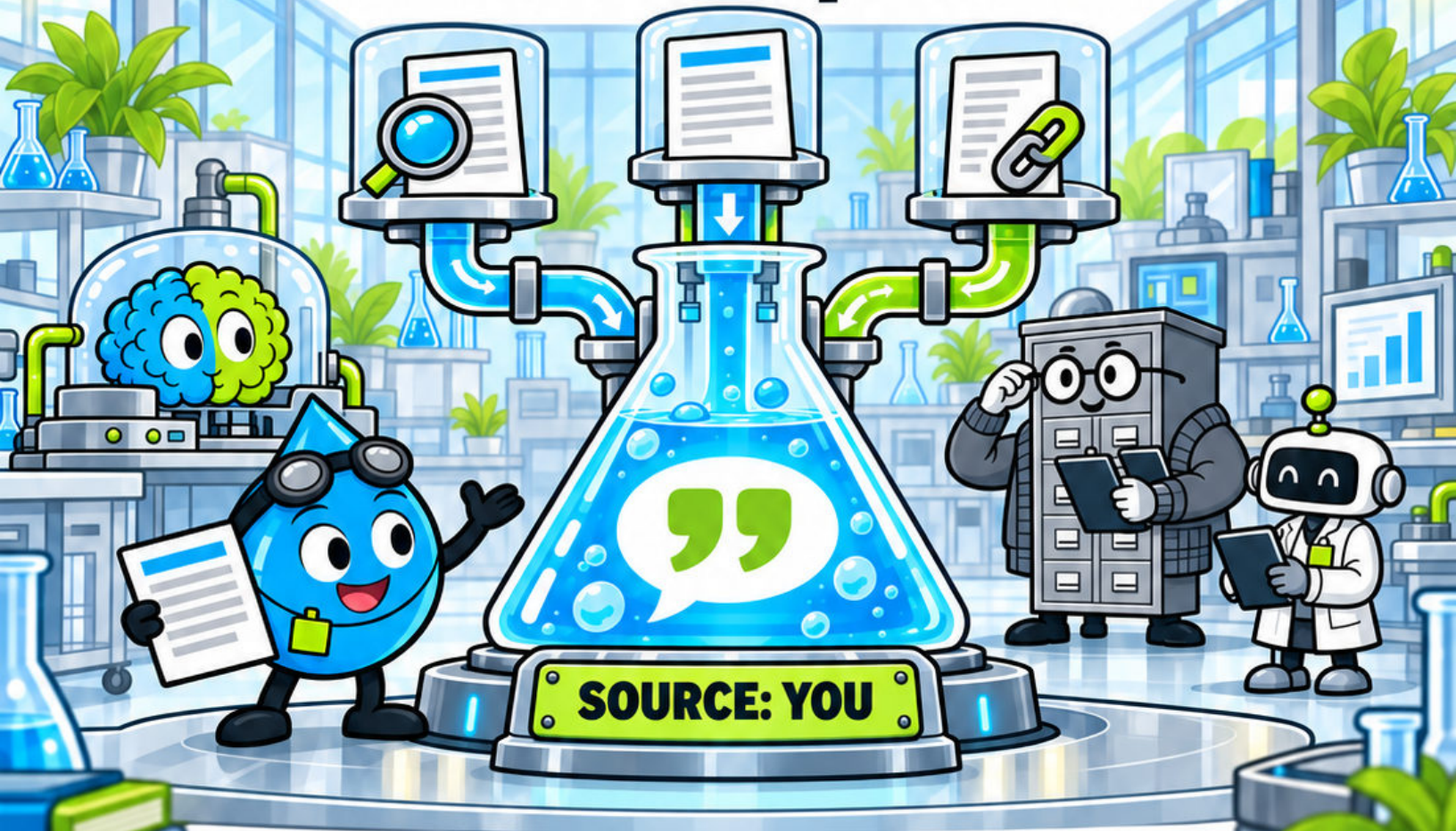


2026 FIELD MANUAL

THE COMPLETE GUIDE TO GEO

**Generative Engine Optimization
for SEO Experts**



**Inside GEO: How AI Finds,
Understands and Cites Content**

JOURDAN ROMBOUGH
OPTIMIZATION THEORY

VERSION 2.4

CONTENTS

START HERE

Why I Built This Lab	3
How to Use This Guide	4
Meet the Lab Crew	5
GEO in 60 Seconds	6
How an AI Answer Gets Made	7
Anatomy of a Citable Chunk	8
Lab Map	9

HOW GEO WORKS

● Room 01: The Library Floods	10
The Click Drain	11
The New Engine Room	12
SEO vs. GEO: What Changes	13
SEO Is the Power Plant	14
● Room 02: Inside the Answer Machine	15
Anatomy of an AI Engine	16
The Answer Pipeline: Part 1	17
The Answer Pipeline: Part 2	18
Query Fan-Out	19
Passages Need Context	20
The Crawl-to-Click Gap	21
Engines Consult Different Shelves	22
● Room 03: The Periodic Table of GEO	23
The Experiment That Started It All	24
From Vague to Specific	25
Do AI Answers Prefer Fresh Pages?	26
Busted Experiments	27

DOING GEO

● Room 04: Brand DNA	28
The Entity Molecule	29
The DNA Match Test	30
Repair Wrong Answers	31
Fill Real Decision Gaps	32
● Room 05: The Technical Airlock	33
The Badge Wall	34
Clear the JavaScript Fog	35
Specimen Labels	36
Different Files, Different Jobs	37
Markdown for Agents	38
Markdown: What the Evidence Says	39

● Room 06: Growing Cultures	40
Web Mentions and AI Visibility	41
Help the Community	42
Useful Video	43
Evidence to Coverage	44
● Room 07: Hazard Zone	45
Why Formulas Don't Travel	46
The Clean Room Pledge	47

MEASURING GEO

● Room 08: Under the Microscope	48
The Prompt Test Rack	49
The Gauge Panel	50
Built in This Lab	51
● Room 09: The 90-Day Protocol	52
Audit, Then Build	53
Grow, Measure, Repeat	54

THE TACTICS BOARD

The Tactics Board	55
Every Tactic at a Glance	56
Files, Formats and Access	57
Writing Tactics	58
Off-Site and Paid Tactics	59
The Black-Hat File	60
The Black-Hat File: Case by Case	61

FROM MY OWN LAB

90 Days of AI Traffic on vuss.io	62
----------------------------------	----

TOOLS AND REFERENCE

The Perfect GEO Page	63
The GEO Lab Master Checklist	64
Field Kit: Robots Policy	65
Field Kit: Entity Graph	66
Field Kit: Prompt Panel	67
Field Kit: GA4 AI Traffic	68
Field Kit: Markdown Alternates	69
Field Kit: Test Your Markdown Setup	70
Glossary and Lab References	71
Chunk Made It. Will Yours?	72
Source Drawer: 1 of 4	73
Source Drawer: 2 of 4	74
Source Drawer: 3 of 4	75
Source Drawer: 4 of 4	76

WHY I BUILT THIS LAB



Hi, I'm Jourdan.

Technical SEO for more than 15 years

Founder, Optimization Theory, Cooperstown, NY

Builder of AI Citation Trace, PageRank Simulator and Black Hat SEO Terminal

linkedin.com/in/jourdanrombough

I've worked in technical SEO for more than 15 years. For most of that time the goal was easy to explain: help the right page rank, earn the click, and give the visitor what they came for.

Then the answers started arriving before the links. People now ask ChatGPT, Gemini, Perplexity or Google's AI Mode and get a written reply, sometimes with a source attached and sometimes without one. I wanted to know how those replies are built: where the facts come from, why one page is quoted and another is passed over, and what a site owner can honestly do about it.

So I read the research, built small tools to test ideas, and checked my own sites. This guide is the result: GEO explained one piece at a time, with pictures, plain language, and a named source behind every number.

If you already know SEO, much of this will feel familiar, and the differences will stand out. Start at the front and go room by room. By the end you'll know how AI answers choose their sources, and what makes a page worth choosing.

Jourdan Rombough

Version 2.4 | October 30, 2026



**JOURDAN BUILT THE LAB. I'M THE
PASSAGE YOU'LL FOLLOW THROUGH IT.**

HOW TO USE THIS GUIDE



Go room by room. Each room explains one part of GEO.

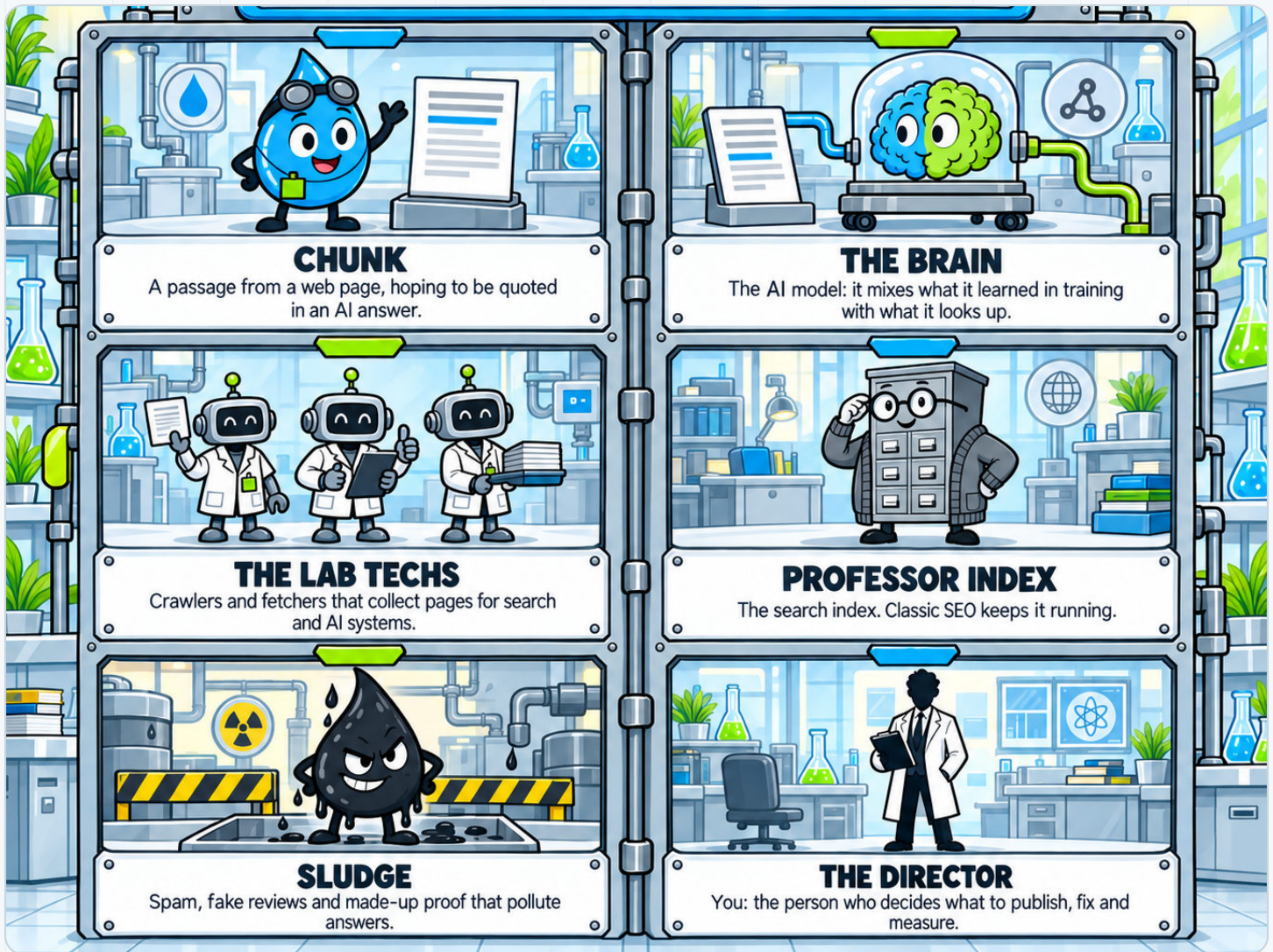
Green checks mark practices that help. Use them.

Yellow flags a myth or a risky shortcut. Skip it.

Every number names its study. All sources: Source Drawer.

Follow Chunk, a passage hoping to be quoted.

MEET THE LAB CREW



WHO THEY ARE IN REAL LIFE

- **CHUNK** is a passage on your page: a paragraph, list item or table row an engine can lift and quote.
- **THE LAB TECHS** are crawlers and fetchers such as Googlebot, Bingbot, OAI-SearchBot, ChatGPT-User, Claude-SearchBot and PerplexityBot.
- **SLUDGE** is spam, fake reviews, hidden prompts and cloaking. Its case file is the Black-Hat File on page 60.
- **THE BRAIN** is the language model behind ChatGPT, Gemini, Claude or Copilot, answering from training and from what it looks up.
- **PROFESSOR INDEX** is the search indexes answers draw on, such as Google's and Bing's. A page that isn't indexed is hard to retrieve.
- **THE DIRECTOR** is you: choose the questions, fix the pages and run the tests. Room 09 is the plan.

GEO IN 60 SECONDS

GEO (generative engine optimization) is the work of making your content easy for AI answer engines to find, understand, trust and cite. Six ideas cover most of it:

01 Start with the task

People ask AI tools to help them decide things: which drone to buy, which software to choose, which agency to hire. Find the questions your audience really asks. A page that directly helps with one gives an answer engine something to use.

[Room 01: The Library Floods, page 10](#)

02 Keep the route open

Before an AI system can quote a page, something has to fetch and read it. Check that the crawlers you want are allowed, pages load without errors, and key facts sit in the HTML. A blocked page can't be cited from a fresh visit, however good it is.

[Room 05: The Technical Airlock, page 33](#)

03 Make the evidence useful

Answers are stitched together from passages. Write ones that make sense alone: name the subject, answer plainly, add the number or example that proves it, and say where it came from. In the GEO benchmark, adding quotes, statistics and citations raised a simulated engine's visibility score.

[Room 03: The Periodic Table of GEO, page 23](#)

04 Keep identity coherent

Engines piece together who you are from your site, profiles, directories and press. When names, addresses or services disagree, answers can blur you with someone else or repeat old facts. Keep one approved set of facts and make every profile match.

[Room 04: Brand DNA, page 28](#)

05 Earn corroboration

AI answers often cite sources you don't control: reviews, Reddit, YouTube, news and research. You can't buy that trust. You can earn it by publishing findings people can check and by helping where your audience already talks.

[Room 06: Growing Cultures, page 40](#)

06 Measure the right outcome

One screenshot of an AI answer proves little; answers change from run to run. Track a fixed set of questions over time and count how often you're named, cited, accurate and visited. A mention isn't a click, and a click isn't a sale.

[Room 08: Under the Microscope, page 48](#)

First step: write down ten questions your audience asks and record how AI answers handle them today, before you change anything.

HOW AN AI ANSWER GETS MADE

Ask ChatGPT, Gemini, Perplexity or Google's AI Mode a question with search turned on, and several steps happen before the reply appears. Products differ, but most search-backed answers follow a pattern like this one.



1 ASK

A person asks a question.

2 SEARCH

The system runs several related searches.

3 RETRIEVE

It gathers candidate pages and passages.

4 SELECT

It keeps the passages that fit best.

5 WRITE

The model writes an answer from them.

6 CITE

Some sources get a visible link.

WHERE GEO FITS

BE FINDABLE

Stages 2 and 3. Pages that crawlers can reach and search can index, matching the sub-questions people ask. Rooms 02 and 05.

BE THE BEST PASSAGE

Stage 4. Self-contained passages with a named subject, a direct answer, evidence and a source. Room 03.

BE DESCRIBED RIGHT

Stages 5 and 6. Consistent facts about you across the web, and independent sources that back them up. Rooms 04 and 06.

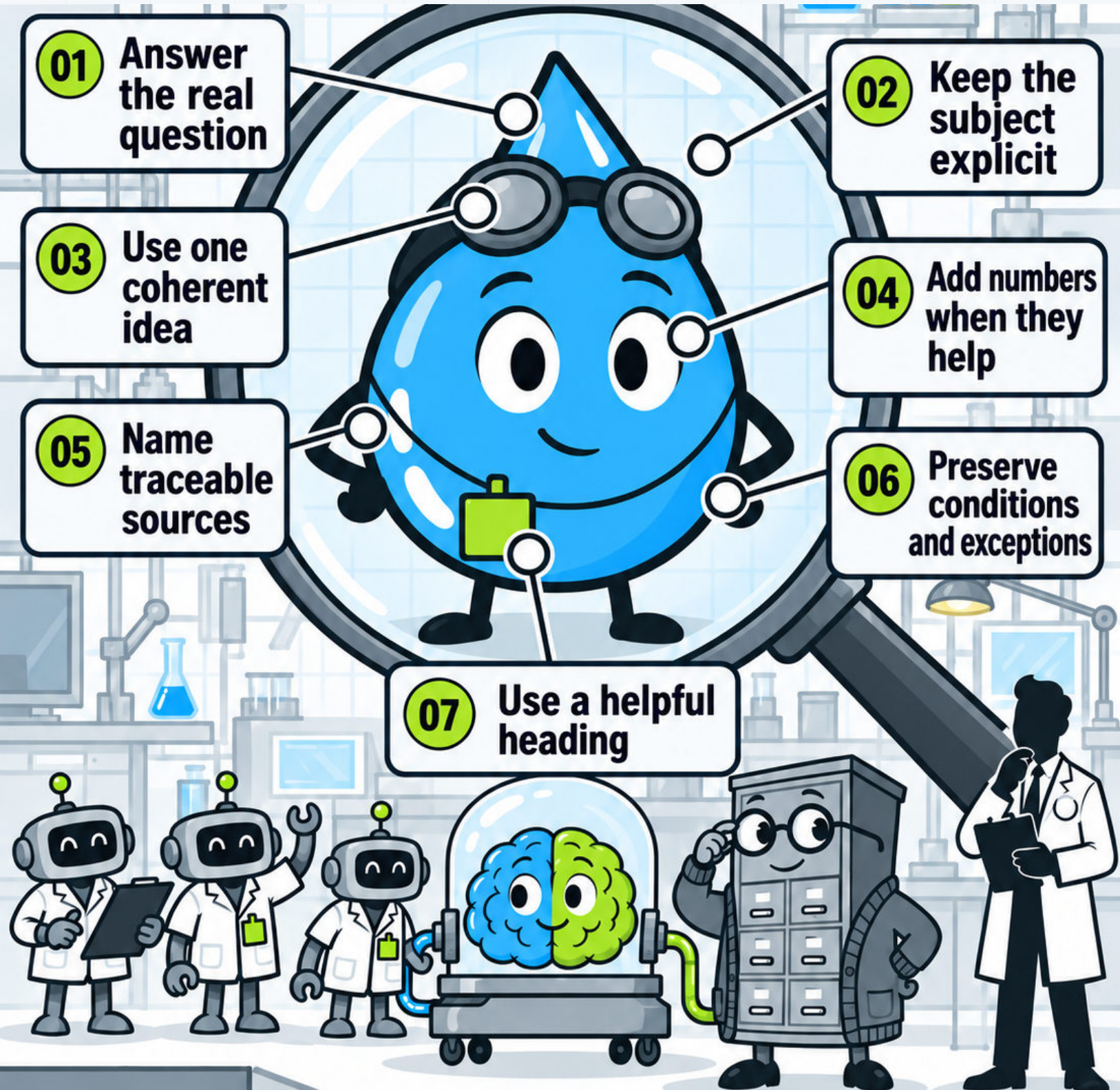


**I CAN'T WRITE THE ANSWER.
I CAN BE THE BEST EVIDENCE IN IT.**

A simplified pattern. Some products answer from training alone, repeat steps, or show no links. Room 02 (page 15) walks through each stage.

