kymata labs

RESEARCH PAPER 03 / AI & THE INFORMATION ECONOMY

The Search Engine Is Eating Its Suppliers.

The answer may be free. The next fact still has a production bill.



Kymata Labs Research
2026-11-15 · Edition 1.0

Independent analysis
kymatalabs.com

Small studio. Open worlds.

THE LONGER VIEW

Executive summary.

An answer interface can create value for a reader while changing how the underlying information is funded. This paper examines a 2026 Google AI Overviews research release, historical attribution research and current crawler-control documentation. It distinguishes source use, credit, visits and financial contribution—four events that a single visibility score cannot represent. A transparent hypothetical publisher model shows when more valuable traffic can offset fewer visits and when additional direct or licensing contribution is required. The evidence establishes exposure and mechanisms, not an industry-wide causal loss of traffic or revenue. The strongest case for answer engines belongs in the analysis: better answers, wider discovery and alternative compensation can benefit both readers and producers. The unresolved test is whether those mechanisms replenish the original reporting and expertise on which future answers depend.

> Analysis of original research repositories, saved outputs and official documentation. The 2026 search study covers a historical, trends-derived query cohort; its full raw pipeline was not independently rerun. No new search crawl, publisher interview or causal estimate of traffic or revenue loss is claimed. Financial scenarios are explicitly hypothetical. Source access: November 16, 2026 UTC.

01 / An answer still has a production bill
02 / What the available measurements actually show
03 / The discovery bargain has several moving parts
04 / A footnote does not complete the transaction
05 / A small publisher's arithmetic
06 / The strongest case for the answer engine
07 / Charging for access is a mechanism, not a settlement
08 / Publisher controls need separate decisions
09 / Measure the relationship you want to preserve
10 / The scarce input is the next piece of work
11 / Evidence and method
Q&A / Questions worth asking
Sources / References

© 2026 Kymata Labs LLC. All rights reserved.
Read the web edition: kymatalabs.com/research/the-search-engine-is-eating-its-suppliers

An answer still has a production bill

A search engine can answer a question without asking the reader to visit the people who did the work. That is a useful product feature. It can also change who gets paid.

The weather explanation, product comparison, local investigation or technical guide inside an answer did not necessarily begin inside the answer engine. Someone gathered information, tested a claim, interviewed a source, maintained a reference page or paid for expertise. The platform can make that work easier to consume while moving the moment of consumption away from its producer.

There is no contradiction here. A better experience for the reader can weaken the business that supplies the information. Equally, an answer can introduce a valuable source to someone who would never have found it. Both possibilities deserve measurement. Neither is settled by counting citations.

The title of this paper names a risk in the business model, not a measured verdict that AI search is bankrupting publishers. The evidence inspected here establishes answer generation, selective attribution and exposure of advertising-supported sources. It does not establish an industry-wide causal loss of traffic or revenue. That missing step matters too much to fill with rhetoric.

Our argument is narrower and more consequential: **an answer interface needs a way to replenish the work it draws from.** Referral traffic is one way. Direct subscriptions, service revenue, licensing and paid access are others. A system that produces excellent answers while returning too little value to indispensable sources can succeed as a product and fail as an information market.

What the available measurements actually show

Haofei Xu, Umar Iqbal and Jacob M. Montgomery released data and code for a 2026 study of Google AI Overviews. Its collection covered March 13 through April 21, 2026: 55,393 query executions drawn from US Google Trends categories, 7,583 AI Overview observations and 19 topical categories. We inspected the repository, analysis code and saved notebook outputs. These are historical research observations, not a measurement of search conditions today. [1]

The overall activation rate was 13.7%. That average conceals a substantial difference in the query population. Queries beginning with one of the study's 15 interrogative words produced an Overview in 2,716 of 4,198 executions, or 64.7%. Other queries produced one in 4,867 of 51,195 executions, or 9.5%. Sports supplied 51.4% of all executions. The sample therefore cannot establish what percentage of every person's searches receive AI answers. Reweighting repeated queries changed the overall estimate from 13.69% to 16.03%. [1]

