality, and UX don't go away; they evolve alongside AI.

No, SEO isn't dying, it's becoming more strategic, more semantic, more valuable. AI-driven retrievability is already redefining visibility. Are you ready to adapt?

More Resources:

This post was originally published on Duane Forrester Decodes.

Featured Image: Stock-Asso/Shutterstock

https://www.searchenginejournal.com/which-seo-jobs-ai-will-reshape-which-might-disappear/552687/

Google Backtracks On Plans For URL Shortener Service via @sejournal, @martinibuster

Google announced that they will continue to support some links created by the deprecated goo.gl URL shortening service, saying that 99% of the shortened URLs receive no traffic. They were previously going to end support entirely, but after receiving feedback, they decided to continue support for a limited group of shortened URLs.

Google URL Shortener

Google announced in 2018 that they were deprecating the Google URL Shortener, no longer accepting new URLs for shortening but continuing to support existing URLs. Seven years later, they noticed that 99% of the shortened links did not receive any traffic at all, so on July 18 of this year, Google announced they would end support for all shortened URLs by August 25, 2025.

After receiving feedback, they changed their plan on August 1 and decided that they would move ahead with ending support for URLs that do not receive traffic, but continue servicing shortened URLs that still receive traffic.

Google's announcement explained:

> *"While we previously announced discontinuing support for all goo.gl URLs after August 25, 2025, we've adjusted our*

> *approach in order to preserve actively used links.*
>
> *We understand these links are embedded in countless documents, videos, posts and more, and we appreciate the input received.*
>
> *…If you get a message that states, "This link will no longer work in the near future", the link won't work after August 25 and we recommend transitioning to another URL shortener if you haven't already.*
>
> *…All other goo.gl links will be preserved and will continue to function as normal."*

If you have a goog.gl redirected link, Google recommends visiting the link to check if it displays a warning message. If it does move the link to another URL shortener. If it doesn't display the warning then the link will continue to function.

Featured Image by Shutterstock/fizkes

https://www.searchenginejournal.com/google-backtracks-on-plans-for-url-shortener-service/552783/

Google Confirms It Uses Something Similar To MUVERA via @sejournal,

@martinibuster

Google’s Gary Illyes answered questions during the recent Search Central Live Deep Dive in Asia about whether or not they use the new Multi-Vector Retrieval via Fixed-Dimensional Encodings (MUVERA) retrieval method and also if they’re using Graph Foundation Models.

MUVERA

Google recently announced MUVERA in a blog post and a research paper: a method that improves retrieval by turning complex multi-vector search into fast single-vector search. It compresses sets of token embeddings into fixed-dimensional vectors that closely approximate their original similarity. This lets it use optimized single-vector search methods to quickly find good candidates, then re-rank them using exact multi-vector similarity. Compared to older systems like PLAID, MUVERA is faster, retrieves fewer candidates, and still improves recall, making it a practical solution for large-scale retrieval.

The key points about MUVERA are:

- MUVERA converts multi-vector sets into fixed vectors using Fixed Dimensional Encodings (FDEs), which are single-vector representations of multi-vector sets.
- These FDEs (Fixed Dimensional Encodings) match the original multi-vector comparisons closely enough to support accurate retrieval.
- MUVERA retrieval uses MIPS (Maximum Inner Product Search), an established search technique used in retrieval, making it easier to deploy at scale.
- Reranking: After using fast single-vector search (MIPS)

to quickly narrow down the most likely matches, MUVERA re-ranks them using Chamfer similarity, a more detailed multi-vector comparison method. This final step restores the full accuracy of multi-vector retrieval, so you get both speed and precision.
- MUVERA is able to find more of the precisely relevant documents with a lower processing time than the state-of-the-art retrieval baseline (PLAID) it was compared to.