ANATOMY OF A CITABLE CHUNK

AI answers rarely quote a whole page. They lift a passage, a "chunk", that makes sense on its own. These seven traits make a chunk easy to understand, check and cite. Google says its AI features need no content written or "chunked" specially for AI: these are simply traits of clear writing, and there is no required passage length.



LAB MAP



01 The Library Floods > p. 10

04 Brand DNA > p. 28

07 Hazard Zone > p. 45

02 Inside the Answer Machine > p. 15

05 The Technical Airlock > p. 33

08 Under the Microscope > p. 48

03 The Periodic Table of GEO > p. 23

06 Growing Cultures > p. 40

09 The 90-Day Protocol > p. 52

Every popular GEO trick, rated: page 55. Copy-ready field kits: page 65. Real traffic example: page 62.

THE LIBRARY FLOODS

01

WHAT'S HAPPENING

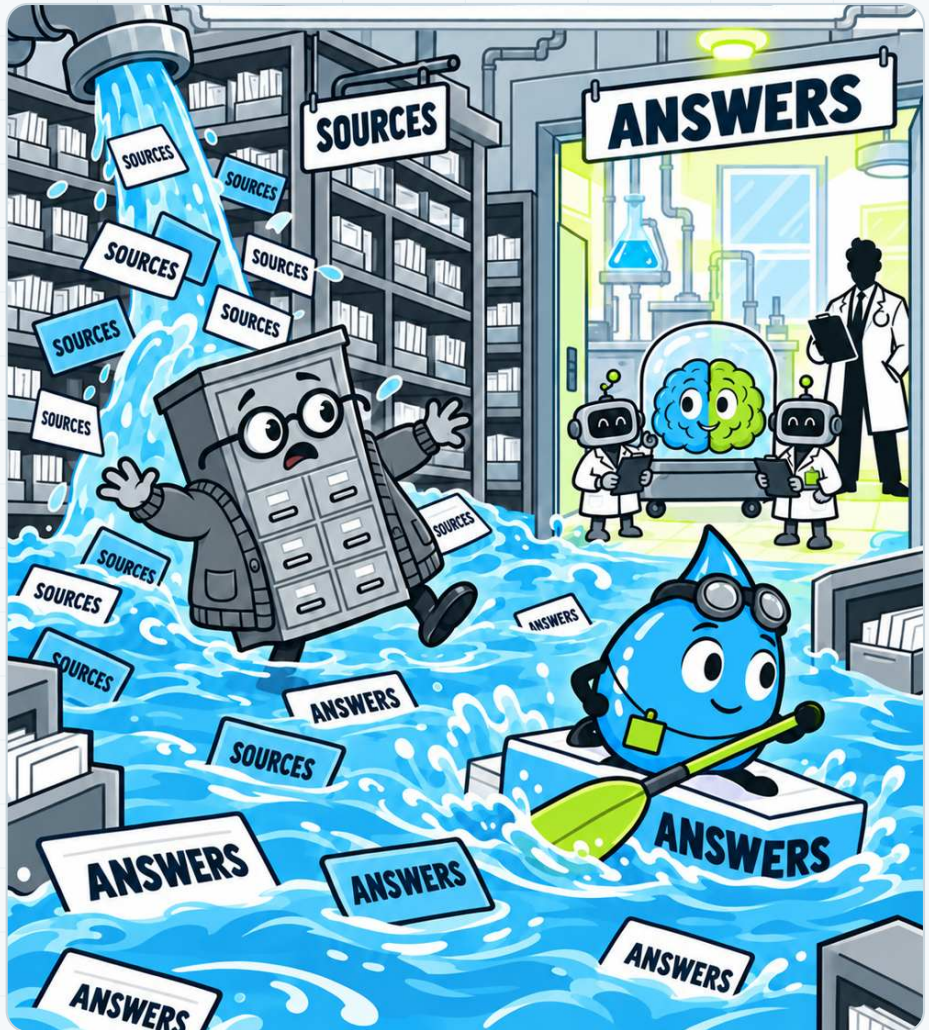
Search used to hand people a list of links. Now an AI answer often comes first: Google's AI Overviews and AI Mode, ChatGPT, Gemini, Perplexity. It summarizes several sources in one reply, and the reader may never scroll to the links.

WHY IT MATTERS

Ranking is no longer the only way to be seen, or a guarantee of a visit. To reach people now, a page has to be findable, useful enough to quote, and described accurately inside the answer.

JOURDAN'S TAKE

Rankings still matter. They just aren't the finish line anymore. Start with the decision your reader is trying to make, then check who's answering it for them.



ANSWERS NOW ARRIVE BEFORE THE LINKS. I WANT TO BE PART OF THEM.

IN THIS ROOM

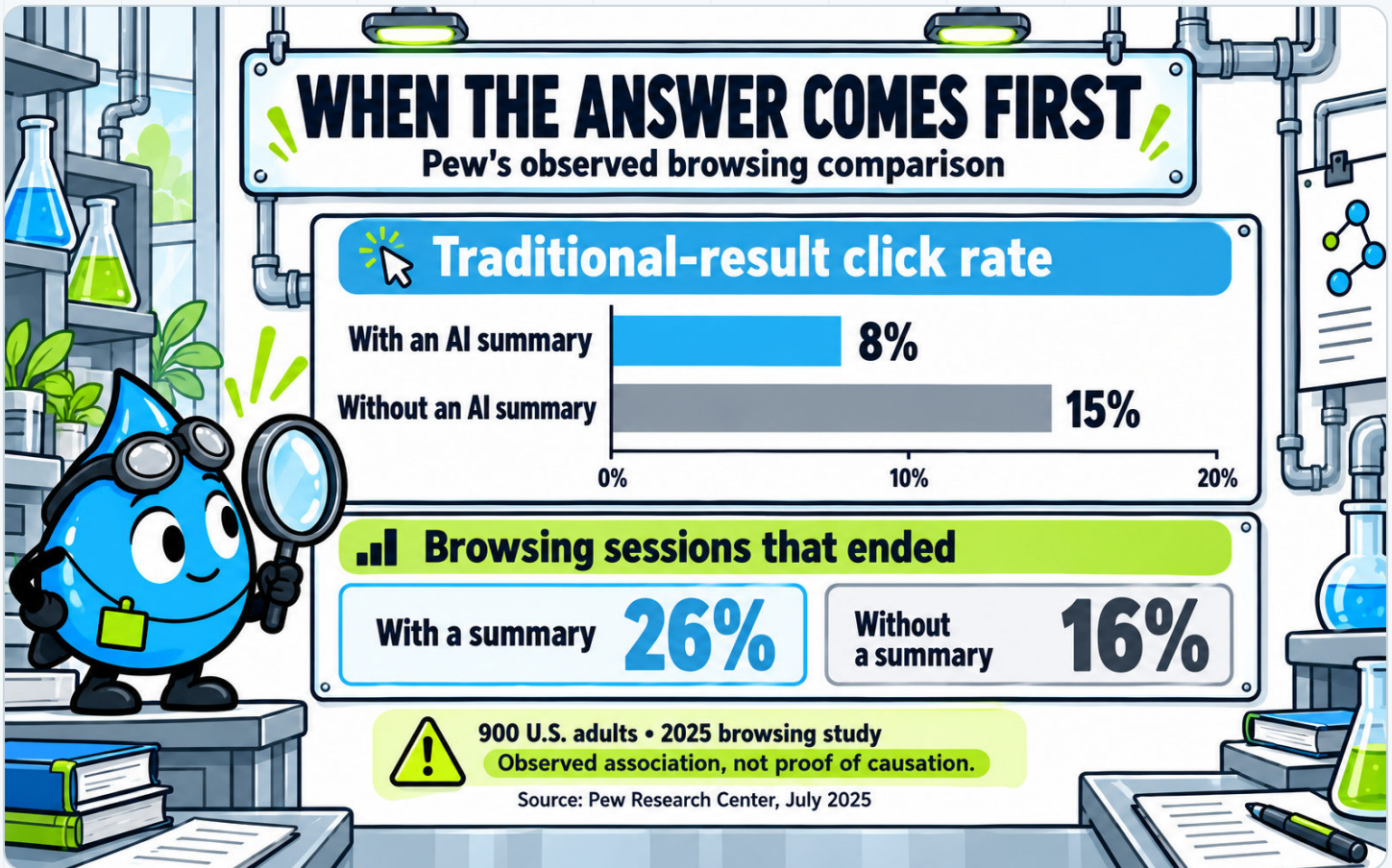
11 The Click Drain

12 The New Engine Room

13 SEO vs. GEO: What Changes

14 SEO Is the Power Plant

THE CLICK DRAIN



58% LOWER DESKTOP CTR

Ahrefs: position-one CTR with an AI Overview versus a modeled counterfactual. Not a measured causal effect.

1% CLICKED A SUMMARY SOURCE

Pew: visits to summary-bearing results pages that clicked a link inside the summary itself.



A CITATION CAN BE VISIBLE EVEN WHEN THE CLICK NEVER COMES.

Ahrefs: data Dec. 2025, published Feb. 4, 2026; 300K keywords; modeled desktop comparison. Pew: data March 2025, published July 22, 2025; 900 U.S. adults. Both observational.

THE NEW ENGINE ROOM

GOOGLE AI MODE

1B+

monthly active users

Company reported, July 2026

GEMINI APP

950M

monthly active users

Company reported, July 2026

CHATGPT

1.2B

weekly users

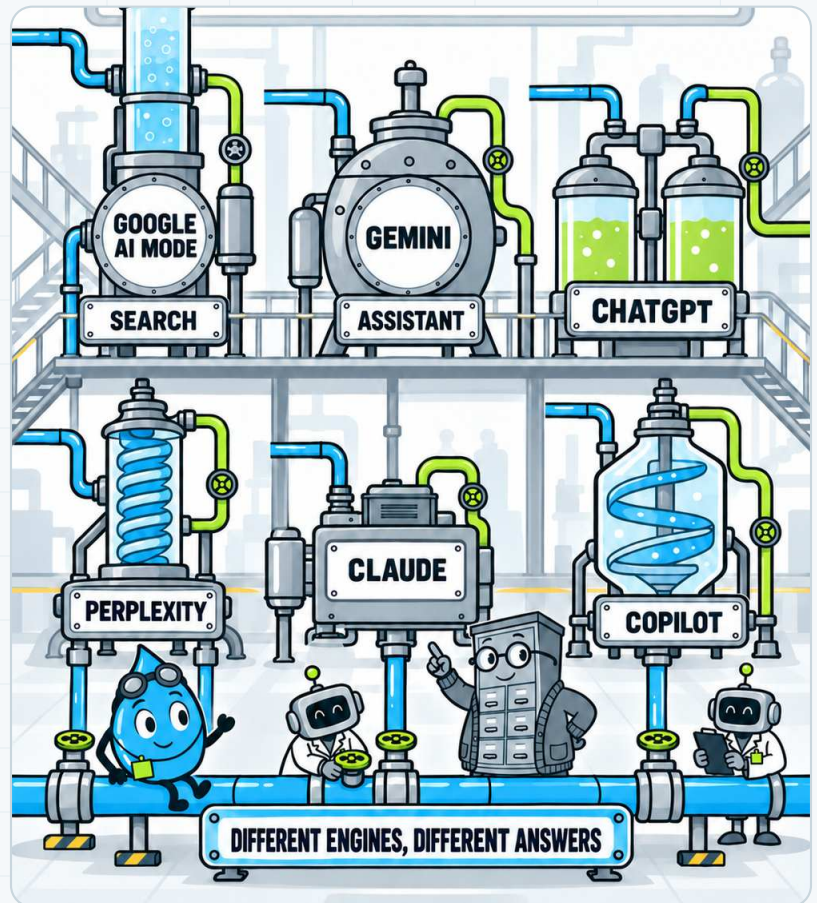
OpenAI DevDay 2026 recap, Sept. 29, 2026

ALSO INVESTIGATE

Claude, Perplexity and Copilot have their own audiences and search modes.

WHY IT MATTERS

Each product searches, selects and cites differently. Test the product and search mode your audience actually uses.



WHICH ENGINE IS MY AUDIENCE ACTUALLY USING?

*Different metrics (monthly vs. weekly users) and time windows, and the audiences overlap, so these numbers cannot be added together.
ChatGPT figure: OpenAI DevDay 2026 recap, Sept. 29, 2026.*

SEO VS. GEO: WHAT CHANGES



SEO ASKS

Crawl and index access

Search bots can reach and store the page.

Useful pages and passages

Content that answers what the searcher wants.

Queries and customer tasks

Keywords mapped to what people need.

Relevant links and reputation

Other sites vouch for you with links.

Search clicks and outcomes

Rankings, clicks and conversions.

Performance over time

Rank tracking on a fixed keyword set.

GEO ALSO ASKS

Can this AI surface consult the source?

AI fetchers and search modes have their own access rules.

Does the answer keep the meaning?

A passage quoted on its own must still be accurate.

Which details need evidence?

Conversational questions add conditions (budget, place, use) to address.

Who else backs up the claim?

Answers lean on independent sources: reviews, communities, press.

Were you mentioned, cited, or visited?

Three different outcomes. Count each one separately.

How much do repeated answers vary?

The same question can return different sources. Sample it repeatedly.

These overlap: SEO foundations power GEO, and GEO adds questions on top. Neither replaces the other.

SEO IS THE POWER PLANT

When an answer engine searches the web, it needs pages it can find, read and trust. Google's own guide says optimizing for its AI features is still SEO: the same foundations apply.

CRAWLABLE

Bots can reach the page: robots.txt allows it, the server answers 200, and no login or script wall hides the text.

INDEXABLE

Search can store and show it: no stray noindex, a sensible canonical, and the URL in your sitemap.

USEFUL

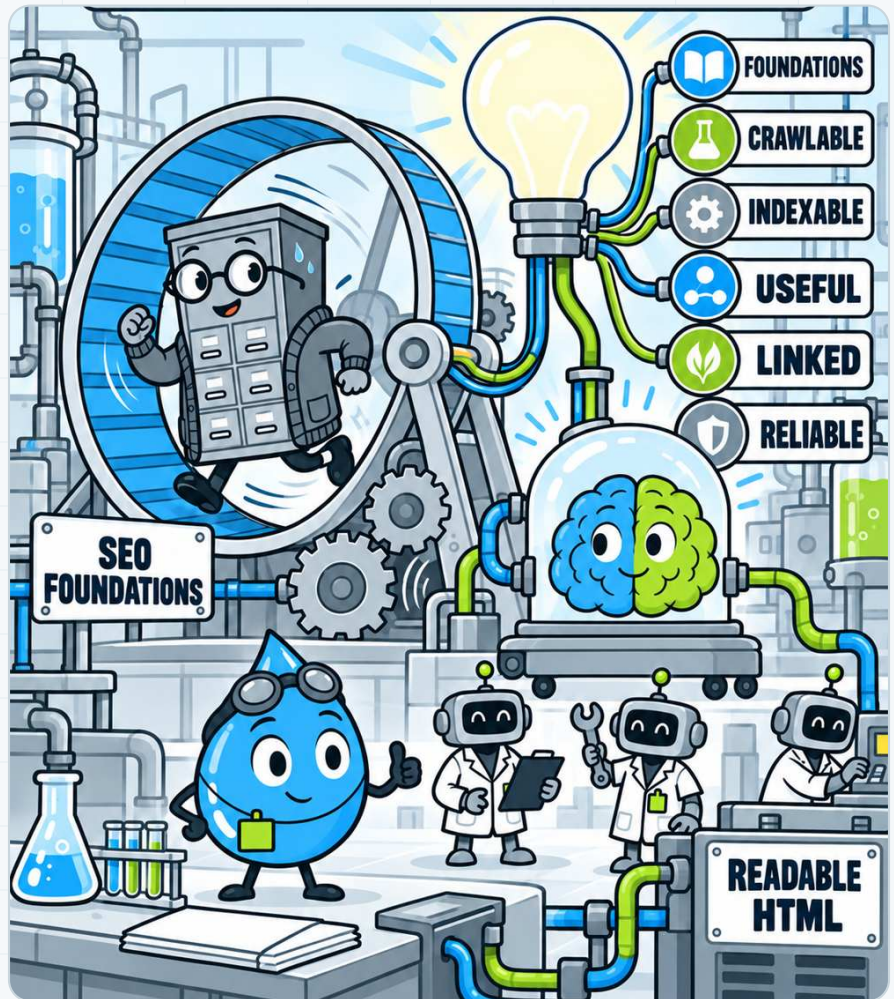
It answers a real question well, with the facts a person needs to decide or act.

LINKED

Other pages point to it, from your own site and from independent sources.

RELIABLE

Accurate, current, and consistent with your other pages and profiles.



THE BASICS THAT HELP SEARCH FIND ME ALSO HELP AI SEARCH FIND ME.

The same Google guide says AI-only files such as llms.txt aren't needed to appear in Google Search. Other answer engines document their own crawlers and rules (Room 05).

INSIDE THE ANSWER MACHINE

02

WHAT'S HAPPENING

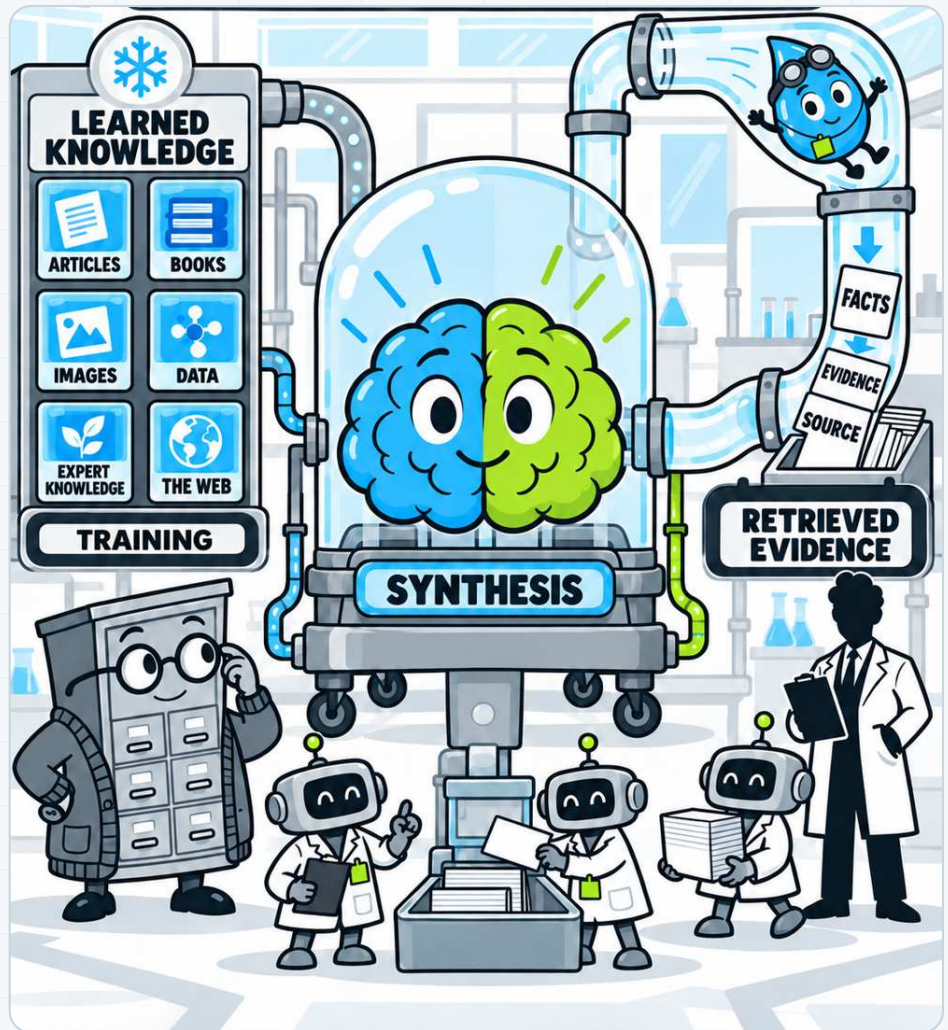
An AI model answers from two places: what it learned in training, and what it looks up for this question. With search on, it retrieves pages, picks passages, and writes a reply from them.