The publisher evidence is especially relevant. Across 61,212 reference occurrences in the Overviews, the released analysis counted 30,994 with detected advertising on pages outside its predefined social-platform list: 50.63% of all reference occurrences. Its detection rule was a positive iframe-ad count. The denominator includes references on that social-platform list, which were excluded from the ad detection step. These are repeated appearances of referenced URLs across observations, not 61,212 distinct articles, publishers or visits. [1]

This tells us that answer generation frequently drew on pages with an observable advertising business. It does not tell us how many visitors those pages lost, how much an ad impression would have earned, whether readers subscribed later, or whether the information would otherwise have gone unread. The study supplies evidence of exposure to substitution. It does not supply the substitution rate.

Nor did Google advertising saturate the observed answers. The study's saved sponsored-ad indexes contained 164 Overview search pages with ads in its near-after category, 39 with ads above an Overview, and none inside one. Those figures should not be stretched into a claim that every answer monetizes content beside an advertisement. [1]

The evidence is useful precisely because it stops short of the dramatic conclusion. There are identifiable suppliers. Some finance their pages with advertising. The answer can deliver information without delivering a visit. The size and distribution of the financial consequence require another layer of observation.

FIGURE 01 OBSERVED COHORT · MARCH–APRIL 2026

Exposure is measurable. Loss is another question.

AI Overview activation varied sharply across a trends-derived query population. The source and query counts below are different units.

01

13.7%

All query executions

7,583 Overviews across 55,393 executions in 19 US Google Trends categories.

02

64.7%

Leading question word

2,716 of 4,198 executions; other queries: 4,867 of 51,195, or 9.5%.

03

50.63%

References with detected ads

30,994 of 61,212 reference occurrences. Repeated URLs count; social-platform references remain in the denominator.

Sports supplied 51.4% of executions. Ad detection used positive iframe-ad counts outside a predefined social-platform list. No click or revenue loss was measured.

Xu, Iqbal and Montgomery, 2026 released research notebooks; definitions and arithmetic inspected. [1]

The discovery bargain has several moving parts

The familiar search bargain joins two exchanges. A publisher makes material available for discovery. A search service helps readers find it. When readers arrive, the publisher may earn advertising revenue, sell a subscription, demonstrate expertise or establish a relationship that pays later.

This bargain was never a guarantee of payment. A search result could answer a simple question, a snippet could satisfy curiosity, and a visitor could leave without producing meaningful revenue. Still, the destination page was often the place where the reader encountered the full product and the producer's terms.

An answer interface moves more of that encounter into the intermediary. It chooses which sources to retrieve, which propositions to retain, how to combine them, whose name to display and how much reason remains to leave. The publisher can become an input into a product whose presentation and audience relationship it does not control.

Four events must therefore stay separate: a source is used; it is credited; a reader visits; the visit produces enough contribution to support further work. None automatically entails the next. A citation can provide accountability without attracting a visit. A visit can provide attention without revenue. Revenue can be positive while still falling short of production costs.

The reverse distinction is equally important. Fewer visits can be economically acceptable if the remaining readers are more likely to subscribe, buy or commission work. A specialist consultancy may value one qualified inquiry more than thousands of casual page views. A public reference project may judge success through reach and usefulness rather than advertising yield.

This is why a single number called “AI visibility” cannot answer a publisher's business question. Visibility is an intermediate result. The useful question is what changes in the producer's capacity to keep supplying reliable information.

The vulnerability is greatest when a source's value is easy to extract but expensive to produce. A short factual conclusion may rest on weeks of reporting or testing. Its length inside an answer is a poor guide to its replacement cost. Paying only for the amount of text displayed could reward cheap repetition while underfunding the expensive discovery that made the sentence possible.

Substitution can also look easier to the platform than it is to the wider information system. Several pages may repeat the same original finding. Replacing one cited URL with another keeps the answer working, but it does not necessarily replace the reporting behind both. Counting available pages can therefore overstate the number of independent suppliers. This is an analytical risk, not a measured feature of the study's corpus.

An agreement concerned with durable supply should ask where new information enters the chain. A reference page may contribute valuable synthesis and maintenance; an original investigation may contribute facts that every later summary relies on. Both can deserve support. A compensation scheme that sees only the final fetched page may miss that distinction, just as a citation that points to a retelling may conceal who first established the claim.