Google Confirms That They Use MUVERA

José Manuel Morgal (LinkedIn profile) related his question to Google's Gary Illyes and his response was to jokingly ask what MUVERA was and then he confirmed that they use a version of it:

This is how the question and answer was described by José:

> *"An article has been published in Google Research about MUVERA and there is an associated paper. Is it currently in production in Search?*
>
> *His response was to ask me what MUVERA was haha and then he commented that they use something similar to MUVERA but they don't name it like that."*

Does Google Use Graph Foundation Models (GFMs)?

Google recently published a blog announcement about an AI breakthrough called a Graph Foundation Model.

Google's Graph Foundation Model (GFM) is a type of AI that learns from relational databases by turning them into graphs,

where rows become nodes and the connections between tables become edges.

Unlike older models (machine learning models and graph neural networks (GNNs)) that only work on one dataset, GFMs can handle new databases with different structures and features without retraining on the new data. GFMs use a large AI model to learn how data points relate across tables. This lets GFMs find patterns that regular models miss, and they perform much better in tasks like detecting spam in Google's scaled systems. GFMs are a big step forward because they bring foundation-model flexibility to complex structured data.

Graph Foundation Models represent a notable achievement because their improvements are not incremental. They are an order-of-magnitude improvement, with performance gains of 3x to 40x in average precision.

José next asked Illyes if Google uses Graph Foundation Models and Gary again jokingly feigned not knowing what José was talking about.

He related the question and answer:

> *"An article has been published in Google Research about Graph Foundation Models for data, this time there are not paper associated with it. Is it currently in production in Search?*
>
> *His answer was the same as before, asking me what Graph Foundation Models for data was, and he thought it was not in production. He did not know because there are not associated paper and on the other hand, he commented me that he did not control what is published in Google Research blog."*

Gary expressed his opinion that Graph Foundation Model was not currently used in Search. At this point, that's the best information we have.

Is GFM Ready For Scaled Deployment?

The official Graph Foundation Model announcement says it was tested in an internal task, spam detection in ads, which strongly suggests that real internal systems and data were used, not just academic benchmarks or simulations.

Here is what Google's announcement relates:

> *"Operating at Google scale means processing graphs of billions of nodes and edges where our JAX environment and scalable TPU infrastructure particularly shines. Such data volumes are amenable for training generalist models, so we probed our GFM on several internal classification tasks like spam detection in ads, which involves dozens of large and connected relational tables. Typical tabular baselines, albeit scalable, do not consider connections between rows of different tables, and therefore miss context that might be useful for accurate predictions. Our experiments vividly demonstrate that gap."*

Takeaways

Google's Gary Illyes confirmed that a form of MUVERA is in use at Google. His answer about GFM seemed to be expressed as an opinion, so it's somewhat less clear, as it's related as Gary saying that he thinks it's not in production.

Featured Image by Shutterstock/Krakenimages.com

https://www.searchenginejournal.com/google-confirms-that-it-uses-something-similar-to-muvera/552769/

Merging SEO And Content Using Your Knowledge Graph To AI-Proof Content via @sejournal, @marthavanberkel

New AI platforms, powered by generative technologies like Google's Gemini, Microsoft's Copilot, Grok, and countless specialized chatbots, are rapidly becoming the front door for digital discovery.

We've entered an era of machine-led discovery, where AI systems aggregate, summarize, and contextualize content across multiple platforms.

Users today no longer follow a linear journey from keyword to website. Instead, they engage in conversations and move fluidly between channels and experiences.

These shifts are being driven by new types of digital engagement, including:

- **AI-generated overviews, such as AI Overviews in Google**, that pull data from many sources.
- **Conversational search, such as ChatGPT and Gemini,** where follow-up questions replace traditional browsing.
- **Social engagement,** with platforms like TikTok equipped with their own generative search features, engaging entire generations in interactive journeys of discovery.

The result is a new definition of discoverability and a need to rethink how you manage your brand across these experiences.

It's not enough to optimize your brand's website for search

engines. You must ensure your website content is machine-consumable and semantically connected to appear in AI-generated results.

This is why forward-thinking organizations are turning to schema markup (structured data) and building content knowledge graphs to manage the data layer that powers both traditional search and emerging AI platforms.

