

# Google Zero: What the Open Web Loses When AI Answers Replace the Click

Second Order Labs™ · May 2026

## ABSTRACT

Publishers blocking AI crawlers think they're losing. They're actually pulling the foundation out from under the AI industry and triggering a feedback loop that ends in model collapse.

*Keywords:* Google Zero, AI Overviews, Zero-Click Search, Generative Engine Optimization, Model Collapse, Data Licensing

*“Google is finally prioritizing the user experience of getting an answer over the ecosystem experience of sending traffic.”*

— Ben Thompson, Stratechery, Google I/O and the AI Search Era

Publishers blocking AI crawlers think they are protecting themselves. They are. But they are also doing something bigger. They are cutting off the raw material AI companies need.

When a news site blocks GPTBot in robots.txt, the immediate motive is obvious: stop the scrape, protect the archive. Most publisher coverage and SEO analysis stops there. That misses the supply problem on the other side. Every blocked crawler removes another source of fresh human writing from the next training run.

Trade press coverage and publisher lawsuits usually frame Google Zero as a fight between aggregators and creators. The traffic numbers are ugly enough to support that framing. SparkToro found that for every 1,000 US Google searches, 593 end without a click to the open web.<sup>[1]</sup> Gartner expects traditional search volume to drop 25% by 2026.<sup>[2]</sup> Those are not side effects. They are evidence that the old bargain is breaking.

What matters more is what follows. Scraping wipes out clicks. Once clicks vanish, creators stop publishing for free. What fills the gap is cheap synthetic junk.

Then the models train on that junk. I call that the **Synthesis Starvation Cycle**, and it is a self-inflicted wound.

## **I. The Implicit Contract That Just Got Torn Up**

The open web worked because everyone accepted a simple trade. Creators published free, indexable work. Search engines indexed it and sent traffic back. That traffic turned into money and status. The link was the payment.

What exactly did AI Overviews break? They kept the answer and cut out the return trip. Ben Thompson put it cleanly: "Google is finally prioritizing the user experience of getting an answer over the ecosystem

experience of sending traffic."<sup>[5]</sup> That is great for the person searching. It is bad for the site that paid to produce the answer. The Atlantic got to the point faster: "If Google is answering the questions, why would anyone click through to the site that actually did the work?"<sup>[4]</sup>

Google still talks as if the old deal survives. Sundar Pichai told The Verge he remains "optimistic that the ecosystem will thrive," and Google says links inside AI Overviews "get more clicks than if the page had appeared as a traditional web listing."<sup>[5][6]</sup> Publishers, outside studies, and court filings keep saying the opposite. Both things cannot be true. The cleaner explanation is that Google is trapped. The direct answer makes search better for users and worse for the web that supplies the answer.