FIGURE 02 CONCEPTUAL MECHANISM · NOT CONVERSION DATA

Four steps between use and funding.

An answer can use a source without completing the economic relationship that supports its next piece of work.

01

01

Source used

Information enters the answer. Use does not guarantee visible attribution.

02

02

Source credited

The producer is identified. A citation does not guarantee a visit.

03

03

Reader visits

Someone reaches the source. A visit does not guarantee contribution.

04

04

Work funded

Revenue or other support can help pay for the next investigation or update.

Licensing or paid access can create another path from use to funding. No transition rate or flow volume is implied.

Kymata Labs analytical framework; attribution research and documented access/referral mechanisms. [2] [4] [5]

A footnote does not complete the transaction

Attribution has at least three jobs: identify the producer, let the reader check a claim and offer a route to the original work. A source link can help with all three. It cannot guarantee any of them.

The 2023 research repository accompanying Nelson F. Liu, Tianyi Zhang and Percy Liang's work on generative-search verifiability makes a useful distinction. Its released annotations separately record whether a statement merits verification, whether the union of citations supports it and whether each individual citation provides complete, partial or no support. It also records perceived usefulness separately. [2]

That separation should become a product requirement. An answer that feels helpful may still attach the wrong evidence to a sentence. A prestigious source can be real while failing to support the wording attributed to it. Even accurate attribution may fail to show which organization originated a finding and which merely repeated it.

The 2026 Overview study supplies a more recent, bounded illustration. Its saved model judgments classified 87,204 of 98,020 claim judgments as consistent with captured source text and 10,816 as inconsistent: 88.97% and 11.03%. The verification cohort covered 7,491 Overviews after exclusions. These were automated assessments relative to supplied text, using categories that separate contradiction, omission and ambiguity. They are not a human-audited estimate that 11% of all Google answers are false. [1]

For publishers, attribution errors have an economic dimension as well as an epistemic one. The producer may provide the underlying work but fail to receive recognizable credit. Or it may receive apparent credit for a formulation it never published. Neither outcome builds a dependable relationship with a future reader.

Optimization does not remove this distinction. The GEO research project reports methods that increased source visibility in its generative-engine evaluation, with improvements up to 40% under the project's measures. That is a claim about visibility in generated responses. It is not evidence of a 40% gain in visits, subscribers or income. [3]

Making material easier to retrieve and cite can be sensible. Treating citation share as a financial result merely moves the uncertainty into a more flattering dashboard.

A small publisher's arithmetic

Consider a deliberately hypothetical publication. It receives 100,000 referred visits in a month. Each visit produces an average contribution of $0.10 after the variable costs associated with that visit. The referred audience therefore contributes $10,000 toward reporting, editing and other fixed costs.

Suppose those fixed monthly costs are $9,000. The operation has a $1,000 cushion. This is a simplified business, not a description of a real publisher, an estimate of industry advertising rates or a forecast of AI's effect.

Now suppose referred visits fall to 60,000 while their average contribution improves to $0.12. The remaining audience is 20% more valuable per visit. Its total contribution is nevertheless $7,200. The publication is $1,800 short of the assumed fixed costs and $2,800 below its previous contribution.

“Better traffic” is a meaningful counterargument only after this multiplication. A higher conversion rate does not automatically offset a smaller audience. In this scenario, contribution per remaining visit must rise to $0.15 merely to cover the $9,000 cost base. It must reach approximately $0.1667 to restore the original $10,000 contribution.

There are several ways to fill the gap. An additional $2,800 in net monthly licensing contribution restores the baseline. So would 20,000 additional direct visits contributing $0.14 each, provided those visits are genuinely additional and their acquisition costs are already accounted for. Net compensation of $1,800 covers the assumed costs but leaves no original cushion. The business implications differ even when every option is described as compensation.