Semantic structured data transforms your content into a machine-readable network of information, enabling your brand to be recognized, connected, and potentially included in AI-driven experiences across channels.

In this article, we'll explore how SEO and content teams can partner to build a content knowledge graph that fuels discoverability in the age of AI, and why this approach is critical for enterprise brands aiming to future-proof their digital presence.

Table of Contents

Why Schema Markup Is Your Strategic Data Layer

You may be asking, "Schema markup – is that not just for rich results (visual changes in SERP)?"

Schema markup is no longer just a technical SEO tactic for achieving rich results; it can also be used to define the content on your website and its relationship to other entities within your brand.

When you apply markup in a connected way, AI and search can do more accurate inferencing, resulting in more accurate targeting to user queries or prompts.

In May 2025, Google and Microsoft both reiterated that the use of structured data does make your content “machine-readable” that makes you eligible for certain features. [Editor’s note: Although, Gary Illyes recently said to avoid excessive use and that Schema is not a ranking factor.]

Schema markup can be a strategic foundation for creating a data layer that feeds AI systems. While schema markup is a technical SEO approach, it all starts with content.

When You Implement Schema Markup, You’re:

Defining Entities

Schema markup clarifies the “things” your content is about, such as products, services, people, locations, and more.

It provides precise tags that help machines recognize and categorize your content accurately.

Establishing Relationships

Beyond defining individual entities (a.k.a. topics), schema markup describes how those entities connect to each other and to broader topics across the web.

This creates a web of meaning that mirrors how humans understand context and relationships.

Providing Machine-Readable Context

Schema markup assists your content to be machine-readable.

It enables search engines and AI tools to confidently identify, interpret, and surface your content in relevant contexts, which can help your brand appear where it is most relevant.

Enterprise SEO and content teams can work together to implement schema markup to create a content knowledge graph, a

structured representation of your brand's expertise, offerings, and topic authority.

When you do this, the data you put into search and AI platforms is ready for large language models (LLMs) to make accurate inferences, which can help with consumer visibility.

What Is A Content Knowledge Graph?

A content knowledge graph organizes your website's data into a network of interconnected entities and topics, all defined by implementing schema markup based on the Schema.org vocabulary. This graph serves as a digital map of your brand's expertise and topical authority.

Imagine your website as a library. Without a knowledge graph, AI systems trying to read your site have to sift through thousands of pages, hoping to piece together meaning from scattered words and phrases.

With a content knowledge graph:

- **Entities are defined.** Machines can informed precisely who, what, and where you're talking about.
- **Topics are connected.** Machines can better understand and infer how subjects relate. For example, machines can infer that "cardiology" encompasses entities like heart disease, cholesterol, or specific medical procedures.
- **Content becomes query-ready.** your content is assisted to become structured data that AI can reference, cite, and include in responses.

When your content is organized into a knowledge graph, you're effectively supplying AI platforms with information about your products, services, and expertise.

This becomes a powerful control point for how your brand is represented in AI search experiences.

Rather than leaving it to chance how AI systems interpret your web content, you can help to proactively shape the narrative and ensure machines have the right signals to potentially include your brand in conversations, summaries, and recommendations.

Your organization's leaders should be aware this is now a strategic issue, not just a technical one.

A content knowledge graph gives you some influence over how your organization's expertise and authority are recognized and distributed by AI systems, which can impact discoverability, reputation, and competitive advantage in a rapidly evolving digital landscape.

This structure can improve your chances of appearing in AI-generated answers and equips your content and SEO teams with data-driven insights to guide your content strategy and optimization efforts.

How Enterprise SEO And Content Teams Can Build A Content Knowledge Graph

Here's how enterprise teams can operationalize a content knowledge graph to future-proof discoverability and unify SEO and content strategies:

1. Define What You Want To Be Known For

Enterprise brands should start by identifying their core topical authority areas. Ask:

- Which topics matter most to our audience and brand?
- Where do we want to be the recognized authority?
- What new topics are emerging in our industry that we should own?