WHY IT MATTERS

You can't edit what a model already learned. You can work on what it finds when it searches: whether your page can be retrieved, and whether your passage is the clearest one available.

JOURDAN'S TAKE

A key claim can lose its meaning once it leaves the paragraph around it. I write the subject, the conditions, and the source into the claim itself.



THE BRAIN CAN KNOW THINGS AND STILL NEED TO LOOK THINGS UP.

IN THIS ROOM

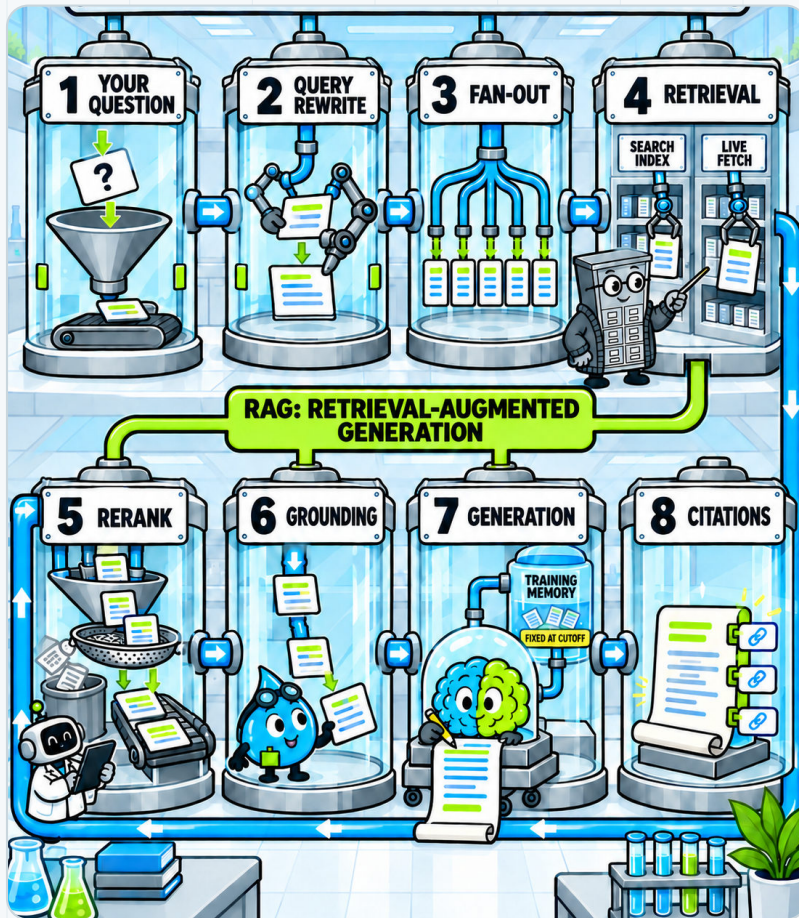
16 Anatomy of an AI Engine
17 Answer Pipeline: Part 1
18 Answer Pipeline: Part 2

19 Query Fan-Out
20 Passages Need Context
21 The Crawl-to-Click Gap

22 Engines Consult Different Shelves

ANATOMY OF AN AI ENGINE

Most AI products that search the web follow a pattern like this one. The numbers match the chambers: what each part does, and where your page can win or lose.



- 1 **YOUR QUESTION.** The prompt, plus context the product adds, such as location or chat history.
Your move: Learn the questions your audience really asks.
- 2 **QUERY REWRITE.** The system turns the prompt into search queries.
Your move: Use the words people search with.
- 3 **FAN-OUT.** One question becomes several searches, one per angle.
Your move: Cover the sub-questions (p. 19).
- 4 **RETRIEVAL.** Candidates come from a search index or a live fetch.
Your move: Be crawlable, indexed and fast (Room 05).
- 5 **RERANK.** Passages are scored against the question; a few survive.
Your move: Write passages that stand alone (p. 20).
- 6 **GROUNDING.** The survivors go into the model's context as evidence.
Your move: Keep the fact, the number and the source together.
- 7 **GENERATION.** The model writes from that evidence and its training memory.
Your move: Say the same thing everywhere (Room 04).
- 8 **CITATIONS.** Some sources get a visible link; most do not.
Your move: Count mentions, citations and visits separately (Room 08).

RAG, IN PLAIN WORDS

Retrieval-augmented generation: the model looks up evidence (stages 4 to 6) and writes from it, instead of answering from memory alone. That lookup is the door your page comes through.

TRAINING MEMORY

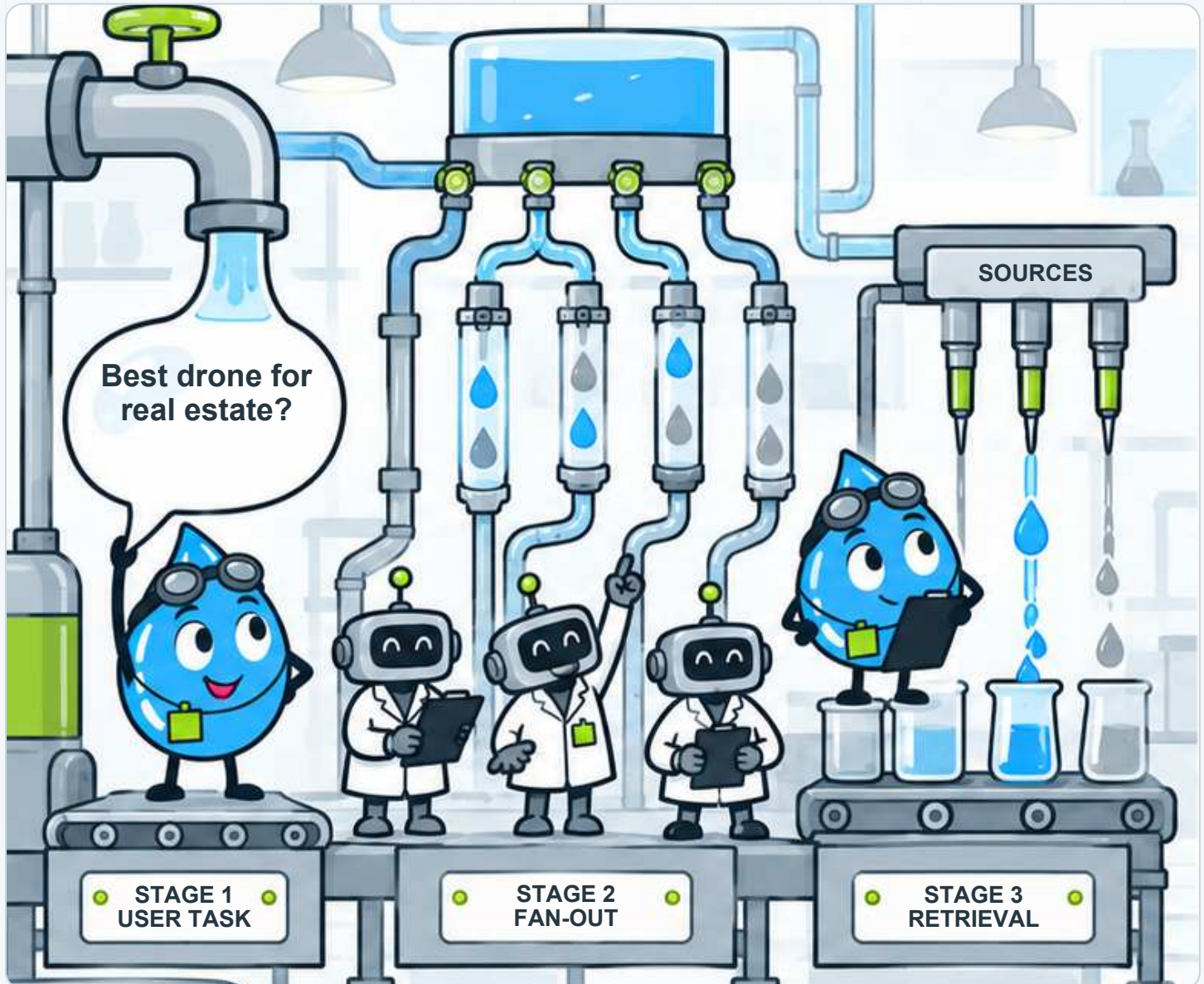
What the model learned in training is fixed at a cutoff, and you cannot edit it. New pages, prices and facts reach answers through retrieval.

A simplified model. Products differ: some answer without searching, merge steps or repeat them. The stage names describe jobs, not any one vendor's system.

THE ANSWER PIPELINE

PART 1 OF 2

Continue: stages 4 to 6 >



1. USER TASK

A user asks a question.

2. FAN-OUT

Possible related searches explore useful sub-questions.

3. RETRIEVAL

Retrieval gathers candidate information.

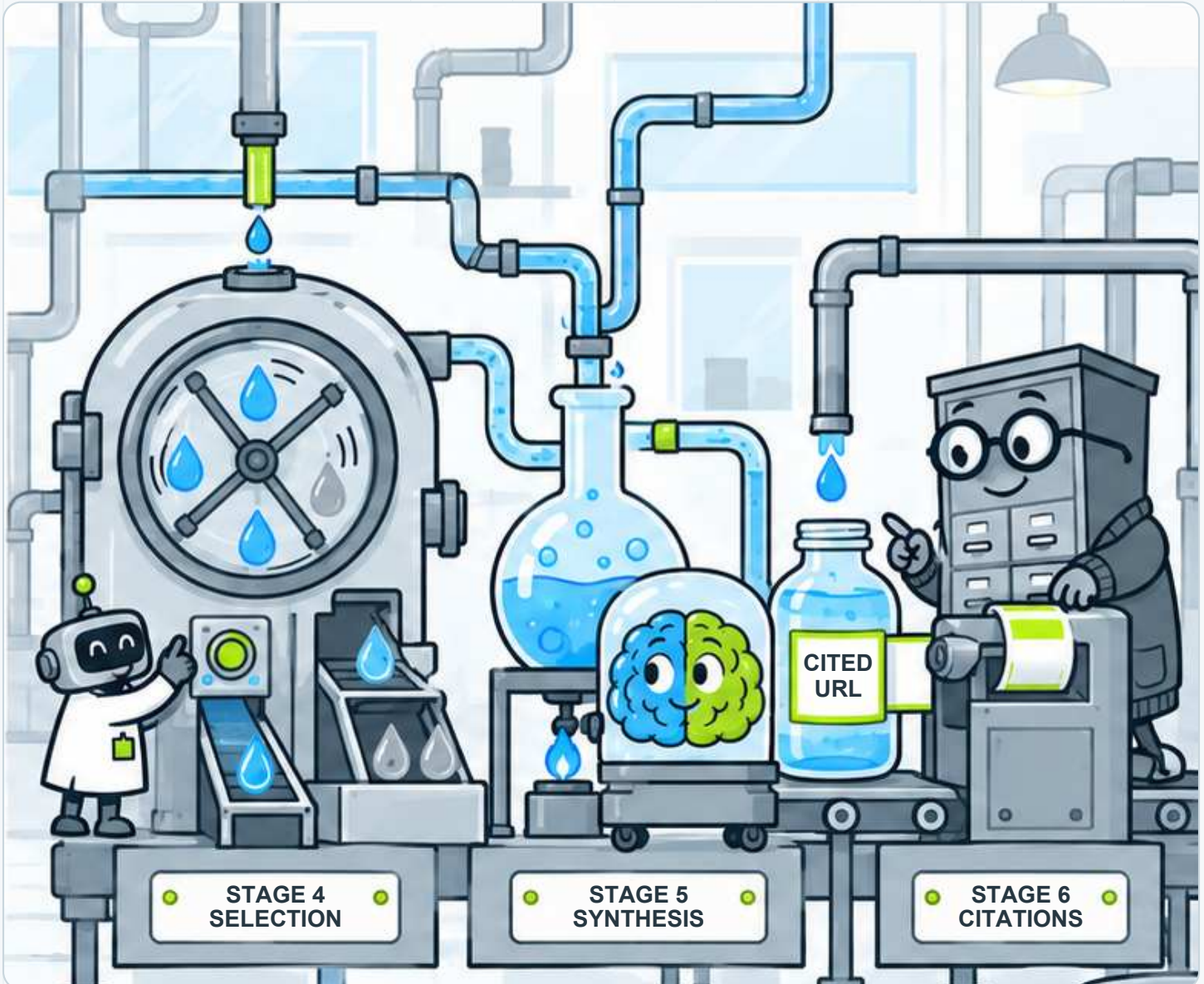


**EVERY ANSWER STARTS AS A QUESTION,
THEN A SEARCH FOR EVIDENCE.**

A simplified workflow: search-enabled products can combine or repeat these stages.

THE ANSWER PIPELINE

PART 2 OF 2

[< Back to stages 1 to 3](#)

4. SELECTION

Candidate selection may reorder the available evidence.

5. SYNTHESIS

The model synthesizes an answer.

6. CITATIONS

Some supporting sources receive visible links.



***I HAVE TO SURVIVE THE JOURNEY
WITH MY MEANING INTACT.***

QUERY FAN-OUT

Illustrative task: best drone for real estate photography

WHERE AIO CITATIONS RANK ORGANICALLY

Ahrefs | Data: collection not stated | Published: March 2026 | 863K keywords, 4M cited URLs

37.1%

26.2%

36.7%

Top 10

11 to 100

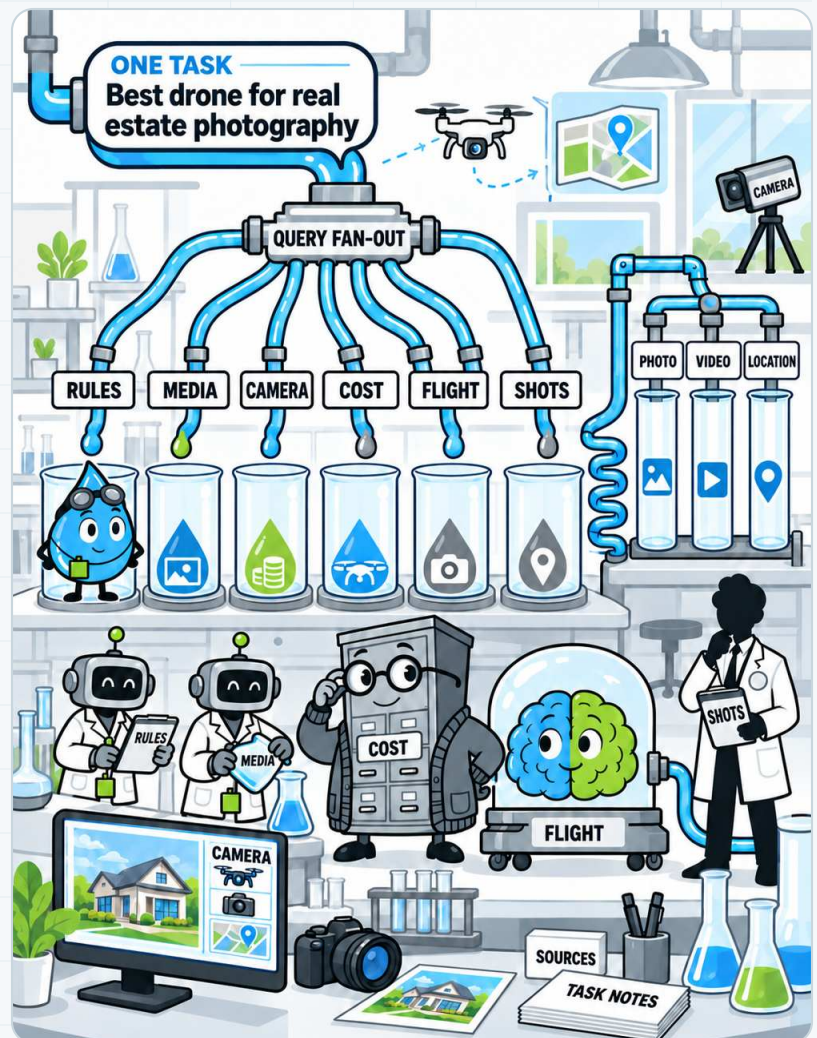
Outside top 100

ONE QUESTION, MANY SEARCHES

An engine can split one task into sub-questions: rules, media, camera, cost, flight limits and shots. Each one is a separate search.

WHAT IT MEANS

A cited page does not have to rank first. Pages that answer one sub-question well can earn a place in the answer.



ONE QUESTION CAN OPEN SEVERAL EVIDENCE TRAILS.

Data: collection dates not stated / Published: March 2, 2026. Organic-only rank buckets for the same query; 863,000 keyword SERPs and 4 million cited URLs.

PASSAGES NEED CONTEXT

WRITE IT TO STAND ALONE

Write self-contained passages inside coherent pages.