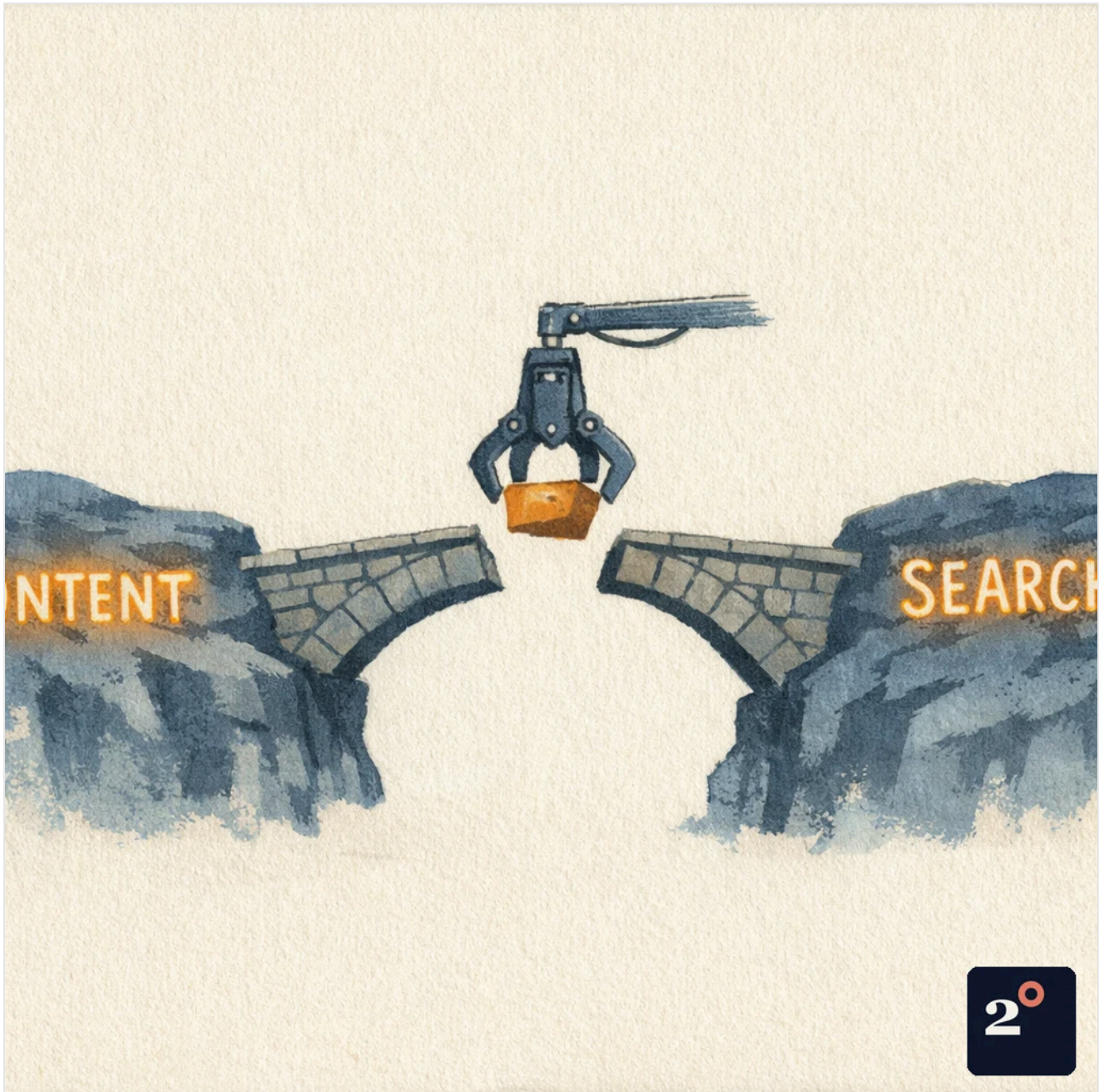


Figure 1: The link was the keystone of an unwritten contract; AI answers remove it without building a replacement.

## II. The Synthesis Starvation Cycle: AI's Self-Inflicted Wound

Most search-industry coverage ignores the feedback loop in the data supply chain. The real danger is not that publishers lose traffic. It is that AI systems degrade the source they depend on.

Scraping and synthesis collapse referral traffic. Creators respond by putting work behind paywalls, requiring logins, selling API access, or quitting. The

gap does not stay empty for long. As 404 Media reported, "the open web is being backfilled by cheap, AI-generated SEO slop designed to game the exact algorithms meant to surface human knowledge."<sup>[7]</sup> Then the models ingest that slop.

We already know where that ends. A 2024 Nature paper found that "use of model-generated content in training causes irreversible defects in the resulting models, where tails of the original content distribution disappear."<sup>[8]</sup> That is model collapse in plain

terms. The rare, weird, expert, low-volume human material disappears first. That is exactly the material the web has always been best at producing.

*The publishers are not just defending their content; they are starving the machine that needs them.*

### III. Why Blocking Crawlers Became the New SEO

For years, publisher survival meant welcoming bots. You wanted Googlebot to crawl everything. Rank was the goal. Traffic was the reward.

Now look at what publishers tracked by Nieman Lab are doing. They are blocking GPTBot and Apple's scraper in robots.txt.<sup>[10]</sup> OpenAI's own docs explain how to block GPTBot and say the pages it crawls can be used to improve future models.<sup>[9]</sup> OpenAI leaves some room in the wording, but the point is obvious. The crawler exists to feed the model.

Cloudflare saw where this was going and turned it into a product. One click, block AI bots.<sup>[11]</sup> So an infrastructure company now gets to decide which machines can read the web. And the aggregate effect of widespread blocking by publishers and platforms is straightforward: less fresh human material for the next model generation. Each individual block is rational. Taken together, they choke the input stream.