These strategic priorities shape the pillars of your content knowledge graph.

2. Use Schema Markup To Define Key Entities

Next, use schema markup to:

- Identify key entities tied to your priority topics, such as products, services, people, places, or concepts.
- Connect those entities to each other through Schema.org properties, such as “about,” “mentions,” or “sameAs.”
- Ensure consistent entity definitions across your entire site so that AI systems can reliably identify and understand entities and their relationships.

This is how your content becomes machine-readable and more likely to be accurately included in AI-driven results and recommendations.

3. Audit Your Existing Content Against Your Content Knowledge Graph

Instead of just tracking keywords, enterprises should audit their content based on entity coverage:

- Are all priority entities represented on your site?
- Do you have “entity homes” (pillar pages) that serve as authoritative hubs for those priority entities?
- Where are there gaps in entity coverage that could limit your presence in search and AI responses?
- What content opportunities exist to improve coverage of priority entities where these gaps have been identified?

A thorough audit provides a clear roadmap for aligning your content strategy with how machines interpret and surface information, ensuring your brand has the potential to be discoverable in evolving AI-driven search experiences.

4. Create Pillar Pages And Fill Content Gaps

Based on your findings from Step 3, create dedicated pillar pages for high-priority entities where needed. These become the authoritative source that:

- Defines the entity.
- Links to supporting content, including case studies, blog posts, or service pages.
- Signals to search engines and AI systems on where to find reliable information about that entity.

Supporting content can then be created to expand on subtopics and related entities that link back to these pillar pages, ensuring comprehensive coverage of topics.

5. Measure Performance By Entity And Topic

Finally, enterprises should track how well their content performs at the entity and topic levels:

- Which entities drive impressions and clicks in AI-powered search results?
- Are there emerging entities gaining traction in your industry that you should cover?
- How does your topical authority compare to competitors?

This data-driven approach enables continuous optimization, helping you to stay visible as AI search evolves.

Why SEO And Content Teams Are The Heroes Of The AI Search Evolution

In this new landscape, where AI generates answers before users ever reach your website, schema markup and content knowledge

graphs provide a critical control point.

They enable your brand to signal its authority to machines, support the possibility of accurate inclusion in AI results and overviews, and inform SEO and content investment based on data, not guesswork.

For enterprise organizations, this isn’t just an SEO tactic; it’s a strategic imperative that could protect visibility and brand presence in the new digital ecosystem.

So, the question remains: What does your brand want to be known for?

Your content knowledge graph is the infrastructure that ensures AI systems, and by extension, your future customers, know the answer.

More Resources:

Featured Image: Urbanscape/Shutterstock

https://www.searchenginejournal.com/merging-seo-content-using-your-knowledge-graph-to-ai-proof-content/548460/

2025 Core Web Vitals Challenge: WordPress Versus Everyone via @sejournal,

@martinibuster

The Core Web Vitals Technology Report shows the top-ranked content management systems by Core Web Vitals (CWV) for the month of June (July's statistics aren't out yet). The breakout star this year is an e-commerce platform, which is notable because shopping sites generally have poor performance due to the heavy JavaScript and image loads necessary to provide shopping features.

This comparison also looks at the Interaction to Next Paint (INP) scores because they don't mirror the CWV scores. INP measures how quickly a website responds visually after a user interacts with it. The phrase "next paint" refers to the moment the browser visually updates the page in response to a user's interaction.

A poor INP score can mean that users will be frustrated with the site because it's perceived as unresponsive. A good INP score correlates with a better user experience because of how quickly the website performs.

Table of Contents

Core Web Vitals Technology Report

The HTTP Archive Technology Report combines two public datasets:

1. Chrome UX Report (CrUX)
2. HTTP Archive

1. Chrome UX Report (CrUX)
CrUX obtains its data from Chrome users who opt into providing

usage statistics reporting as they browse over 8 million websites. This data includes performance on Core Web Vitals metrics and is aggregated into monthly datasets.