The general calculation is simple. Let baseline referred visits be V and contribution per visit be m. Let r be the fraction of those visits retained and q the multiplier on contribution per retained visit. Let L be net additional licensing income and D net additional contribution from direct relationships. The new contribution is:

$$V \times m \times r \times q + L + D$$

Preserving the baseline requires that expression to reach V × m. With no additional compensation or direct contribution, the threshold is r × q = 1. Retaining 60% of visits requires a contribution multiplier of about 1.67. Retaining 80% requires 1.25. These are arithmetic thresholds under stated assumptions, not empirical predictions.

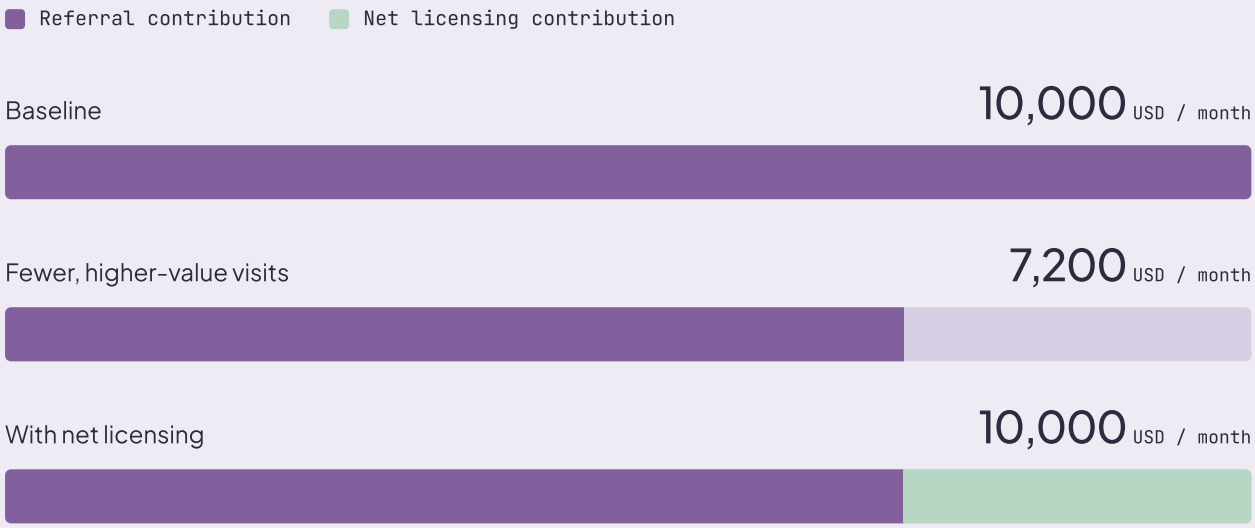
The model deliberately holds production costs fixed. That is appropriate for asking whether the same work can continue. Cutting reporting until the accounts balance may keep a company alive while reducing what the information system receives. A cost saving and a sustainable replacement for lost funding are different outcomes.

It also leaves out long-term subscriber value unless that value is consistently included in m and q. Counting future subscription value in the new channel while valuing the old channel only by its first advertisement would manufacture an improvement. The time horizon, cost allocation and attribution rule must match.

FIGURE 03 HYPOTHETICAL PUBLISHER · NOT OBSERVED TRAFFIC

Better traffic still has to pay the bill.

Holding the work constant, better contribution per visit may still leave a gap when fewer visits arrive.



01

100,000 × $0.10

Baseline

Monthly referred visits × contribution per visit = $10,000.

02

60,000 × $0.12

Fewer, higher-value visits

Contribution improves 20% per visit, but the total falls to $7,200.

03

+$2,800 net

Restore the baseline

Additional net licensing contribution restores $10,000; $1,800 only covers the assumed $9,000 fixed cost base.

All visits, rates and costs are invented. Contribution is after visit-related variable costs. Compare consistent horizons and charge acquisition costs.

Original Kymata Labs scenario. With no other contribution, retaining 60% of visits requires about 1.67× contribution per visit to restore the baseline.

The strongest case for the answer engine

The defense of AI search deserves more than a paragraph about convenience. Readers do not owe a website an unnecessary visit. They want an explanation, a comparison or a route through unfamiliar material. If an answer saves time and helps them understand the subject, it has created value.