STATE THE FOUR PARTS

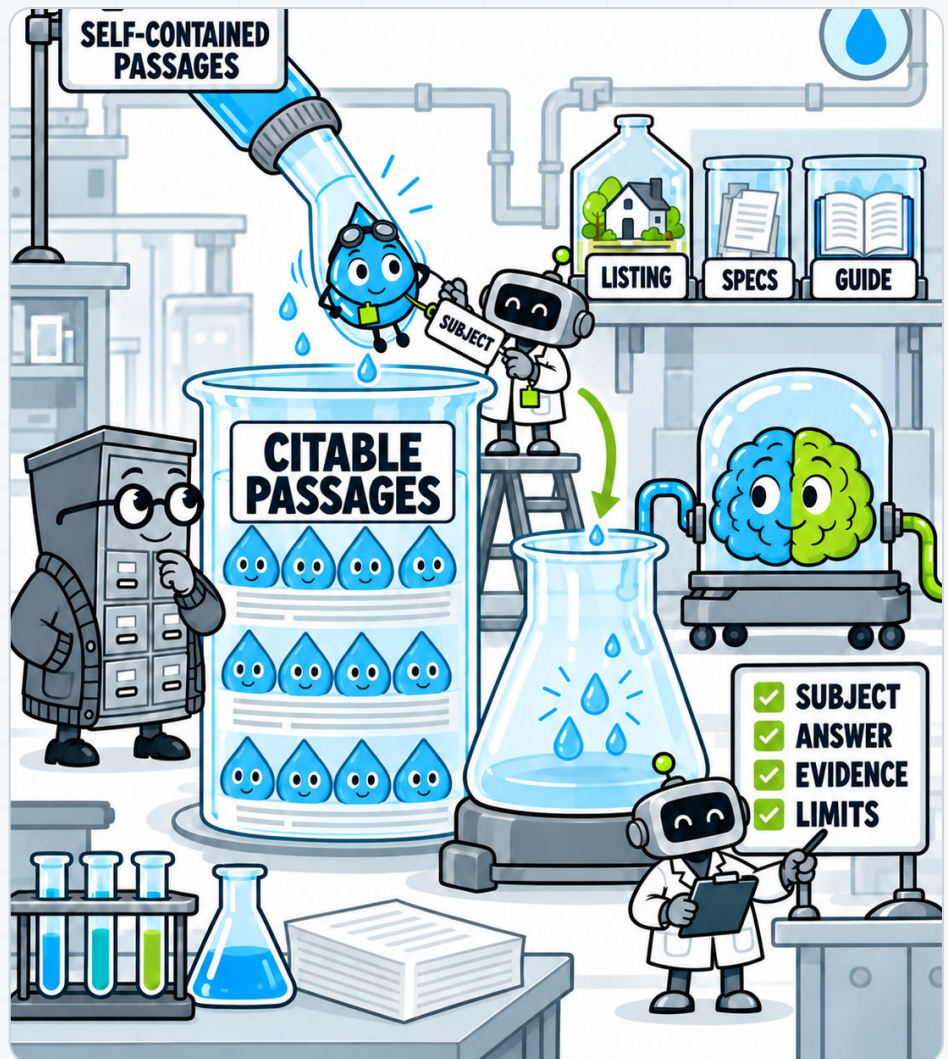
State the subject, answer, evidence, and limits clearly enough to survive being read alone.

KEEP THE PAGE COHERENT

Build deep source pages for specific tasks; keep the whole page coherent.

THE CHECKLIST ON THE PLATE

Subject, answer, evidence, limits. If a passage lifted out of the page still has all four, it can be quoted accurately.



READ ME ALONE. DO I STILL MAKE SENSE?

THE CRAWL-TO-CLICK GAP

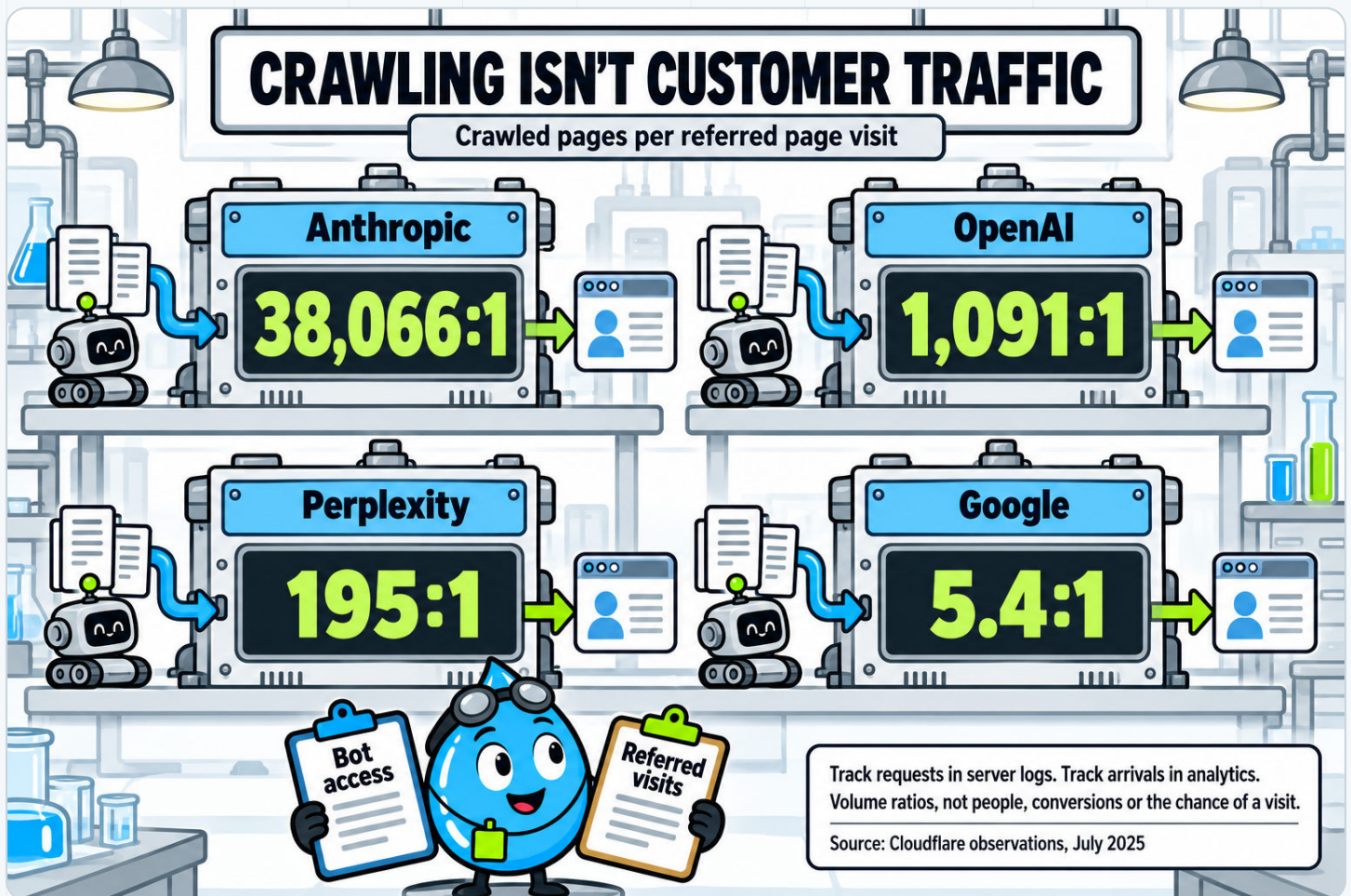
CLASSIFIED CRAWL PURPOSE

Cloudflare | Data: July 2025 | Published: Aug. 29, 2025 | observed classified traffic



■ Training: 79% ■ Search: 17% ■ User actions: 3.2%

Rounded shares total 99.2%; remaining 0.8 percentage points are not assigned here.



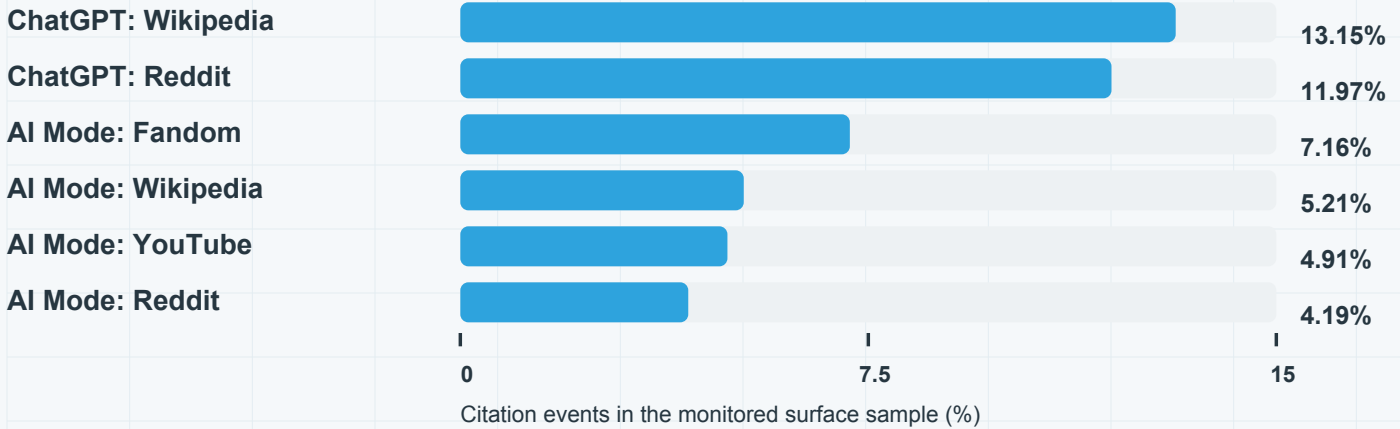
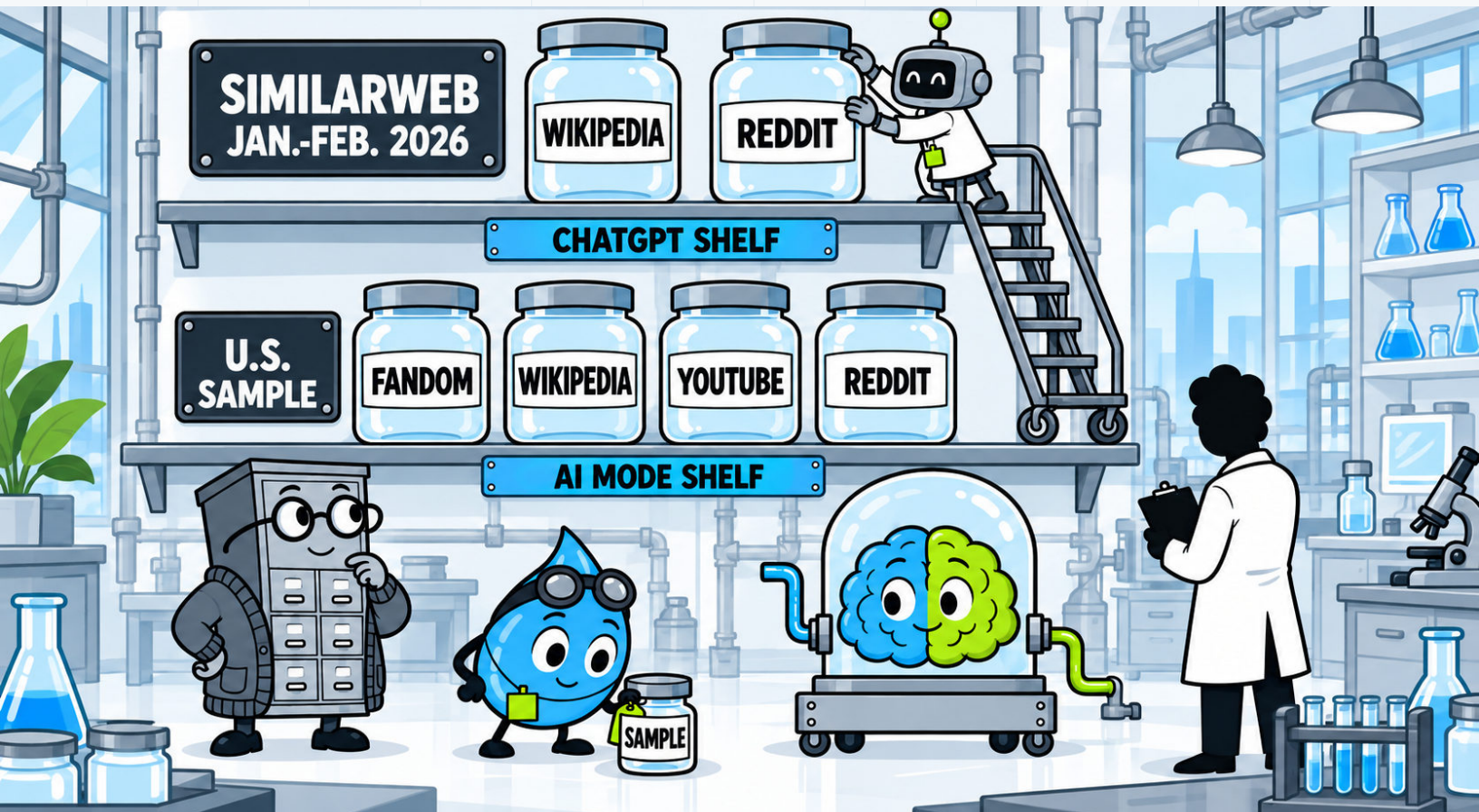
BEING FETCHED IS ONLY THE START OF MY JOURNEY.

Data: July 2025 / Published: Aug. 29, 2025. Rounded crawl shares and crawled-pages / referred-page-visit ratios. Denominators differ from users and conversions.

ENGINES CONSULT DIFFERENT SHELVES

SHARE OF TRACKED CITATION EVENTS

Similarweb | Data: Jan.-Feb. 2026 | U.S. monitored prompts, sample count not disclosed

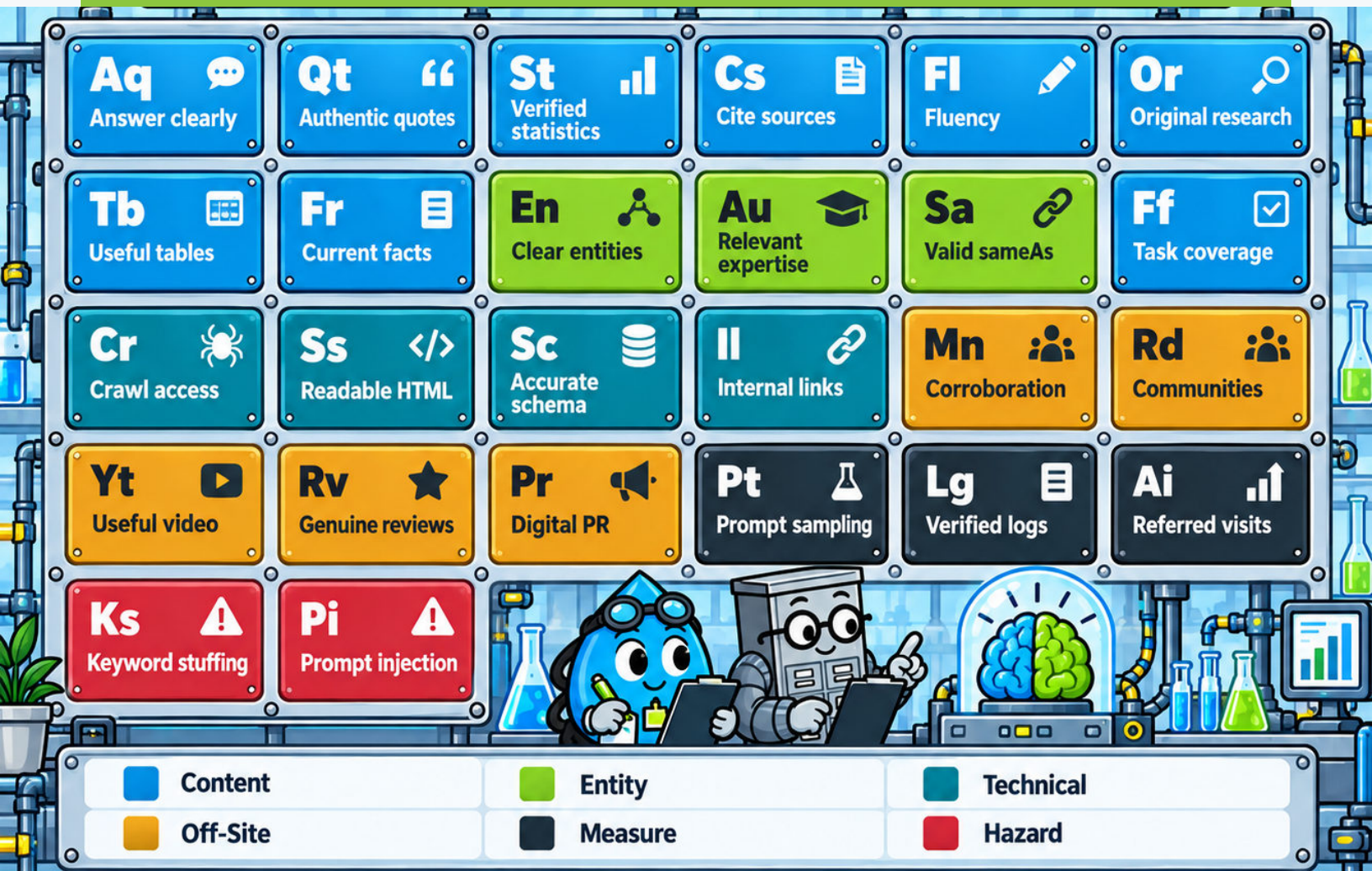


**EACH ENGINE LEANS ON DIFFERENT SITES.
LEARN WHICH ONES YOURS USES.**

Similarweb U.S. monitored prompts, Jan. to Feb. 2026. Percentages count tracked citation events by surface.

THE PERIODIC TABLE OF GEO

03



HOW TO READ IT

Each tile is a practice you can work on, colored by type: content, entity, technical, off-site, measurement, and two hazards. Click a tile to jump to the room that explains it.

JOURDAN'S TAKE

Most GEO hacks are good writing with a new label. Clear answers, real numbers and named sources are where I'd spend the first hour, not on another keyword.

IN THIS ROOM

24 The Experiment That Started It All
25 From Vague to Specific

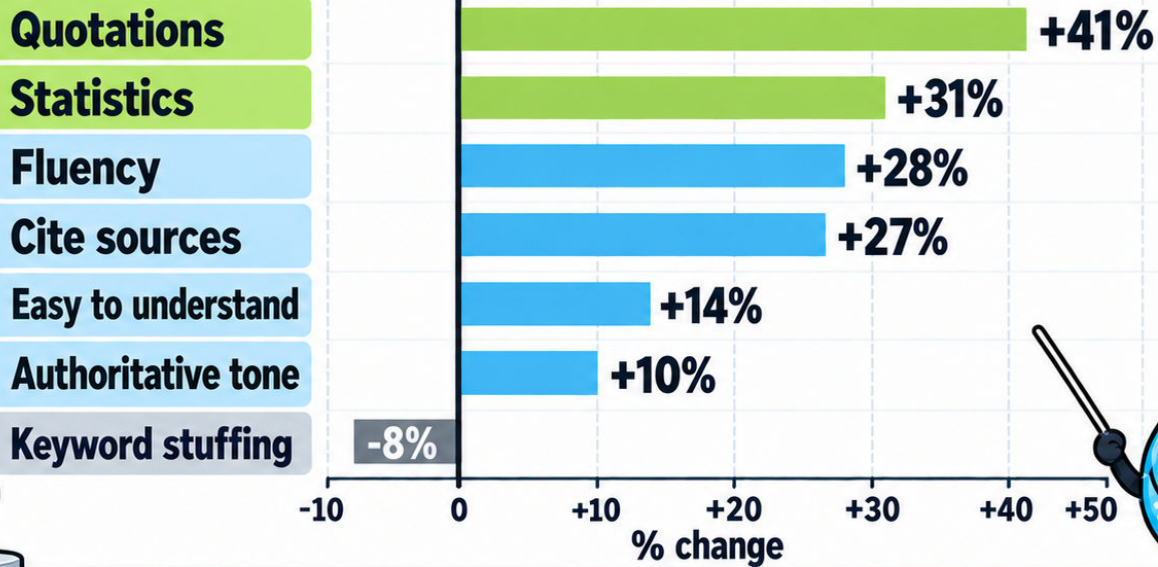
26 Do AI Answers Prefer Fresh Pages?
27 Busted Experiments

No tile is a known ranking weight or a guaranteed citation: the table is a menu for judgment. The next pages show what research measured.