2. HTTP Archive
HTTP Archive obtains its data from lab tests by tools like WebPageTest and Lighthouse that analyze how pages are built and whether they follow performance best practices. Together, these datasets show how websites perform and what technologies they use.

The CWV Technology Report combines data from HTTP Archive (which tracks websites through lab-based crawling and testing) and CrUX (which collects real-user performance data from Chrome users), and that’s where the Core Web Vitals performance data of content management systems comes from.

#1 Ranked Core Web Vitals (CWV) Performer

The top-performing content management system is Duda. A remarkable 83.63% of websites on the Duda platform received a good CWV score. Duda has consistently ranked #1, and this month continues that trend.

For Interaction to Next Paint scores, Duda ranks in the second position.

#2 Ranked CWV CMS: Shopify

The next position is occupied by Shopify. 75.22% of Shopify websites received a good CWV score.

This is extraordinary because shopping sites are typically burdened with excessive JavaScript to power features like product filters, sliders, image effects, and other tools that shoppers rely on to make their choices. Shopify, however,

appears to have largely solved those issues and is outperforming other platforms, like Wix and WordPress.

In terms of INP, Shopify is ranked #3, at the upper end of the rankings.

#3 Ranked CMS For CWV: Wix

Wix comes in third place, just behind Shopify. 70.76% of Wix websites received a good CWV score. In terms of INP scores, 86.82% of Wix sites received a good INP score. That puts them in fourth place for INP.

#4 Ranked CMS: Squarespace

67.66% of Squarespace sites had a good CWV score, putting them in fourth place for CWV, just a few percentage points behind the No. 3 ranked Wix.

That said, Squarespace ranks No. 1 for INP, with a total of 95.85% of Squarespace sites achieving a good INP score. That’s a big deal because INP is a strong indicator of a good user experience.

#5 Ranked CMS: Drupal

59.07% of sites on the Drupal platform had a good CWV score. That’s more than half of sites, considerably lower than Duda’s 83.63% score but higher than WordPress’s score.

But when it comes to the INP score, Drupal ranks last, with only 85.5% of sites scoring a good INP score.

#6 Ranked CMS: WordPress

Only 43.44% of WordPress sites had a good CWV score. That’s over fifteen percentage points lower than fifth-ranked Drupal.

So WordPress isn’t just last in terms of CWV performance; it’s last by a wide margin.

WordPress performance hasn’t been getting better this year either. It started 2025 at 42.58%, then went up a few points in April to 44.93%, then fell back to 43.44%, finishing June at less than one percentage point higher than where it started the year.

WordPress is in fifth place for INP scores, with 85.89% of WordPress sites achieving a good INP score, just 0.39 points above Drupal, which is in last place.

But that’s not the whole story about the WordPress INP scores. WordPress started the year with a score of 86.05% and ended June with a slightly lower score.

INP Rankings By CMS

Here are the rankings for INP, with the percentage of sites exhibiting a good INP score next to the CMS name:

1. Squarespace 95.85%
2. Duda 93.35%
3. Shopify 89.07%
4. Wix 86.82%
5. WordPress 85.89%
6. Drupal 85.5%

As you can see, positions 3–6 are all bunched together in the eighty percent range, with only a 3.57 percentage point difference between the last-placed Drupal and the third-ranked Shopify. So, clearly, all the content management systems deserve a trophy for INP scores. Those are decent scores, especially for Shopify, which earned a second-place ranking for CWV and third place for INP.

Takeaways

- **Duda Is #1**
  Duda leads in Core Web Vitals (CWV) performance, with 83.63% of sites scoring well, maintaining its top position.
- **Shopify Is A Strong Performer**
  Shopify ranks #2 for CWV, a surprising performance given the complexity of e-commerce platforms, and scores well for INP.
- **Squarespace #1 For User Experience**
  Squarespace ranks #1 for INP, with 95.85% of its sites showing good responsiveness, indicating an excellent user experience.
- **WordPress Performance Scores Are Stagnant**
  WordPress lags far behind, with only 43.44% of sites passing CWV and no signs of positive momentum.
- **Drupal Also Lags**
  Drupal ranks last in INP and fifth in CWV, with over half its sites passing but still underperforming against most competitors.
- **INP Scores Are Generally High Across All CMSs**
  Overall INP scores are close among the bottom four platforms, suggesting that INP scores are relatively high across all content management systems.