Some source pages are cumbersome, repetitive or built around delaying the answer. Preserving every existing page-view opportunity would preserve those incentives too. A serious argument for sustainable information cannot depend on making discovery worse so that an incumbent advertising model remains

untouched.

Answer systems can also widen discovery. In the 2026 study, mean overlap between Overview citation domains and the full captured search-result links was 70.2%. The complement, 29.8%, describes the average share of cited domains absent from those captured links within an observation. It is not a measured increase in publisher reach, but it shows why an answer need not simply reproduce the visible ranking. [1]

The source-quality results also resist an easy condemnation. On the study's available credibility scores, cited sources scored higher than the full captured result-link set. Yet the advantage changed with the comparison: in query-paired analyses, Overview source averages were lower than the top one, three or five captured links and higher than the top ten or full set. The scores covered only part of the sources, and the captured links were not a separately classified list of organic results. “AI picks better sources” is therefore as incomplete as “AI always degrades search.” [1]

More valuable referrals are plausible too. An answer may help a reader decide which expert to consult before sending them there. The resulting visit can arrive with clearer intent. An information producer with a subscription, software product or advisory service may benefit even if casual visits decline. That outcome should count as success when it appears in matched financial data.

Finally, traffic is not the only possible payment system. A platform can purchase access, license an archive, pay for updates or share revenue. If those arrangements fund the underlying work, there is no economic principle requiring every reader to traverse the original webpage.

The case against extraction is strongest when it accepts all of this. The question is whether the value created by the new interface returns through a mechanism large, durable and broad enough to maintain its necessary sources. Reader benefit is evidence that a product is useful. It is not, by itself, evidence that the supply chain is funded.

Charging for access is a mechanism, not a settlement

Cloudflare's documentation describes a concrete alternative to free crawling. Its Pay Per Crawl feature lets a site owner set a price for crawler access. A crawler can present payment intent in request headers, or receive an HTTP 402 Payment Required response containing pricing. Cloudflare acts as merchant of record. The inspected documentation labels the feature a closed beta. [5]

This is a useful change in the available tools. It connects a request for content with a potential payment instead of assuming that later referral traffic will compensate the producer. But a documented payment mechanism is not evidence of widespread participation, a sufficient price or sustainable publisher income.

A crawl is also not the same unit as an answer, an audience or an original discovery. One fetched page can inform multiple subsequent uses; one generated answer can combine many pages. A flat request price may be administratively simple while remaining a rough proxy for the value exchanged. Caching, updates, contractual permissions and repeat use all affect the arrangement.

Licensing raises a related distribution problem. A deal that works for a large archive need not reach a small specialist or the local source whose reporting others repeat. Compensation based only on existing scale can reinforce the organizations that already have bargaining power. A sustainable market must leave room for useful new entrants and costly, narrow reporting.

There is no universal payment formula in the evidence reviewed here. A sensible agreement would specify what is being bought: retrieval, training, display, quotation, reuse, freshness or some combination. It would also specify the evidence a producer can inspect, the treatment of corrections, the duration of permission and the conditions for exit.

These details are not peripheral administration. They determine whether the producer can price the service and continue investing in it. An opaque promise that access creates exposure is the old uncertainty in contractual clothing.

Publisher controls need separate decisions

Blocking a crawler is not a complete publishing strategy. Neither is allowing every crawler in the hope that visibility will eventually pay. The useful decision depends on the operator, purpose, content and return.

Cloudflare's crawler-management documentation offers allow, block and charging controls. Its examples explicitly recognize that citations, referrals or existing agreements can provide value. Its analytics documentation separately describes crawler requests and referral visits. That separation is a good operational starting point: the machine collecting a page and the person arriving from an answer are different events. [4][6]

Purpose matters too. The inspected managed-robots documentation distinguishes search indexing, real-time AI input and model training. Its Content Signals vocabulary names these as search, ai-input and ai-train. These are useful categories for expressing preferences; their existence should not be mistaken for a universal legal ruling or a guarantee that every service will honor them. [7]

Even compliance dashboards require interpretation. Cloudflare warns that its robots.txt violation table compares past requests against current directives. A newly changed rule can therefore cause an earlier legitimate request to appear as a violation. A publisher should preserve the policy that applied at the time before accusing a service of disregarding it. [8]

The practical aim is to make decisions reversible and observable. Begin with a defined set of pages, record crawler access and human referrals, note the rule change, and inspect downstream outcomes. A publisher with a strong direct audience may make a different choice from a new reference site seeking recognition. The purpose of controls is to express that strategy, not to substitute for one.