Figure 2: The web bifurcates into bright, gated silos and a dark forest of synthetic replacements.

### IV. The Web Splits Into Silos and Slop

The web is dividing into two ugly categories. Paid silos full of authenticated human material. And a public web increasingly stuffed with filler.

Reddit and Stack Overflow can sell access. The indie forum owner cannot.

Reddit said in its S-1 that its growing archive of conversation will become more valuable for training AI models.<sup>[12]</sup> Stack Overflow launched OverflowAPI to sell a live feed of its public knowledge base for training and fine-tuning.<sup>[13]</sup> News Corp said it was negotiating for "fair value" because its content would be "a vital input for AI models."<sup>[14]</sup> Human conversation, especially authenticated conversation, now has a price tag.

Tier	Who	Strategy	Leverage
Licensed data silos	Reddit, Stack Overflow, News Corp	Sell access via deals and APIs	High: negotiating power to license content
Long-Tail Commons	Independent blogs, forums, niche experts	Block crawlers or go dark	Near zero: no negotiating power

Independent bloggers, forum operators, and niche experts are getting nothing. They cannot cut licensing deals. Their options are narrower: block the bots, gate content, require login access, or stop publishing. The

long tail is where the web's odd, specific, hard-won knowledge lives. It is also the first thing to disappear when publishing no longer pays.

## V. The Citation Economy and What Operators Should Do

For SEO-driven publishers, the click-maximizing playbook is breaking down as zero-click search rises. The next contest is not for the visit. It is for the citation.

Researchers from Princeton and IIT Delhi gave that shift a name: "Generative Engine Optimization (GEO), a novel paradigm to optimize content for visibility in generative search engines."<sup>[15]</sup> Strip away the academic label and the idea is simple. You are trying to become the source the model mentions, not the blue link the user clicks. That has value. But it is weaker value. A citation can build recall. It does not replace the economics of an actual visit.

Perplexity's Publishers Program is the clearest admission that plain scraping does not hold up. The company says that "when Perplexity generates revenue from an interaction where a publisher's content is referenced, that publisher earns a share."<sup>[16]</sup> Revenue

sharing is not generosity. It is recognition that answer engines need publishers to keep publishing, and the click no longer does that job.<sup>[17]</sup>

The legal fight sharpens the same point. The New York Times argues that OpenAI and Microsoft are using its reporting to build "substitutive products without permission or payment."<sup>[18]</sup> That word, substitutive, matters. If courts decide that an AI answer replacing the reason to visit the original work is not fair use but market substitution, the economics of scraping change fast.

Sam Altman says "the world needs is a better way to find, synthesize, and act on information."<sup>[19]</sup> Fine. But synthesis depends on a supply of human-created information worth synthesizing. Big platforms should charge for access. Small publishers should gate access and stop pretending traffic will come back. The web's power is not moral outrage or legal theory. It is refusal. If answer engines keep suppressing clicks, the next generation of models will train on a thinner, worse web, and their answers will rot with it.

### KEY FINDINGS

SparkToro found that for every 1,000 US Google searches, 593 end in zero clicks to the open web.

Gartner projects traditional search engine volume will drop 25% by 2026 as users shift to AI chatbots.

A 2024 Nature paper showed that training models on AI-generated content causes irreversible defects, with the tails of the original data distribution disappearing first.

Reddit, Stack Overflow, and News Corp now sell access to their data, while independent creators have no licensing leverage and can only block crawlers or go dark.

Perplexity's Publishers Program shares revenue when cited content drives an interaction, an admission that answer engines need publishers to keep publishing.