Find the Looker Studio rankings for here (must be logged into a Google account to view).

Featured Image by Shutterstock/Krakenimages.com

https://www.searchenginejournal.com/2025-core-web-vitals-cms-rankings/552679/

Google URL Removal Bug Enabled Attackers To Deindex URLs via @sejournal, @martinibuster

Google recently fixed a bug that enabled anyone to anonymously use an official Google tool to remove any URL from Google search and get away with it. The tool had the potential to be used to devastate competitor rankings by removing their URLs completely from Google's index. The bug was known by Google since 2023 but until now Google hadn't taken action to fix it.

Tool Exploited For Reputation Management

A report by the Freedom of the Press Foundation recounted the case of a tech CEO who had employed numerous tactics to "censor" negative reporting by a journalist, ranging from legal action to identify the reporter's sources, an "intimidation campaign" via the San Francisco city attorney and a DMCA takedown request.

Through it all, the reporter and the Freedom of the Press Foundation prevailed in court, and the article at the center of the actions remained online until it began getting removed through abuse of Google's Remove Outdated Content tool. Restoring the web page with Google Search Console was easy, but the abuse continued. This led to opening a discussion on the Google Search Console Help Community.

The person posted a description of what was happening and

asked if there was a way to block abuse of the tool. The post alleged that the attacker was choosing a word that was no longer in the original article and using that as the basis for claiming an article is outdated and should be removed from Google's search index.

This is what the report on Google's Help Community explained:

> *"We have a dozen articles that got removed this way. We can measure it by searching Google for the article, using the headline in quotes and with the site name. It shows no results returned.*
>
> *Then, we go to GSC and find it has been "APPROVED" under outdated content removal. We cancel that request. Moments later, the SAME search brings up an indexed article. This is the 5th time we've seen this happen."*

Four Hundred Articles Deindexed

What was happening was an aggressive attack against a website, and Google apparently was unable to do anything to stop the abuse, leaving the user in a very bad position.

In a follow-up post, they explained the devastating effect of the sustained negative SEO attack:

> *"Every week, dozens of pages are being deindexed and we have to check the GSC every day to see if anything else got removed, and then restore that.*
>
> *We've had over 400 articles deindexed, and all of the articles were still live and on our sites. Someone went in and submitted them through the public removal tool, and they got deindexed."*

Google Promised To Look Into It

They asked if there was a way to block the attacks, and Google's Danny Sullivan responded:

> *"Thank you – and again, the pages where you see the removal happening, there's no blocking mechanism on them."*

Danny responded to a follow-up post, saying that they would look into it:

> *"The tool is designed to remove links that are no longer live or snippets that are no longer reflecting live content. We'll look into this further."*

How Google's Tool Was Exploited

The initial report said that the negative SEO attack was leveraging changed words within the content to file a successful outdated content removal. But it appears that they later discovered that another attack method was being used.

Google's Outdated Content Removal tool is case-sensitive, which means that if you submit a URL containing an uppercase letter, the crawler will go out to specifically check for the uppercase version, and if the server returns a 404 Not Found error response, Google will remove all versions of the URL.

The Freedom of the Press Foundation writes that the tool is case insensitive, but that's not entirely correct because if it were insensitive, the case wouldn't matter. But the case does matter, which means that it is case sensitive.

By the way, the victim of the attack could have created a workaround by rewriting all requests for uppercase URLs to lowercase and enforcing lowercase URLs across the entire website.

That’s the flaw the attacker exploited. So, while the tool was case sensitive, at some point in the system Google’s removal system is case agnostic, which resulted in the correct URL being removed.