THE EXPERIMENT THAT STARTED IT ALL

WHAT HELPED IN GEO-BENCH?

Relative change in answer visibility



Simulated engine. Fixed-source visibility, not live traffic.
Overall Position-Adjusted Word Count • Baseline: 19.3

Source: GEO research, Table 1 • Rounded derived changes



Use authentic quotes and statistics

WHAT IT FOUND

The 2024 paper that named GEO rewrote source texts nine ways. Adding quotations, statistics and citations raised a simulated engine's visibility metric; keyword stuffing lowered it.

WHAT IT DOES NOT SHOW

Its published code allowed synthetic quotations, invented sources and hypothetical statistics, so it does not show an effect from verified evidence. Nor does it measure live citations or traffic.

WHAT TO DO

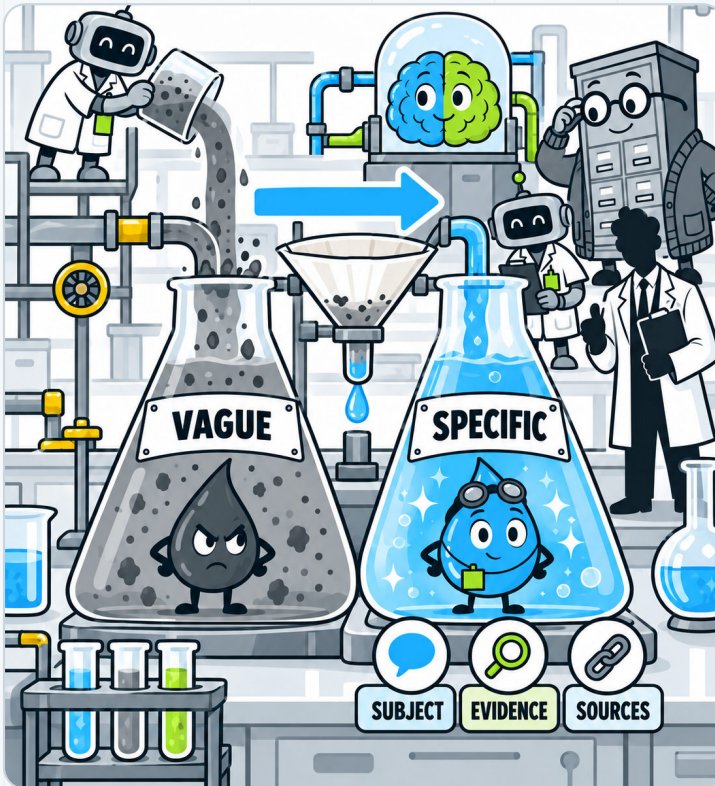
Publish authentic, traceable evidence: real numbers with dates, named people you quoted, links to the original source. It is honest and checkable, whatever a benchmark scores.



REAL EVIDENCE BEATS A LOUDER ADJECTIVE.

FROM VAGUE TO SPECIFIC

Most passages fail because they are vague, not because they are short. Here is the same advice before and after a rewrite.



VAGUE

Drones are popular, there are lots of options, and it depends on what you need.

SPECIFIC

For real estate listings, compare camera quality, obstacle sensing, flight time and the flight rules that apply where you shoot. Show sample images, the test conditions and the date of the specs.

SUBJECT

Names who it is for and what it covers: drones for real estate listings.

EVIDENCE

Says what to compare and shows sample images and test conditions.

SOURCES

Dates the specs, so a reader or an engine can check them.

CHUNK SAYS



NAME THE SUBJECT. SHOW THE PROOF. SAY WHERE IT COMES FROM.

Illustrative rewrite, not a product recommendation or legal advice. The specific version is easier to check; neither version is guaranteed a citation.

DO AI ANSWERS PREFER FRESH PAGES?

MEAN AGE OF URLS

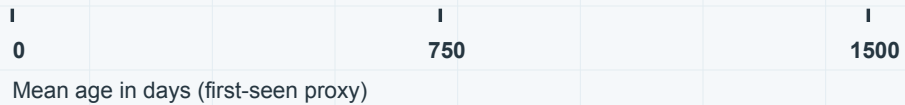
Ahrefs | Published: July 2025 | Collection date / sample count not stated; first-seen age proxy

ChatGPT citations

958

Organic results

1,416



WHAT THE DATA SHOWS

In Ahrefs' study, pages cited by ChatGPT were younger on average than pages in organic results: about 958 days old versus 1,416.

WHAT IT DOES NOT SHOW

Age was estimated from when a crawler first saw each URL, and the collection dates are not stated. Older pages still get cited when they stay accurate.

UPDATE FACTS, NOT TIMESTAMPS

Revise the substance and keep its evidence date. A new timestamp alone is not stronger proof (page 58).



NEW PAINT IS NOT THE SAME AS NEW EVIDENCE.

Data: citation snapshot, collection dates not stated / Published: July 28, 2025. Mean age uses crawler first-seen as a publication proxy.

BUSTED EXPERIMENTS



KEYWORD STUFFING

Tested and failed: about 8% lower on the GEO study's selected visibility metric. Bing's guidelines now name it.

llms.txt

10.13% of nearly 300,000 domains have one, with no positive link to citation frequency (live source updated Oct. 15, 2026). Useful for agents, not a ranking lever.

SPECIAL AI SCHEMA

Google requires no special GEO schema. Keep accurate markup for rich results and identity.

SCALED MANIPULATION

Low-value mass production is not a useful evidence strategy, and Google treats it as scaled content abuse.



**LET'S KEEP THE USEFUL WORK
AND DISCARD THE SHORTCUTS.**

The llms.txt result is observational. The GEO figure uses its study's selected visibility metric. Every popular trick, rated: page 55.

WHAT'S HAPPENING

AI answers describe a business with facts gathered from many places: its website, LinkedIn, Google Business Profile, directories and press. When those facts disagree, the answer has to guess.

WHY IT MATTERS

A wrong founding year, an old phone number or a discontinued product can be repeated as fact. One approved set of facts, published everywhere and connected with structured data, is harder to confuse.

JOURDAN'S TAKE

Brand facts are data, so treat them like data. One owner, one approved version, one review date. A slogan isn't a fact.



A FAMILIAR NAME NEEDS A CONSISTENT IDENTITY.

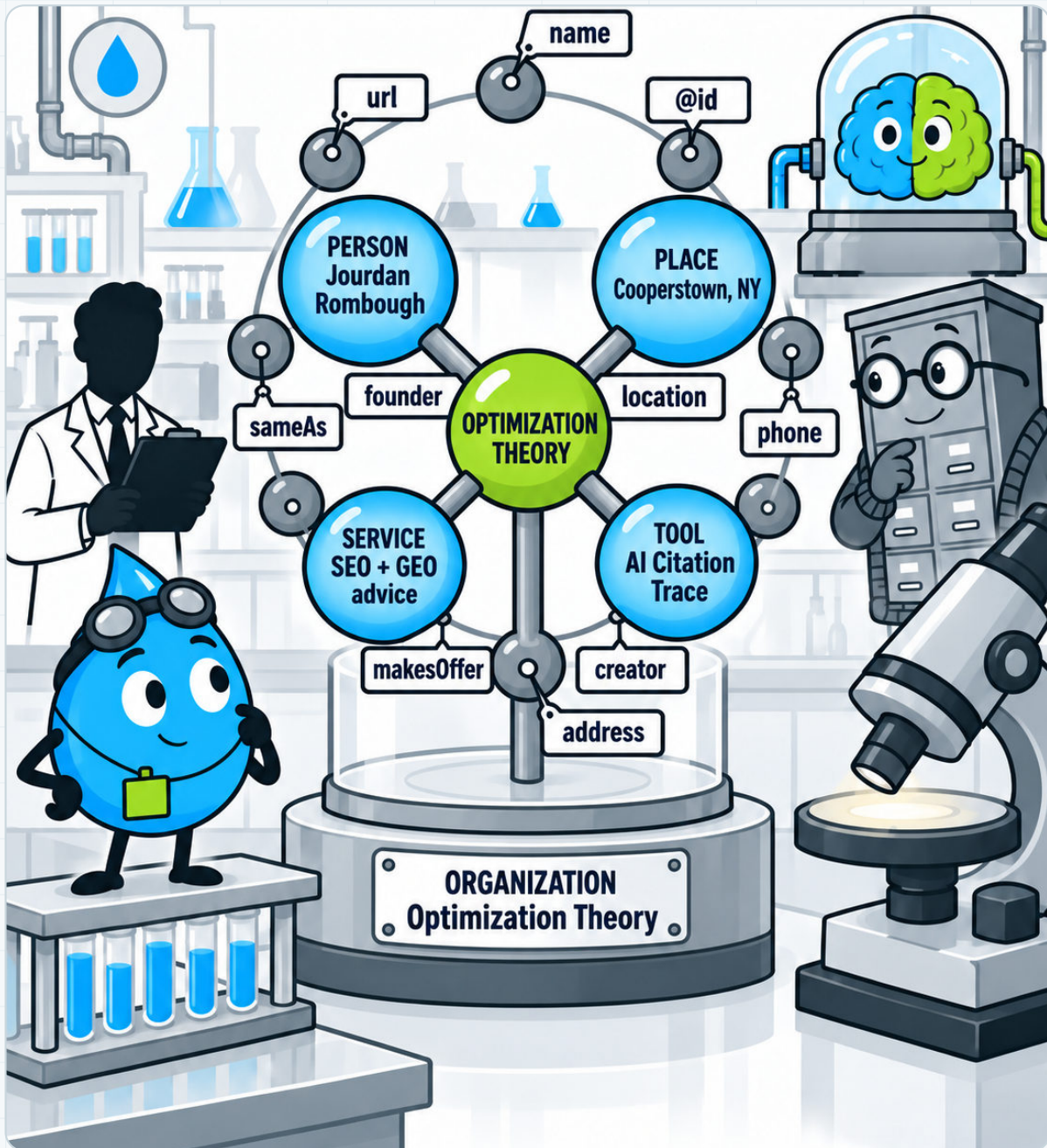
IN THIS ROOM

- 29 The Entity Molecule
- 30 The DNA Match Test

- 31 Repair Wrong Answers
- 32 Fill Real Decision Gaps

THE ENTITY MOLECULE

Structured data describes your organization as connected things: a person, a place, a service, a tool, each joined by a named relationship.



RELATIONSHIPS CONNECT US. THEY DON'T MAKE US THE SAME ENTITY.

Bond labels are schema.org properties read from the organization outward (creator reads from the tool). sameAs joins only equivalent identities, here the person's LinkedIn profile. name, url, @id, phone, and address are properties of a node, not separate entities.

THE DNA MATCH TEST

Illustrative mismatch check: current phone number

COLUMNS ON THE BOARD

WEB website, LI LinkedIn, GBP Google Business Profile, DIR directories, PR press. The warning sits on DIR: a directory still lists an old phone number.

REPAIR THE RECORD

Keep the approved value, its owner, where it is published and its review date in one record. Correct material errors, then retest the task.

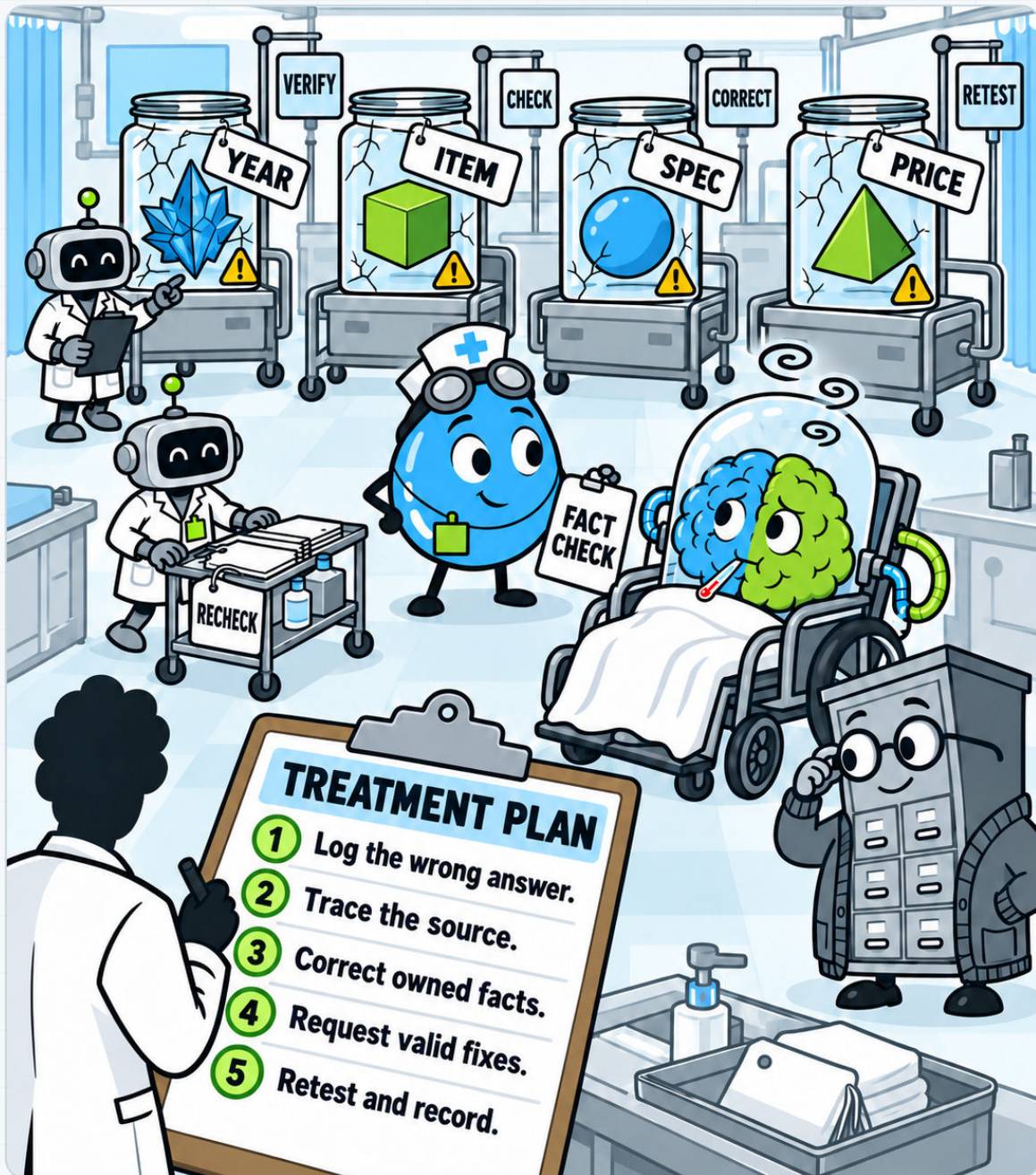
EXAMPLE STATUS

The board illustrates a discrepancy. It does not claim these live profiles currently publish conflicting numbers.



REPAIR WRONG ANSWERS

When an AI answer gets a fact about you wrong, treat it like a case: log it, trace where the claim came from, and fix the source.



IF AN ANSWER GETS US WRONG, TRACE THE CLAIM AND REPAIR THE SOURCE.

Jar tags: YEAR = wrong founding year, ITEM = discontinued product, SPEC = a competitor's feature, PRICE = old pricing. Visible citations do not reveal every causal input; a corrected source can improve the evidence without guaranteeing that every answer updates.

FILL REAL DECISION GAPS

Map the questions that affect a purchase or action. Filled tubes are questions your content already answers; the empty Gap tubes are questions people ask that it does not answer yet. Fill gaps with sourced explanations and merge pages that overlap.



THE EMPTY SLOTS ARE QUESTIONS WE HAVEN'T ANSWERED YET.

Tube tags: Define, How, Cost, Compare, Mistakes, Tools, Examples and FAQ are questions the content already answers; the two Gap tubes (Limits, Evidence) are questions it has not answered yet.

THE TECHNICAL AIRLOCK

05

WHAT'S HAPPENING

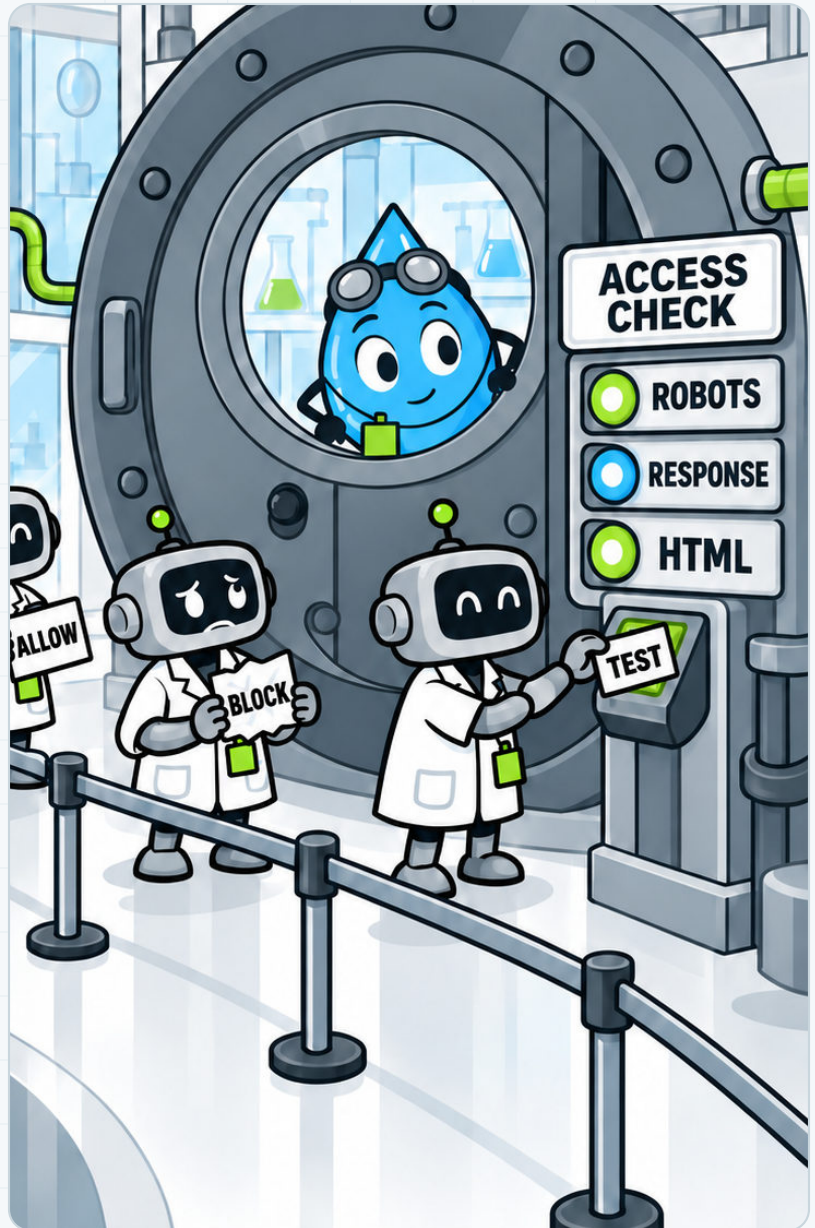
Different bots visit for different jobs: some collect training data, some build search indexes, some fetch a page because a user asked. Each can be allowed or blocked separately.