Measure the relationship you want to preserve

Publishers need a ledger that joins production to distribution. For a stable group of pages, record their topic, original publication date, substantive updates, production cost and revenue role. Separate original reporting from summaries of others' work. Otherwise, a site can appear financially healthy while the

expensive work underneath it quietly disappears.

Then connect four kinds of evidence. Access logs show which known crawlers requested material. Answer observations show where and how it was used or credited. Referral records show identifiable human visits. Financial records show subscriptions, advertising contribution, sales or other outcomes. Each has missing data; none should silently stand in for the others.

A before-and-after traffic chart is insufficient to attribute a change to AI. Seasonality, ranking updates, competitor launches, publication frequency and changing demand can all move the same line. Compare consistent page cohorts, record material changes and use a credible comparison group where possible. Report the remaining ambiguity instead of giving a confounded estimate a causal headline.

Platform operators can make this analysis less speculative. They can provide source-level usage reporting, distinguish citation display from reader action, preserve clear links to original work and offer corrections that propagate when source material changes. Compensation systems should report their unit of payment and the period in which it was earned. Those are auditable features, not public-relations promises.

Readers have a smaller but real role. Opening the original source is useful when context, qualifications or consequential decisions matter. Supporting a publication directly can fund work that many people encounter elsewhere. The point is not to turn every quick question into a moral test. It is to recognize that effortless consumption does not make production effortless.

The scarce input is the next piece of work

The deepest economic question is not what an answer engine owes for yesterday's page. It is what makes tomorrow's page worth producing.

A publisher can maintain an archive while reducing original reporting. An expert can leave a useful tutorial online while deciding not to update it. A local institution can continue to exist while losing the capacity to investigate difficult subjects. Those changes would not necessarily appear immediately in citation counts or answer-quality demonstrations.

This is a prospective incentive argument, not a claim that the reviewed studies measured such deterioration. It also differs from the argument about models training on synthetic material. The concern here is whether people are paid to discover, verify and maintain information in the first place.

Better answers are worth building. So are better ways of paying for their inputs. The responsible test is whether reader value and continuing source production can rise together. Where referrals suffice, show the numbers. Where they do not, make the replacement explicit.

The answer may be free at the point of use. The next fact still has a production bill.

Evidence and method

Research access was recorded on November 16, 2026 UTC (November 15 in Los Angeles). The empirical search findings above come from the authors' public repository, saved notebook outputs and inspected analysis code at release commit b622031334fe875f8f30d669691b3a0710a5cb57. We checked definitions and arithmetic; we did not independently rerun the complete raw-data pipeline, perform new search crawls or purchase model calls. The study's trends-derived query population is not representative of all searches. Its automated fidelity judgments assess captured-source support, not universal truth.

The other sources provide historical research definitions and current vendor documentation. Documentation establishes described capabilities, not adoption or revenue outcomes. The publisher scenario is entirely hypothetical. No publisher interview, proprietary traffic result, measured industry revenue loss or current demand trend is claimed.

Questions worth asking.

Does this paper prove AI Overviews reduce publisher revenue?

No. The inspected 2026 study measures answer activation, sources, captured-source support and advertising exposure. It does not measure a causal publisher click or revenue loss. The economic model is explicitly hypothetical.

Can fewer visits still be better for a publisher?

Yes, if each remaining visit contributes enough value, or another income source fills the gap. Retaining 60% of visits requires about 66.7% more contribution per visit to preserve the old total without additional income. Use the same cost rules and time horizon for both channels.

Is getting cited in AI answers enough?

A citation can provide recognition and a route to verification. It does not guarantee that the claim is supported, a reader visits, or the publisher earns revenue. Track those outcomes separately.

Can publishers charge AI crawlers?