Here’s how the Freedom of the Press Foundation described it:

> *“Our article… was vanished from Google search using a novel maneuver that apparently hasn’t been publicly well documented before: a sustained and coordinated abuse of Google’s “Refresh Outdated Content” tool.*
>
> *This tool is supposed to allow those who are not a site’s owner to request the removal from search results of web pages that are no longer live (returning a “404 error”), or to request an update in search of web pages that display outdated or obsolete information in returned results.*
>
> *However, a malicious actor could, until recently, disappear a legitimate article by submitting a removal request for a URL that resembled the target article but led to a “404 error.” By altering the capitalization of a URL slug, a malicious actor apparently could take advantage of a case-insensitivity bug in Google’s automated system of content removal.”*

Other Sites Affected By Thes Exploit

Google responded to the Freedom of the Press Foundation and admitted that this exploit did, in fact, affect other sites.

They are quoted as saying the issue only impacted a “tiny fraction of websites” and that the wrongly impacted sites were reinstated.

Google responded by email to note that this bug has been fixed.

https://www.searchenginejournal.com/google-url-removal-bug-enabled-attackers-to-deindex-urls/552667/

Industry Pioneer Reveals Why SEO Isn’t Working & What To Refocus On via @sejournal, @theshelleywalsh

Bill Hunt is a true pioneer in the industry, with more than 25 years of experience working on the websites of some of the largest multinationals. Having built two large digital/search agencies, one of which was acquired by Ogilvy, Bill has now moved into consulting focused on repositioning search to leverage marketing for shareholder growth.

His approach is not myopic surface-level SEO, but as an enterprise specialist who looks at what users actually want from their online experience. He connects the dots between search visibility, user experience, and business value for real results.

Bill is currently writing a series for Search Engine Journal about connecting search visibility to business value, and I spoke to him for IMHO to find out why he thinks SEO is currently not working.

> *“SEOs are creatures of habit. To succeed now, we need to unlearn and relearn how discovery actually works.”*

The Real Problems Aren’t What You Think

I started out by asking Bill why SEO isn’t working, and his key message was not that SEO is broken, but there is paralysis, distraction from AI hype, and a neglect of fundamentals:

> *“I think there are three key problems right now. One is paralysis. We see that clients put search on pause, especially organic search, because they just don’t know what to do.*
>
> *The second is the distraction with all the hype around the AI thing.*
>
> *I mean, there’s a different acronym every day. So, which do we do? Are we chasing answers? Are we doing LLM index files or whatever craziness comes out?*
>
> *And then the third is that there’s such a distraction from all this that a lot of the fundamentals aren’t being covered. And I think that’s where the problem is.”*

Bill emphasized that the impact varies significantly by business type. Information-based businesses have been significantly affected because AI now directly answers queries that previously drove traffic to their sites. However, many other businesses might not be negatively impacted if they understand what’s actually changed.

Three Fundamental Shifts To Pay Attention To

Bill went on to talk about how three core changes have reshaped search, and understanding them is crucial for

adaptation:

- **Intent understanding has evolved:** Everything is about what did they search for? What are they hoping to see?
- **Friction must be removed:** Platforms reward the path of least resistance.
- **Monetization is leading the way:** It's not just about helpful, but also about profitable.

Bill used an example from his work with Absolut Vodka.

> *"When I was working with Absolut Vodka, we had a drink site that was really just an awareness driver, and every month we sat down and we looked at Google's search results and said, 'If we were Google, what would we be changing around drinks or recipes or things like that?'*
>
> *And so, by looking at the results, we could see, little by little, [that] somebody [was] looking for yellow cocktails. What should Google present?"*
>
> *Rather than just optimizing for rankings, his team studied Google's interface changes and adapted their visual content accordingly.*
>
> *"We started focusing on the drink, bringing it front and center, amplifying the colors, the ingredients, and more and more people clicked.*
>
> *We were generating millions and millions of visits because every step that Google was making to create a different user experience, we were trying to accommodate it."*

Bill believes that the idea of intent is still crucial. Considering how users just want to get to an answer, we must think about how they discover information and how we then present information to them.