---

## REFERENCES

- [1] SparkToro, 2024 Zero-Click Search Study. <https://sparktoro.com/blog/2024-zero-click-search-study-for-every-1000-google-searches-only-360-clicks-to-the-open-web/> <https://sparktoro.com/blog/2024-zero-click-search-study-for-every-1000-us-google-searches-only-374-clicks-go-to-the-open-web-in-the-eu-its-360/>
- [2] Gartner, Predicts Traditional Search Engine Volume Will Drop 25% by 2026. <https://www.gartner.com/en/newsroom/press-releases/2024-02-19-gartner-predicts-traditional-search-engine-volume-will-drop-25-percent-by-2026> <https://www.gartner.com/en/newsroom/press-releases/2024-02-19-gartner-predicts-search-engine-volume-will-drop-25-percent-by-2026-due-to-ai-chatbots-and-other-virtual-agents>
- [3] Ben Thompson, Google I/O and the AI Search Era, Stratechery. <https://stratechery.com/2024/google-io-and-the-ai-search-era/>
- [4] The Atlantic, The Internet Is About to Get Much Worse. <https://www.theatlantic.com/technology/archive/2024/05/google-ai-overviews-search/678436/> <https://www.theatlantic.com/technology/archive/2024/05/google-ai-overviews-search/678436/>
- [5] Sundar Pichai, interview with The Verge. <https://www.theverge.com/24158374/google-ceo-sundar-pichai-ai-search-gemini-future-of-the-internet-web-decoder> <https://www.theverge.com/24158374/google-ceo-sundar-pichai-ai-search-gemini-future-of-the-internet-web-decoder>
- [6] Google, Generative AI in Search developer blog. <https://developers.google.com/search/blog/2024/05/generative-ai-in-search>
- [7] 404 Media, AI Spam Is Already Starting to Ruin Google Search. <https://www.404media.co/ai-spam-is-already-starting-to-ruin-google-search/>
- [8] Nature, AI models collapse when trained on recursively generated data. <https://www.nature.com/articles/s41586-024-07566-y> <https://www.nature.com/articles/s41586-024-07566-y>
- [9] OpenAI, GPTBot documentation. <https://platform.openai.com/docs/gptbot> <https://platform.openai.com/docs/gptbot>
- [10] Nieman Lab, A growing number of news publishers are blocking Apple's AI data scraper. <https://www.niemanlab.org/2024/07/a-growing-number-of-news-publishers-are-blocking-apples-ai-data-scraper/> <https://www.niemanlab.org/2024/07/a-growing-number-of-news-publishers-are-blocking-apples-ai-data-scraper/>
- [11] Cloudflare, Declare your AI independence. <https://blog.cloudflare.com/declaring-your-ai-independence-block-ai-bots-scrapers-and-crawlers-with-a-single-click> <https://blog.cloudflare.com/declaring-your-ai-independence-block-ai-bots-scrapers-and-crawlers-with-a-single-click/>

- [12] Reddit, Inc. Form S-1 Registration Statement. <https://www.sec.gov/Archives/edgar/data/1713445/000162828024006294/reddits-1.htm> <https://www.sec.gov/Archives/edgar/data/1713445/000162828024006294/reddits-1.htm>
- [13] Stack Overflow, Announcing OverflowAPI. <https://stackoverflow.blog/2024/02/28/announcing-overflowapi/>
- [14] News Corp, Fiscal 2024 Q3 Earnings Release. <https://investors.newscorp.com/news-releases/news-release-details/news-corporation-reports-third-quarter-results-fiscal-2024> <https://www.sec.gov/Archives/edgar/data/0001564708/000156470824000246/release-q3fy2024.htm>
- [15] GEO: Generative Engine Optimization, arXiv. <https://arxiv.org/abs/2311.09735> <https://arxiv.org/abs/2311.09735>
- [16] Perplexity, Introducing the Perplexity Publishers Program. <https://www.perplexity.ai/hub/blog/perplexity-publishers-program> <https://www.perplexity.ai/hub/blog/introducing-the-perplexity-publishers-program>
- [17] Aravind Srinivas, Lex Fridman Podcast #434. <https://lexfridman.com/aravind-srinivas/> <https://lexfridman.com/aravind-srinivas/>
- [18] The New York Times Company v. Microsoft Corp. and OpenAI. [https://nytco-assets.nytimes.com/2023/12/NYT\\_Complaint\\_Dec2023.pdf](https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf) [https://nytco-assets.nytimes.com/2023/12/NYT\\_Complaint\\_Dec2023.pdf](https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf)
- [19] Sam Altman, Lex Fridman Podcast #419. <https://lexfridman.com/sam-altman-3/> <https://lexfridman.com/sam-altman-2/>