Cloudflare documents Pay Per Crawl in closed beta, with pricing and payment-intent or HTTP 402 flows. That establishes a possible mechanism; it does not establish universal access to the feature, widespread payment or sustainable publisher income.

Should publishers block AI crawlers?

Decide by service, purpose and measured return. Separate indexing, real-time answer retrieval and training; record the rules in effect and compare crawler access, referrals and financial contribution. A blanket rule may fit some publishers and undermine discovery for others.

SOURCES, NOT SCENERY

References.

1. **Measuring Google AI Overviews: Activation, Source Quality, Claim Fidelity, and Publisher Impact — released research repository**
   Haofei Xu, Umar Iqbal and Jacob M. Montgomery; IMC 2026 · 2026; repository release 2026-11-04 · Primary research: original code, saved notebook outputs and study documentation

   DOI 10.1145/3777912.3839818. March 13–April 21, 2026; 55,393 query executions, 7,583 Overviews, 19 US Google Trends categories. Inspected README, scraper/verifier documentation and notebooks01,03–07. Full raw pipeline not independently rerun. Does not measure publisher click or revenue loss.

   https://github.com/haofeix/google-aio-characterization/tree/b622031334fe875f8f30d669691b3a0710a5cb57

2. **Evaluating Verifiability in Generative Search Engines — accompanying material**
   Nelson F. Liu, Tianyi Zhang and Percy Liang · 2023 · Primary research artifact

   Inspected original README and acquired public human-evaluation annotations. Used for explicit distinctions among usefulness, statement support and individual-citation support; no current performance claim.

   https://github.com/nelson-liu/evaluating-verifiability-in-generative-search-engines

3. **GEO: Generative Engine Optimization — research repository**
   Pranjal Aggarwal, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, Karthik R Narasimhan and Ameet Deshpande · 2023 research; accessed 2026-11-16 · Primary research artifact

   Original README reports source-visibility improvements up to 40% under project evaluations. This is not an estimate of traffic, conversion or revenue gains. Original README and project-page source were inspected; the evaluation was not rerun and no paid model calls were made.

   https://github.com/GEO-optim/GEO

4. **Analyze AI traffic**
   Cloudflare · Undated documentation; accessed 2026-11-16 · Official product documentation

   Inspected official cloudflare/cloudflare-docs production source. Distinguishes crawler requests from referral visits; documentation is capability evidence, not measured adoption or publisher outcomes.

   https://github.com/cloudflare/cloudflare-docs/blob/production/src/content/docs/ai-crawl-control/features/analyze-ai-traffic.mdx

5. **What is Pay Per Crawl?**
   Cloudflare · Undated documentation; accessed 2026-11-16 · Official product documentation

   Inspected official repository source. Explicitly closed beta; price per zone; payment-intent headers or HTTP 402 pricing response; Cloudflare merchant of record. No revenue/adoption inference.

   https://github.com/cloudflare/cloudflare-docs/blob/production/src/content/docs/ai-crawl-control/features/pay-per-crawl/what-is-pay-per-crawl.mdx

6. **Manage AI crawlers**
   Cloudflare · Undated documentation; accessed 2026-11-16 · Official product documentation

   Inspected official repository source describing allow, block and charge options and possible value from citations, referrals or agreements.

   https://github.com/cloudflare/cloudflare-docs/blob/production/src/content/docs/ai-crawl-control/features/manage-ai-crawlers.mdx

7. **Managed robots.txt — Content Signals Policy**
   Cloudflare · Undated documentation; accessed 2026-11-16 · Official product documentation

   Inspected official repository source. Describes search, ai-input and ai-train use categories. Signals are not evidence of universal compliance or a legal conclusion.

   https://github.com/cloudflare/cloudflare-docs/blob/production/src/content/docs/bots/additional-configurations/managed-robots-txt.mdx

8. **Track robots.txt**
   Cloudflare · Undated documentation; accessed 2026-11-16 · Official product documentation

   Inspected official repository source. Explicitly warns violation table compares current directives with past requests, which can flag previously legitimate access.

   https://github.com/cloudflare/cloudflare-docs/blob/production/src/content/docs/ai-crawl-control/features/track-robots-txt.mdx