> *"I think that's really it in a nutshell. All of this change*

> *has paralyzed us and distracted us, and we need to recenter and refocus.*
>
> *And that's really a key part of what this series [at SEJ] is about: How do we refocus? How do we rethink this, both from a strategic point of view, from a shareholder value standpoint, and from a simple workflow standpoint?"*

AI Tools Reward Consensus, Not Originality

In a recent LinkedIn post, Bill stated that AI tools don't reward originality; they reward consensus.

As generative AI becomes embedded into how users explore and consume information, Bill warned against assuming that originality is enough to get discovered.

> *"AI systems synthesize consensus. If you're saying something radically different, you won't show up unless you connect it to what people already know."*

So, I asked Bill if you are creating this original content, how do you teach the systems to see you?

Bill's advice is that to succeed in AI search environments, businesses need to:

- Link new ideas to familiar terms.
- Reflect user language and legacy concepts.
- Be explicit in bridging the gap between old and new methods.

Otherwise, you risk being invisible to LLMs and answer engines that rely on summarizing well-established viewpoints.

> *"If you're stating that you're radically different, you're*

> *not going to be shown because you're radically different. So, you have to connect, and this is what I put in that article. You need to connect back to the consensus idea.*
>
> *If you're saying you've got a new way to cut bread, you have to talk about the old way to cut bread and connect it to a more efficient or easier way to do it."*

Is Your Product Even Discoverable?

The most practical insight from our conversation surrounded how people can discover your brand or your product.

Historically, keyword research has been focused on connecting to searches that have existing search volume. But, if somebody doesn't know a product exists to solve a problem, how would they search for it?

> *"I used to tell companies, if somebody doesn't know a product exists to solve a problem, how would they search for it?*
>
> *They would use the problem or symptoms of the problem. If they know a product exists but don't know you exist, how would they search for it?"*

Bill recommended that you run searches for problems related to your product and see if you show up. Search as if you know the solution exists, but not your brand.

And if you don't surface, ask yourself why not?

> *"Take the symptoms people have, go into any tool you want, Google, Perplexity, ChatGPT, Gemini, and search and see if you come up.*
>
> *If you don't come up, the very next question you should ask is, 'Why isn't this product or this company in your result set?' That's probably the single most illuminating thing a*

> *senior executive can do…*
>
> *When it tells you that you don't have the answer, your very next step is, 'How do we then create the answer, and then how do we get it into these?'"*

This kind of query-path analysis is more revealing than traditional keyword research because it aligns with how people actually search, especially in AI environments that interpret broader queries.

Moving Forward: Back To Basics

Despite all the AI disruption, Bill recommends a return to fundamental principles. Companies need to ensure they're indexable, crawlable, and seen as authorities in their space. The same core elements that have always mattered for search visibility.

> *"Who got cited? Who was number one? And Larry and Sergey said, 'Well, if they're cited most frequently as a source for a question, shouldn't they be?'"*

The key difference is that these fundamentals now operate in an AI-enhanced environment where understanding user intent and creating relevant, engaging content matter more than ever.

And if you want to find answers, ask the tools; they can tell you everything you need to know.

> *"I would tell everybody to go do that query and do the follow-up saying why aren't we there? And you'd be surprised how efficient these tools are at telling you what you need to do to close that gap."*

Rather than panicking about AI destroying SEO, organizations should focus on understanding what's actually changed and

adapting their strategies accordingly.

The fundamentals remain solid; they just need to be applied in new ways.

You can watch the full interview with Bill Hunt below:

[embedded content]

Don’t miss the new series that Bill is currently writing for SEJ about how you can connect the dots between search visibility, user experience, and business value that will not only help CMOs but also help search marketers get buy-in from CMOs.

Thank you to Bill Hunt for offering his insights and being my guest on IMHO.

More Resources:

Featured Image: Shelley Walsh/Search Engine Journal

https://www.searchenginejournal.com/industry-pioneer-reveals-why-seo-isnt-working-what-to-refocus-on/552235/