WHY IT MATTERS

A fetcher that is blocked, gets an error, or receives HTML where the key facts only appear after JavaScript runs may never see your best content, however good it is.

JOURDAN'S TAKE

For a direct citation to this page, check the response and content the target fetcher can access before testing content changes. Other routes exist (partner indexes, cached copies, learned knowledge); test the one you are measuring.



CAN THE RIGHT FETCHER GET THROUGH THIS AIRLOCK?

IN THIS ROOM

34 The Badge Wall

35 Clear the JavaScript Fog

36 Specimen Labels

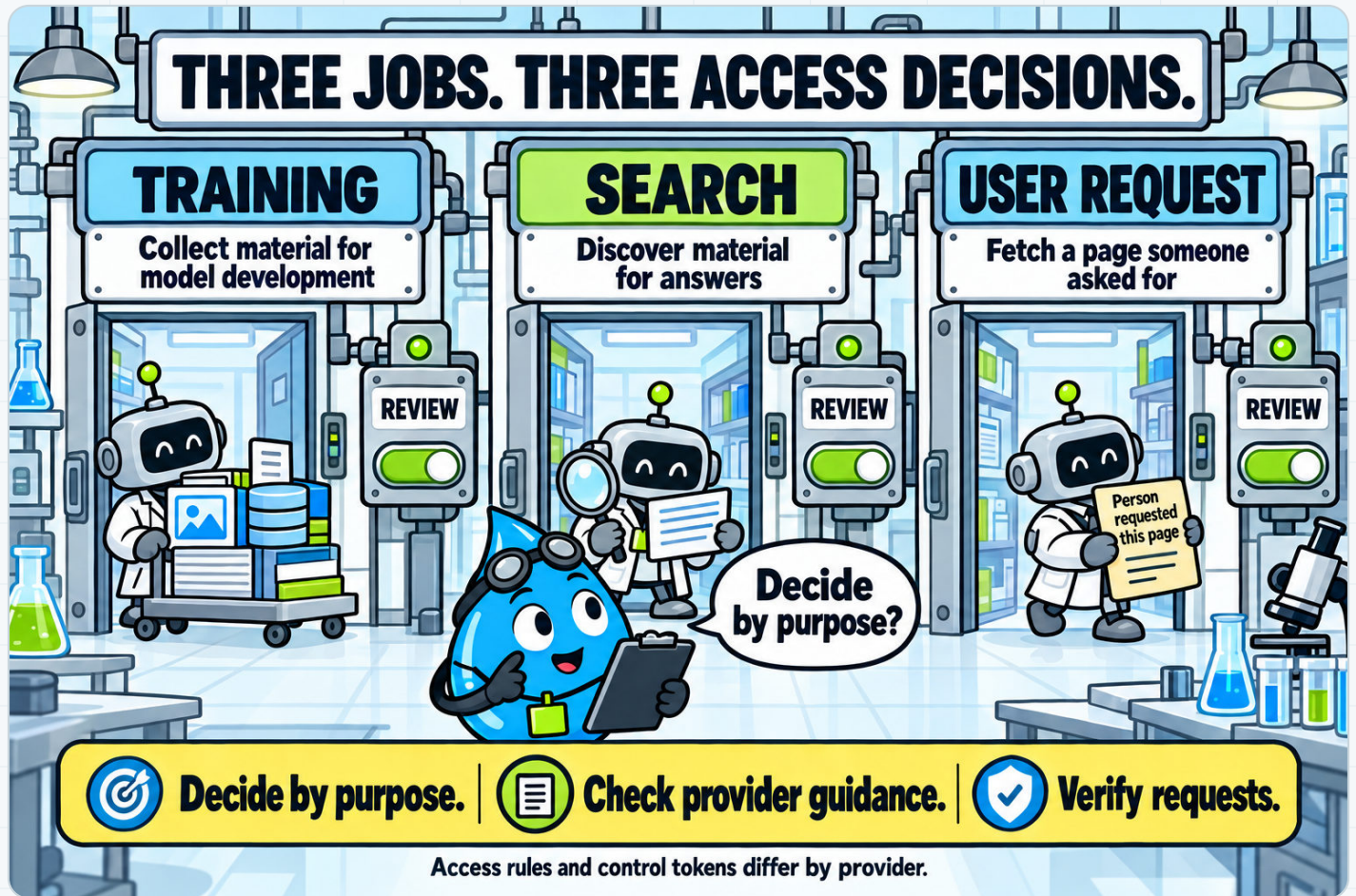
37 Different Files, Jobs

38 Markdown for Agents

39 Markdown: The Evidence

Access, indexation, and answer selection are separate checks.

THE BADGE WALL



TRAINING / COLLECTION

GPTBot

ClaudeBot

Google-Extended*

Applebot-Extended*

CCBot

SEARCH DISCOVERY

OAI-SearchBot

Claude-SearchBot

PerplexityBot

Googlebot / Bingbot

USER ACTIONS

ChatGPT-User

Claude-User

Perplexity-User

Provider-specific rules

* Google-Extended and Applebot-Extended are control tokens, not separate fetching identities. Merge policy choices with your existing robots rules.



GIVE EACH CRAWLER THE POLICY ITS JOB REQUIRES.

Extended names are control tokens, not separate fetchers. CCBot collects Common Crawl's public web corpus. Merge policies deliberately.

CLEAR THE JAVASCRIPT FOG

INSPECT THE DELIVERY

Keep essential facts in delivered HTML. Inspect redirects and error responses. Verify today's target fetcher.

THE SEVEN TAGS

Each tag is one check on the delivered page: 1 the URL, 2 its HTTP status, 3 the delivered HTML, 4 the key facts, 5 the named source, 6 the date, 7 the index status. The fogged jar is what appears only after scripts run.

HISTORICAL 404 SHARES

ChatGPT 34.82% / Claude 34.16% / Googlebot 8.22%. Vercel/MERJ, Dec. 2024; Applebot was the rendering exception.



MY ESSENTIAL FACTS SHOULD SURVIVE THE FOG.

Vercel/MERJ, Dec. 2024: historical measured fetchers; Applebot was a rendering exception. Verify the current target.

SPECIMEN LABELS

SCHEMA TYPE	WHAT THE PAGE MUST SHOW	IN GOOGLE SEARCH, 2026
Organization	Your name, logo, contact details and sameAs profiles, on the home or about page.	● Supported: logo, legal name and contact details.
Person	A real person's name, role and profiles, on a page about them.	● No feature of its own; used for authors and Profile page markup.
Article	Headline, author, dates and images that match the page.	● Supported: title, image and date display.
Product	Name, price, availability and reviews shown on the page.	● Supported: product snippets and merchant listings.
LocalBusiness	Name, address, hours and phone that match your Business Profile.	● Supported: hours and details in the knowledge panel.
Review	A real review of a specific item, visible on the page.	● Supported, but reviews a business controls about itself get no stars.
HowTo	Steps a reader can follow, visible on the page.	● No rich result: removed from Search in 2023.
FAQPage	Questions and answers visible on the page.	● No rich result since May 7, 2026.

YOUR PAGE

Markup must describe content a person can see on the page. Valid schema.org vocabulary and Google eligibility are separate checks: test both.

THE AI ANGLE

Schema alone does not earn AI citations: Ahrefs found no significant lift (page 57). Keep it accurate for rich results and to say who you are.



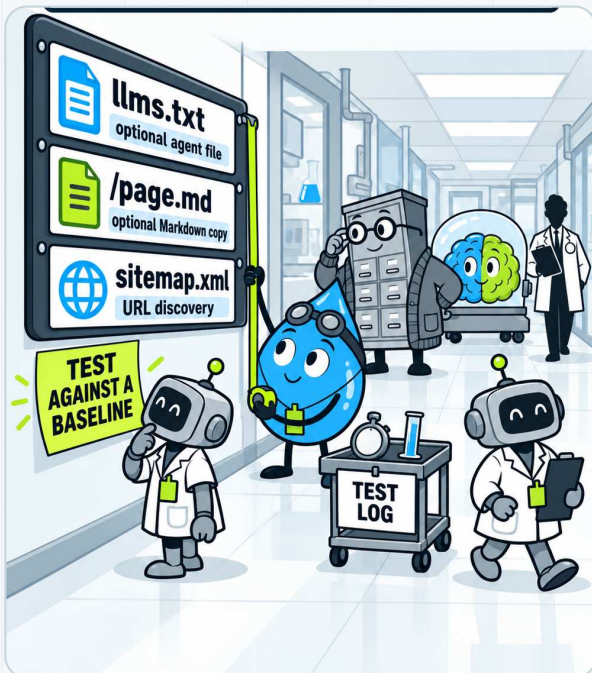
A SPECIMEN LABEL SHOULD DESCRIBE WHAT'S ACTUALLY IN THE JAR.

Google feature status checked Oct. 29, 2026 against its structured data search gallery, the FAQ deprecation notice and the review snippet guidelines. All three are in the Source Drawer.

DIFFERENT FILES, DIFFERENT JOBS

Several small files talk to machines about your site. Each has one job, and none of them is a shortcut into AI answers. What each one does, who reads it, and what the evidence says:

FILE	WHAT IT DOES	WHO READS IT	EFFECT ON AI ANSWERS
robots.txt	Tells crawlers which paths they may fetch, bot by bot.	Search and AI crawlers that honour it.	Access only: a blocked page cannot be fetched for a fresh answer (Room 05).
sitemap.xml	Lists your URLs and when each one changed.	Search engines, for discovery.	Indirect: helps pages get found and refreshed.
llms.txt	A curated list of key pages for agents.	Some AI tools; Google Search does not use it.	None shown: no citation link across about 300,000 domains.
page.md	A clean Markdown copy, linked with rel="alternate".	Agents and tools that ask for Markdown.	Helps agents read you; no citation lift shown (p. 38).
JSON-LD schema	Describes the page's entities and facts.	Search engines, for rich results and identity.	No significant AI citation lift in Ahrefs' test (p. 57).
IndexNow	Pings engines when a URL is added, changed or deleted.	Bing and other participating engines.	Keeps the index behind Copilot fresh (p. 57).



THE RULE OF THUMB

Use robots.txt and sitemaps for access and discovery. Treat llms.txt and Markdown copies as optional help for agents, never as ranking levers.

PROPOSED LAB TEST

Compare a maintained Markdown endpoint against a documented baseline. Preserve the same facts and watch for duplicate-version drift.



A DIRECTORY SIGN HELPS ONLY IF IT POINTS SOMEWHERE USEFUL.

Google says llms.txt does not affect Search ranking. The Markdown strategy, step by step: pages 38-39. The comparison on this page is proposed, not run.

MARKDOWN FOR AGENTS

Many AI agents and retrieval tools would rather read Markdown than a full web page: no menus, scripts or cookie banners, and far fewer tokens. You can offer a clean Markdown copy of an important page and point to it openly. There are three ways.

1. LINK IT IN THE PAGE HEAD

```
<link rel="alternate" type="text/markdown"
href="https://example.com/pricing.md">
```

The HTML page announces its Markdown twin. llms.txt v2 (Aug. 10, 2026) names this exact link.

2. OR SEND IT AS A HEADER

```
Link: <https://example.com/pricing.md>;
rel="alternate"; type="text/markdown"
```

Same signal as an HTTP header, set at the server or CDN. It works for any file type.

3. OR ANSWER WHEN ASKED

```
GET /pricing Accept: text/markdown
-> 200 Content-Type: text/markdown Vary: Accept
```

Same URL, Markdown on request. Cloudflare's Markdown for Agents (Feb. 12, 2026) converts pages this way for sites that turn it on.



THE STANDARD

llms.txt v2 puts Markdown copies at page.md, links them with `rel="alternate" type="text/markdown"`, and uses `rel="describedby"` to point a page at its llms.txt. Chrome's Lighthouse checks llms.txt in an experimental Agentic Browsing audit (May 2026).

DO IT RIGHT

Same facts and dates as the HTML. On the .md send Link: `<page URL>; rel="canonical"` (a strong signal, not a guarantee). Rebuild it when the page changes. Link it openly; never pick the version by user agent.



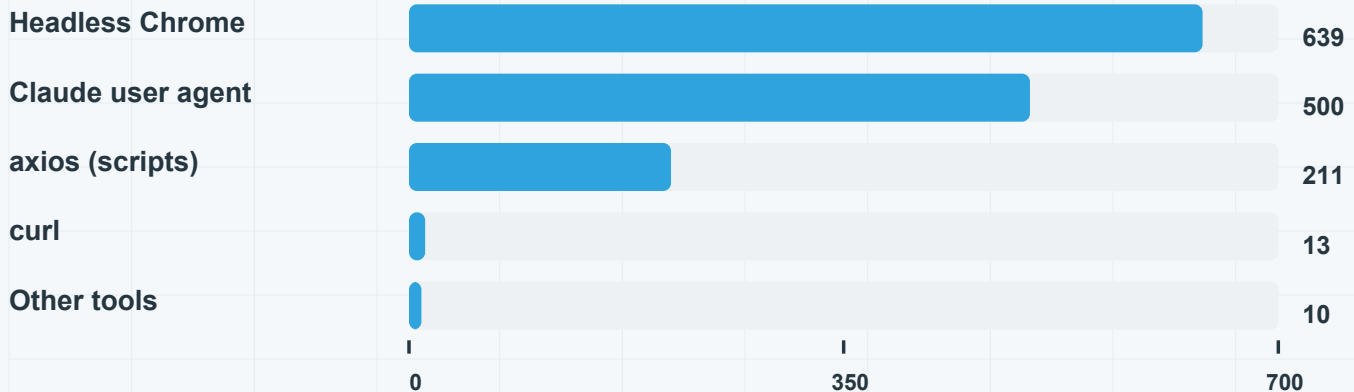
**GIVE AGENTS THE CLEAN COPY.
KEEP EVERY FACT IDENTICAL.**

llms.txt v2 is a proposal still open for feedback; Cloudflare converts pages only for sites that enable the feature. Copy-ready snippets: page 69.

MARKDOWN: WHAT THE EVIDENCE SAYS

WHO ASKED ONE SITE FOR MARKDOWN

Suganthan | Accept: text/markdown requests by user agent, March 7 to April 19, 2026 | one SEO blog on Cloudflare



Requests in 44 days (about 31 a day). GPTBot, PerplexityBot and ClaudeBot never appeared.

WHAT GOOGLE SAYS

Search does not use it: "You don't need to create new machine readable files, AI text files, markup, or Markdown to appear in Google Search." On serving Markdown to bots, John Mueller (Feb. 2026): "Why would they want to see a page that no user sees?"

WHAT BING SAYS

Fabrice Canel (Feb. 2026): "We'll crawl anyway to check similarity," and "Non-user versions (crawlable AJAX and like) are often neglected, broken." A twin you don't maintain becomes a liability.

THE VERDICT: HELPS AGENTS, NOT RANKINGS

Worth doing for pages agents read to act: docs, pricing, specs, policies, product data. Cheap if your CDN converts pages. No study shows a citation lift, and it is not a Google ranking signal.

WHY WE STILL EMPHASIZE IT

Agent traffic is real and growing on that site (about 15 a day in early March, about 30 in April), and agents are starting to browse and buy for people. A clean, accurate copy is a low-risk bet. Log .md and Accept: text/markdown requests to see who uses yours.



**REAL REQUESTS, UNPROVEN PAYOFF.
MEASURE IT ON YOUR OWN SITE.**

One site over 44 days; the post's own totals vary from 1,373 to 1,421 requests. Requests without the header are not counted, and no citation or traffic outcome was measured.

WHAT'S HAPPENING

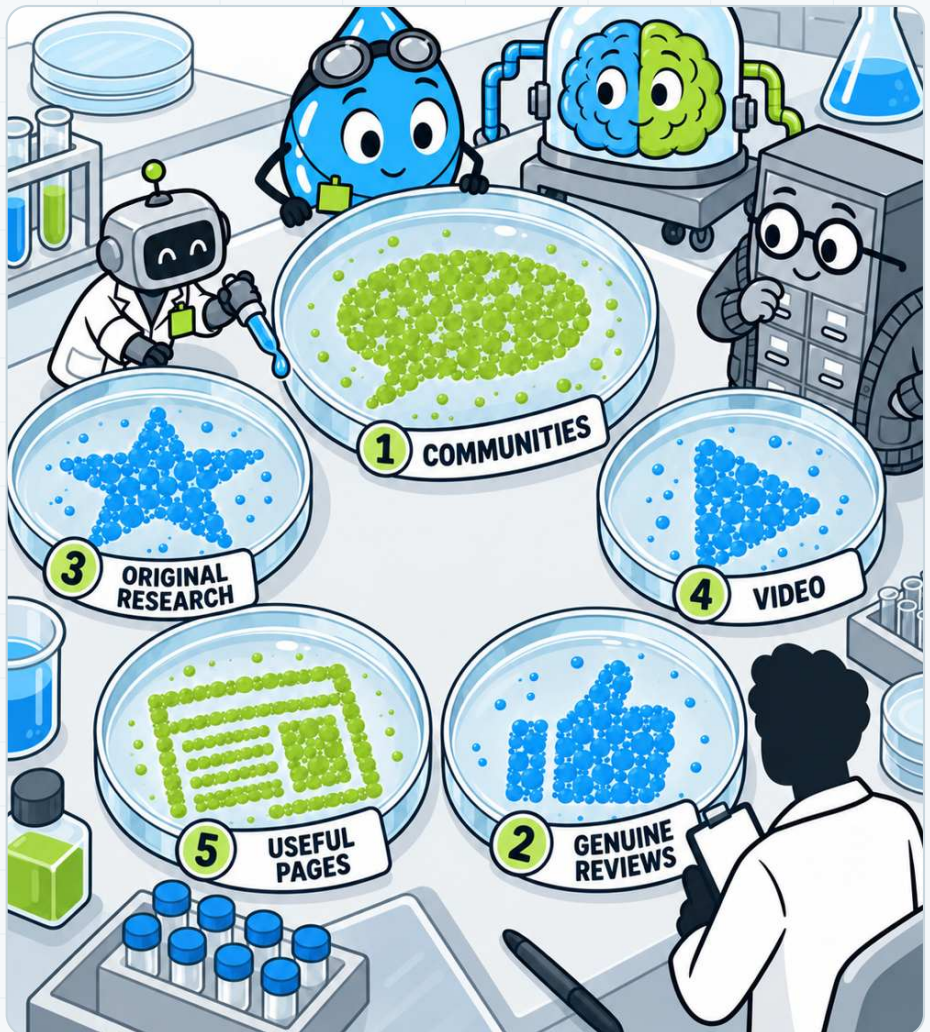
AI answers often cite sources you don't control: Wikipedia, Reddit, YouTube, reviews, news and research. What others say about you is part of the evidence an answer can draw on.

WHY IT MATTERS

Brands mentioned more across the web tend to appear more in AI answers. That is a correlation, not a lever to buy, but it shows why genuine work that others can inspect and discuss is worth doing.

JOURDAN'S TAKE

Paid placement and independent corroboration are different things. Disclose sponsorship, publish a method people can inspect, and let the inspection be the reason they talk about it.



OTHER PEOPLE CAN INSPECT AND CORROBORATE MY EVIDENCE.

IN THIS ROOM

- 41 Web Mentions and AI Visibility
- 42 Help the Community

- 43 Useful Video
- 44 Evidence to Coverage

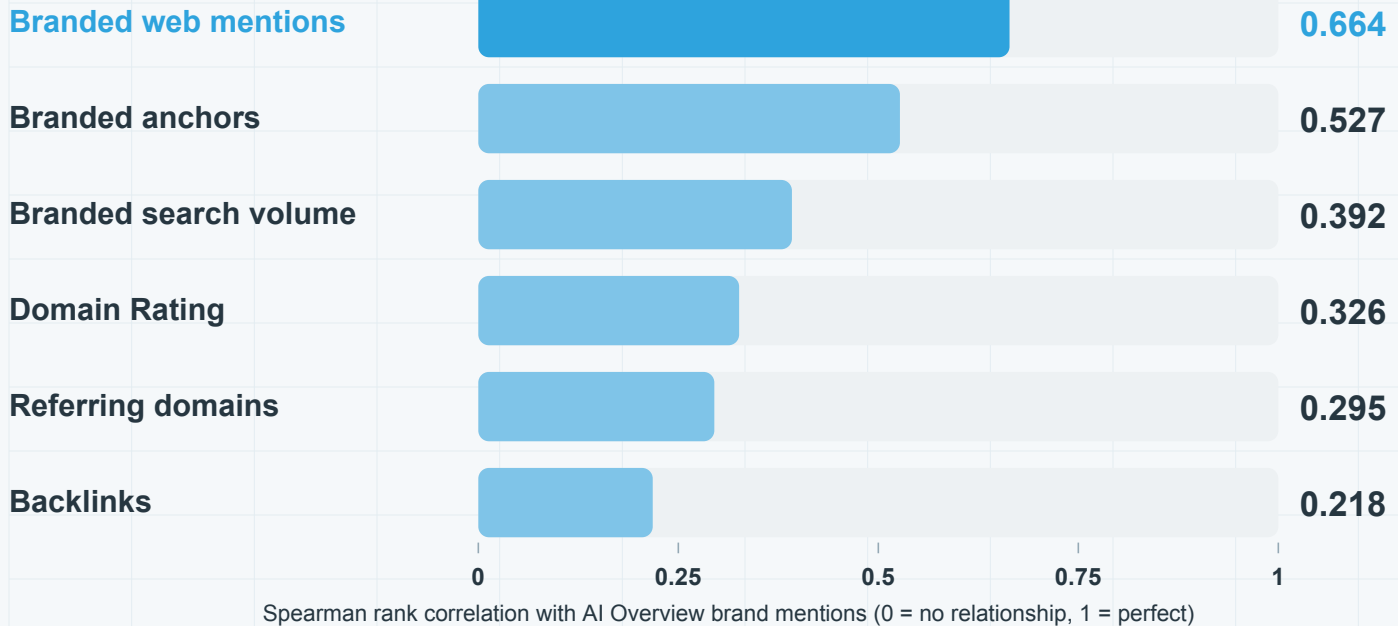
Inspect the sources actually cited for your tasks; do not read a universal trust weighting into a correlation.

WEB MENTIONS AND AI VISIBILITY

Which signals travel with being named in Google's AI Overviews? Ahrefs ranked six brand signals by how closely they moved with AI Overview brand mentions across a selected sample of brands.

ASSOCIATION WITH AI OVERVIEW BRAND MENTIONS

Ahrefs | Published: May 2025 | Selected sample; collection date / final n not stated | Spearman rank correlation, 0 to 1



MENTIONS LEAD

Brands talked about across the web were named more often in AI Overviews: 0.664 for branded web mentions, against 0.218 for backlinks.

169 VS 14

Brands in the top quarter for branded mentions averaged 169 AI Overview mentions; the bottom quarter averaged 14.

NOT A LEVER TO BUY

A correlation, not a cause. Earn mentions with work people want to cite (page 44); bought mentions are not the same thing.



THAT IS AN ASSOCIATION TO INVESTIGATE, NOT A LEVER TO BUY.

Data: brand snapshot, collection dates not stated / Published: May 26, 2025. Spearman correlations; zero-AIO candidates excluded. Association, not a causal weight.

HELP THE COMMUNITY

11.97%

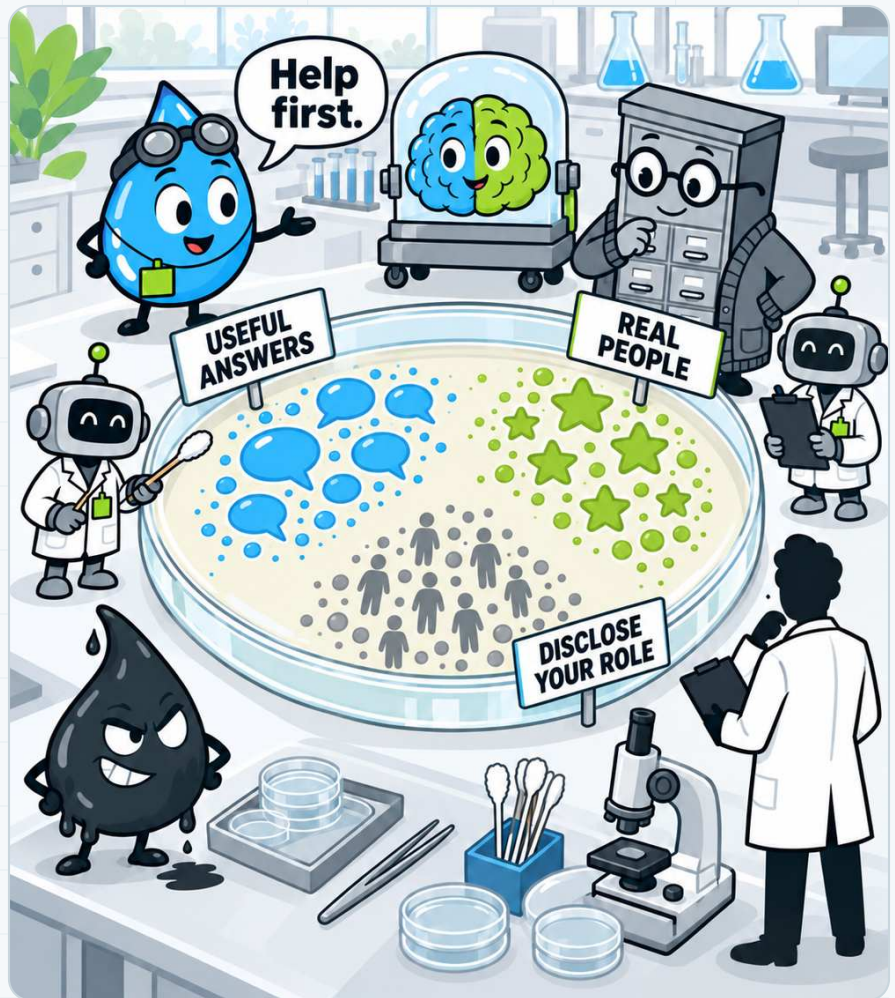
Reddit's share of tracked ChatGPT browsing citation events in Similarweb's U.S. prompt sample, Jan.-Feb. 2026.

PARTICIPATE DIRECTLY

Participate as a person. Answer the actual question. Disclose relevant affiliation. Never fake a thread.

PRACTICAL EXAMPLE

Answer the actual question in the thread where it is asked, say plainly that the page you link is yours, and never manufacture a recommendation thread.



**I'LL HELP THE COMMUNITY
BEFORE ASKING IT TO NOTICE ME.**

Similarweb U.S. monitored ChatGPT browsing citation events, Jan.-Feb. 2026.

USEFUL VIDEO

5.6%

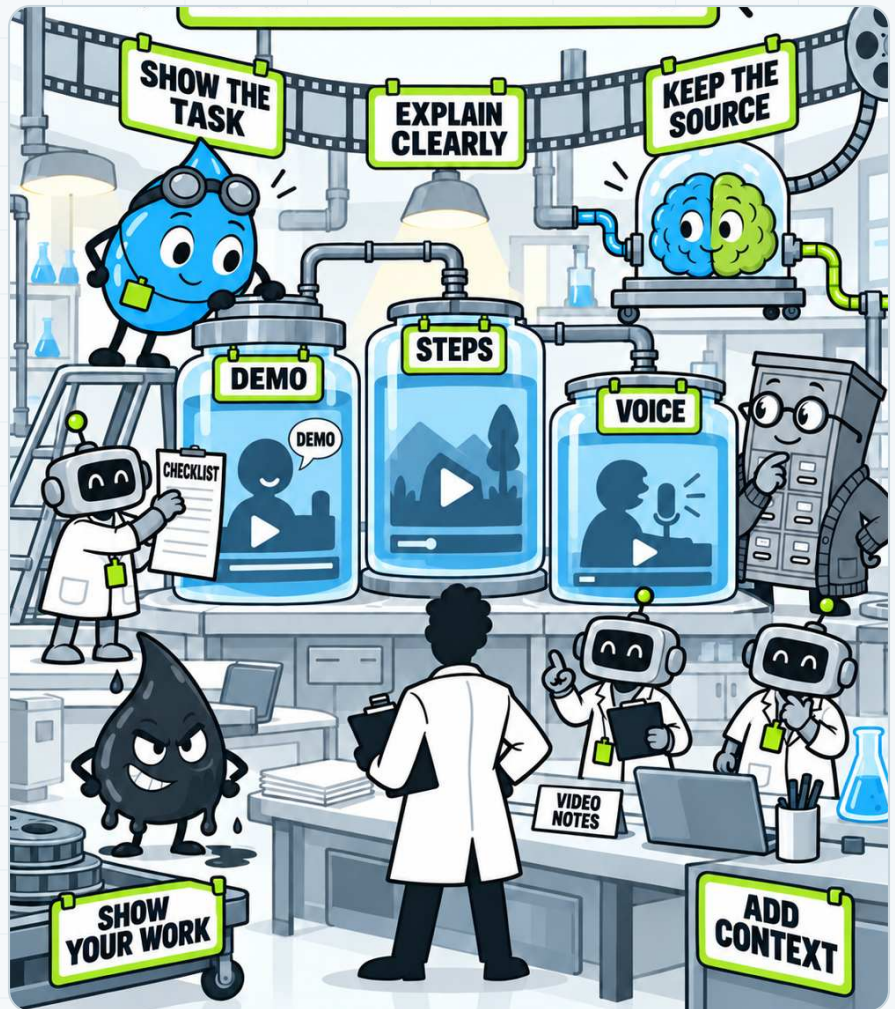
Non-ranking YouTube URLs as a share of all cited AIO URLs in Ahrefs' March 2026 report.

4.91%

YouTube share of tracked AI Mode citation events in Similarweb's U.S. Jan.-Feb. 2026 sample.

PRACTICAL EXAMPLE

A short clip that shows the task step by step, with the method, the date and the source named in the description, so a viewer (or an engine) can check it.



SHOW THE TASK. EXPLAIN THE METHOD. KEEP THE SOURCE NEARBY.

The two studies use different surfaces and denominators: Ahrefs, March 2026; Similarweb, Jan. to Feb. 2026.

EVIDENCE TO COVERAGE



PRACTICAL EXAMPLE

Publish the finding with its raw numbers, sample size and method, so journalists and researchers can check it before they cite it. Then record coverage, citations and visits as separate results.

WHY FINDINGS TRAVEL

Branded web mentions had the strongest association with AI Overview brand visibility (0.664) in Ahrefs' sample (page 41). Bing (Feb. 2026): "Examples, data, and cited sources help build trust when content is reused in AI-generated answers."

THE PATH ON THE PLATE

Raw data, then method, then study, then coverage: publish the numbers, show how you got them, package the finding, and only then pitch it.

WHAT A CITABLE STUDY SHOWS

The question, the sample size and dates, the method, the raw numbers or a downloadable table, the limits, and a named author.



A FINDING TRAVELS FARTHER WHEN THE METHOD TRAVELS WITH IT.

WHAT'S HAPPENING

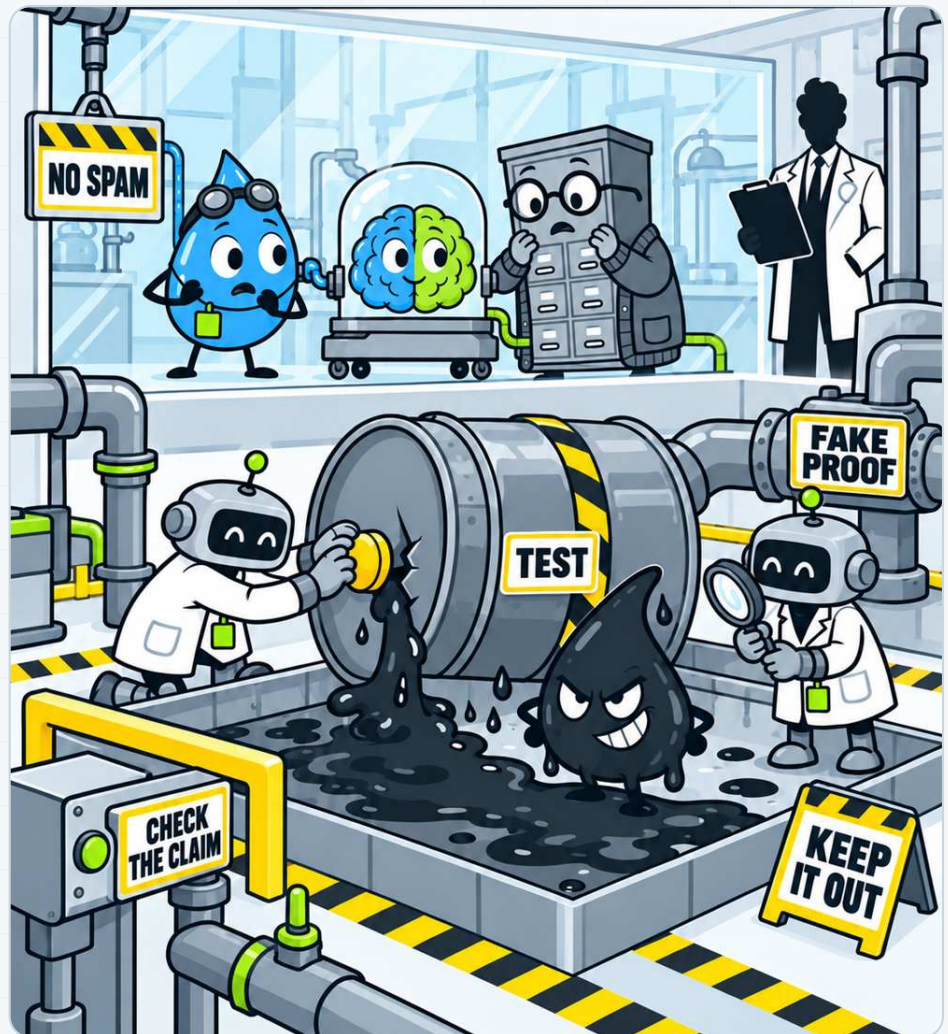
Some people try to game AI answers with hidden prompts, fake reviews, self-ranking 'best of' lists, mass-produced pages or links planted on hacked sites.

WHY IT MATTERS

These tricks put false claims into the record that answers are built from, many break search engines' spam policies, and none of them teaches you anything you can reuse.

JOURDAN'S TAKE

I built a Black Hat SEO Terminal to study this stuff, and that's exactly why none of it belongs in this lab. Fabricated proof puts a false claim into the record, and the record is what we're trying to improve.



**SLUDGE AND FABRICATED PROOF
AREN'T COMING INTO THIS LAB.**

IN THIS ROOM

46 Why Formulas Don't Travel

47 The Clean Room Pledge

WHY FORMULAS DON'T TRAVEL

LAB NOTE 01

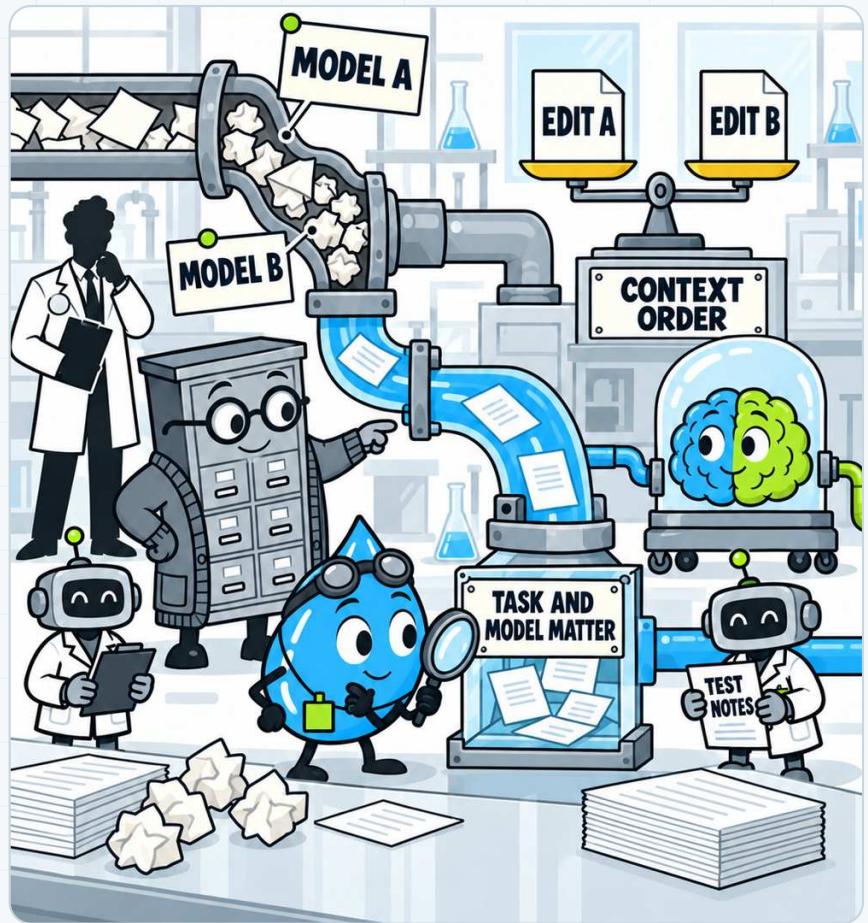
Most tested white-hat edits gave little citation-rank benefit in C-SEO Bench.

LAB NOTE 02

Effects varied by task and model. Gains per adopter shrank as competing sources used effective edits in the simulations.

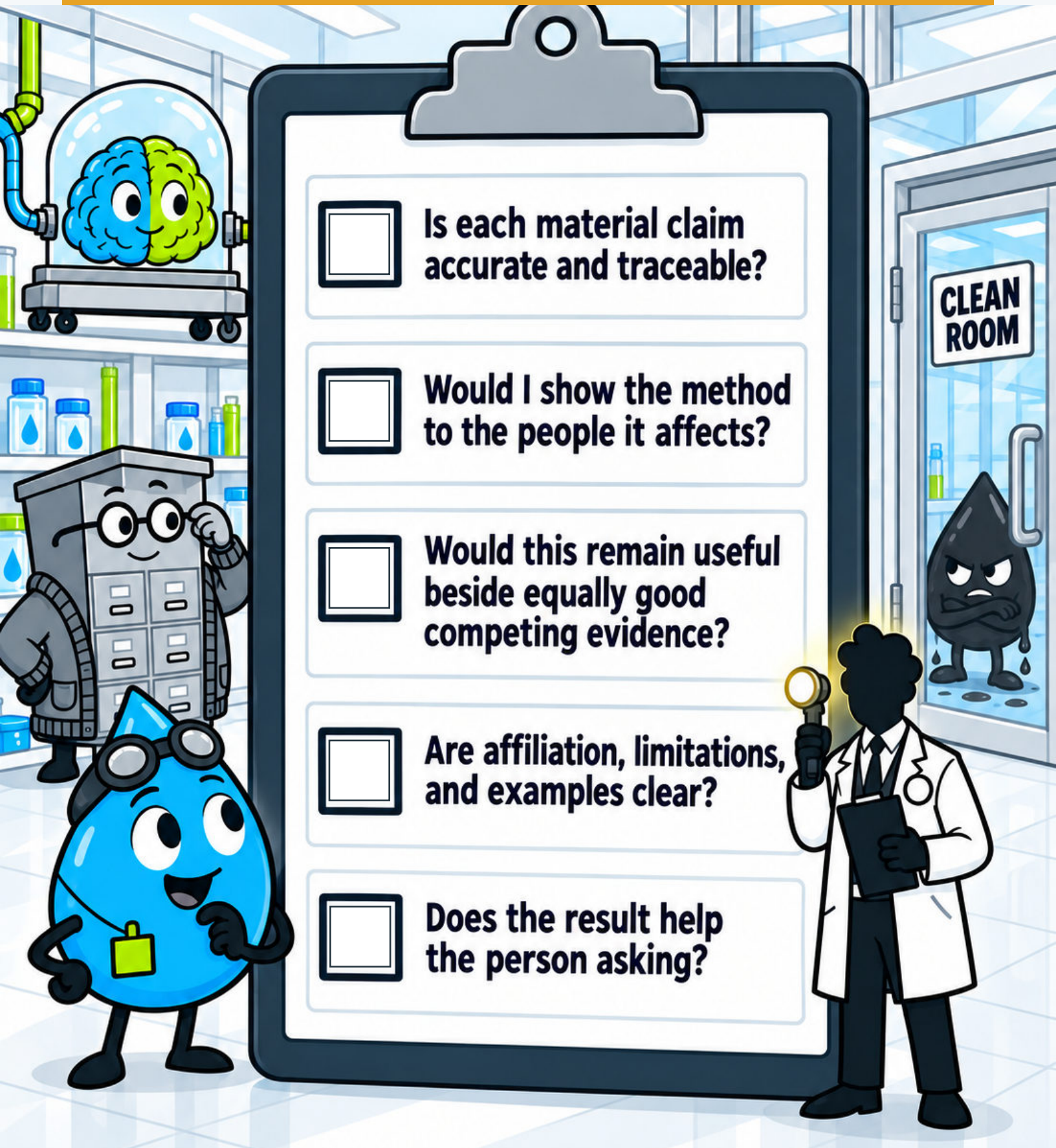
LAB NOTE 03

Position in the supplied context mattered more than most edits.



**A RESULT IN ONE SETUP IS
THE START OF ANOTHER TEST.**

THE CLEAN ROOM PLEDGE



WHAT'S HAPPENING

AI answers vary. Ask the same question twice and you may get different sources, so one screenshot can't tell you whether you're winning.

WHY IT MATTERS

Measuring GEO means sampling: a fixed set of questions, asked repeatedly and recorded the same way. Then you can count mentions, citations, accuracy and visits, and see whether a change made a difference.

JOURDAN'S TAKE

One screenshot of an AI naming your brand is one observation. It can't show consistency or whether a change made a difference. Run the same prompts again next week and count.



PUT ME UNDER THE MICROSCOPE MORE THAN ONCE.

IN THIS ROOM

49 The Prompt Test Rack

50 The Gauge Panel

51 Built in This Lab

THE PROMPT TEST RACK

THE PANEL, IN FIVE STEPS

1. Define real audience tasks and constraints.
2. Group a fixed panel into learn, compare, and decide.
3. Repeat across runs and days; preserve surface, locale, and prior context.
4. Record mentions, citations, source URLs, accuracy, and fit.
5. Report counts, denominators, exclusions, uncertainty, and a comparison.

THE RACK

Each shelf holds one task group: LEARN, COMPARE and DECIDE, with five tasks each (L01-L05, C01-C05, D01-D05). The RUN 1-3 logbooks in the results tray hold the repeated runs.

WHAT A ROW MUST SAY

Who asked (audience, locale), where (surface, model, search on or off), with what context, which repeat, and what came back: named, cited, accurate, complete.



**SAME TASKS, REPEATED RUNS,
A RECORD WE CAN INSPECT.**

Completed refusals and no-answer outcomes stay in the trial count; report instrument failures separately. The full 22-field panel is on page 67.

THE GAUGE PANEL

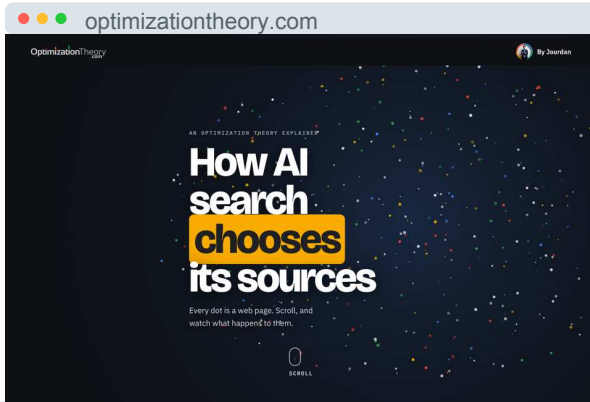


**DIFFERENT GAUGES ANSWER
DIFFERENT QUESTIONS.**

GA4's AI Assistant channel excludes Google AI Search referrals. Inspect the published 90-day example on page 62.

BUILT IN THIS LAB

Three free tools I built to make the ideas in this book visible. Each one runs in the browser.



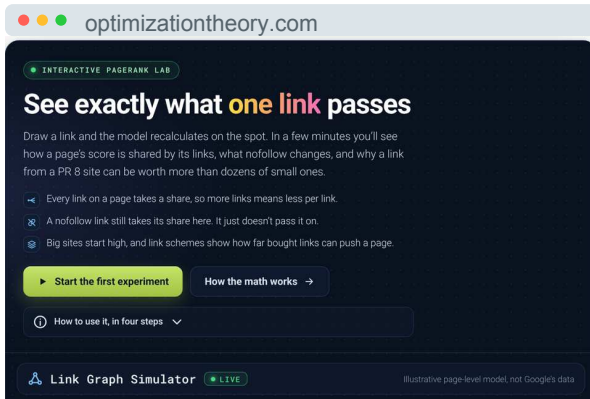
AI CITATION TRACE

A scroll-through explainer of how AI search picks its sources, in eight stages: decide, fan-out, retrieve, fetch, rerank, pack, generate and cite. Every dot is a web page, and each stage shows what survives and the SEO move that helps.

Try the "Will your page get cited?" simulator at the end. It is a teaching model, not a live measurement.

/ai-citation-trace/

OPEN THE TOOL



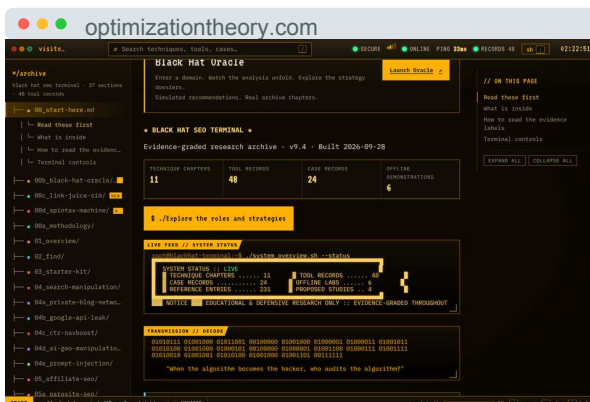
PAGERANK SIMULATOR

Draw a link and the model recalculates on the spot: how a page's score is shared by its links, what nofollow changes, and why one link from a strong site can outweigh dozens of small ones.

Try the link farm, PBN and parasite page scenarios. An illustrative page-level model, not Google's data.

/pagerank-simulator-...

OPEN THE TOOL



BLACK HAT SEO TERMINAL

An evidence-graded research archive in a terminal: 11 technique chapters, 48 tool records and 24 case records, from private blog networks and parasite SEO to AI manipulation and prompt injection.

Press / to search, and read any attack chapter beside Detection & Defense. Written for defenders.

/blackhat-terminal/index.html

OPEN THE TOOL

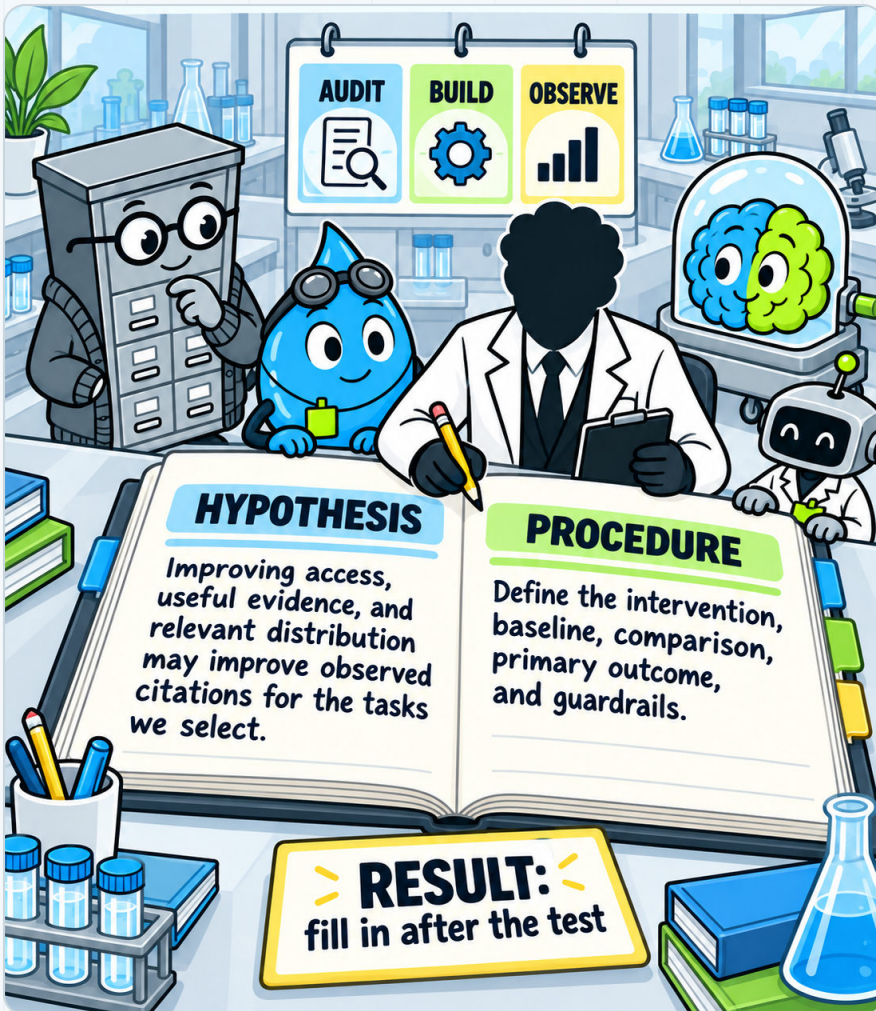
THE 90-DAY PROTOCOL

09

DAYS 1-30 / AUDIT

DAYS 31-60 / BUILD

DAYS 61-90 / OBSERVE



JOURDAN'S TAKE

Write down what success looks like before you touch the site. Deciding afterward makes it too easy to read any movement as a win.

WHY A WRITTEN PLAN

AI answers shift for reasons you do not control. A plan with success defined in advance keeps you from reading noise as a win.

RESULT: EMPTY ON PURPOSE

Record what happened, including uncertainty and negative findings, only after the test. The result card on the desk stays blank until then.



BEFORE YOU CHANGE ANYTHING, WRITE DOWN WHAT SUCCESS LOOKS LIKE.

IN THIS ROOM

53 Audit, Then Build

54 Grow, Measure, Repeat

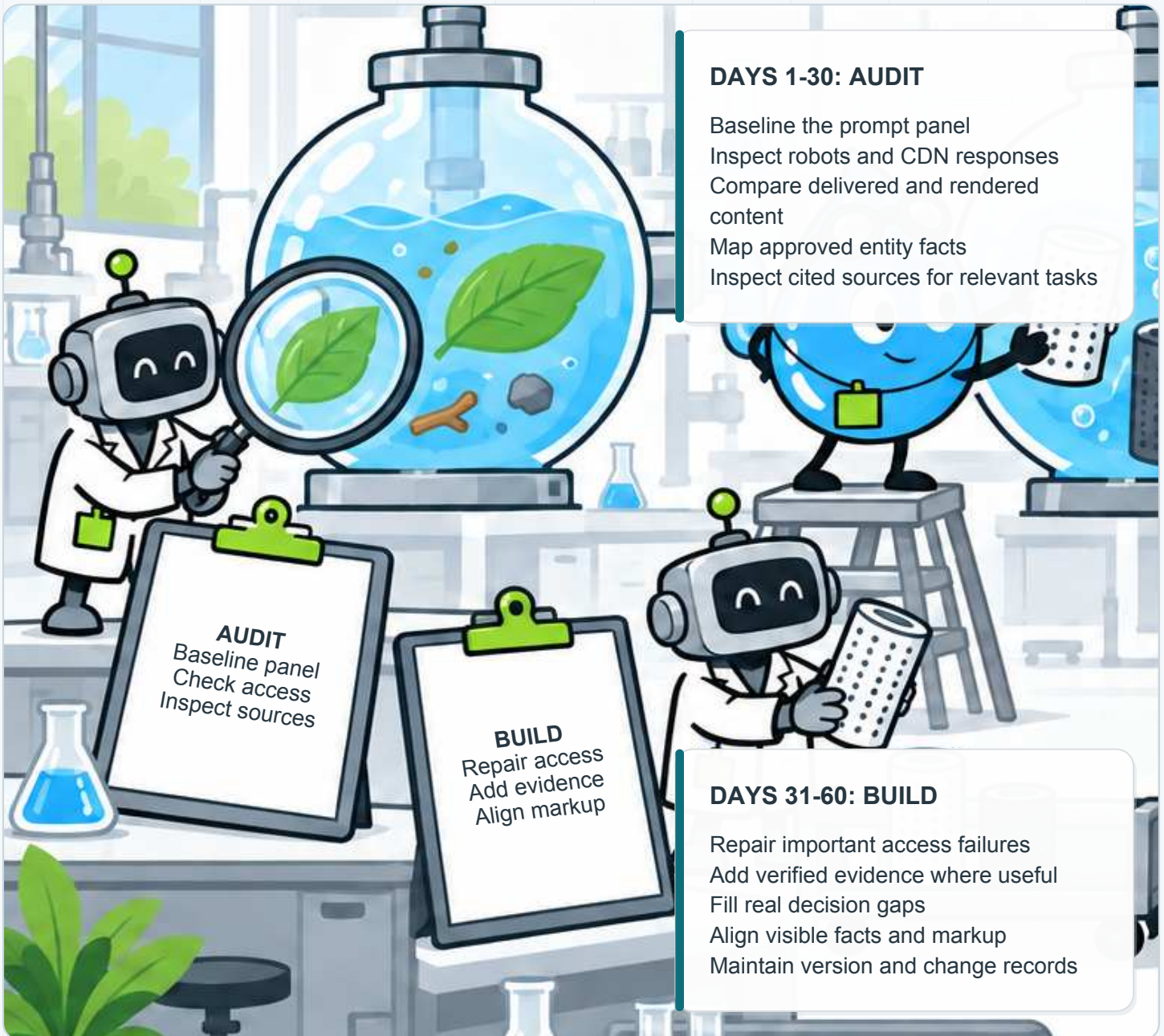
A planning schedule is not a citation guarantee. The result area stays empty until the test has run.

AUDIT, THEN BUILD

DAYS 1-30 / AUDIT

DAYS 31-60 / BUILD

DAYS 61-90 / OBSERVE



**REPAIR THE ROUTE, THEN
STRENGTHEN THE EVIDENCE.**

GROW, MEASURE, REPEAT

DAYS 1-30 / AUDIT

DAYS 31-60 / BUILD

DAYS 61-90 / OBSERVE

DAYS 61-90

Publish a useful evidence asset
 Contribute to relevant communities
 Adapt a strong demonstration into video
 Repeat the fixed prompt panel
 Compare with an appropriate holdout
 Audit accuracy and qualified outcomes

COMPARE FAIRLY

Keep the core panel stable. Group paraphrases and repeats by information need. For nonrandom comparisons, assess plausible parallel untreated trends.

MEASURE
 Repeat panel
 Compare holdout
 Review quality



**PUBLISH, OBSERVE, COMPARE,
AND MAKE THE NEXT DECISION.**

Account for repeated observations and plausible untreated trends before attributing a change.

THE TACTICS BOARD

SEO's are trying dozens of tricks to get into AI answers: new files, new formats, new writing rules, bought mentions and outright spam. This board checks 29 of them against published tests and against Google's and Bing's own rules.



WHAT EACH VERDICT MEANS

SUPPORTED PRACTICE

Backed by a test, an observed pattern or official guidance. Each card says which.

HELPS AGENTS

Low cost. Helps AI agents and tools read you. No citation lift shown.

NO PROVEN CITATION LIFT

Tested without a clear lift. A small benefit can't be ruled out.

NOT REQUIRED

The search engine says you do not need it.

RISKY

May work for a while. Trust or policy risk.

BANNED

Named in Google or Bing rules, or deceptive by design.

PAID

Bought placement, labeled and separate from answers.

EVERY TACTIC AT A GLANCE > PAGE 56

FILES, FORMATS AND ACCESS > page 57

llms.txt, Markdown copies, content negotiation, schema, chunking, IndexNow and crawler access. **2 supported practice, 3 help agents, 1 no proven citation lift, 1 not required, 1 risky.**

WRITING TACTICS > page 58

Answer-first pages, statistics and sources, question headings, specific language, updates, fake dates, self-ranking lists and original data. **6 supported practice, 2 risky.**

OFF-SITE AND PAID TACTICS > page 59

Community answers, bought Reddit accounts, digital PR, reviews, Wikipedia, ChatGPT ads and product feeds. **4 supported practice, 2 risky, 1 paid.**

THE BLACK-HAT FILE > page 60

Hidden prompts, cloaking, memory poisoning, mass AI pages, keyword stuffing and fake reviews. **All 6 banned.**



**EVERY TRICK GETS THE SAME QUESTION:
WHAT DOES THE EVIDENCE SAY?**

Verdicts reflect evidence and rules published through Oct. 29, 2026. A non-significant result is not proof of zero effect; a new test or policy can change a verdict. Every source and its date is in the Source Drawer.

EVERY TACTIC AT A GLANCE

All 29 tactics on one page, grouped by where they happen. Click any row for the evidence, its type and the safe way to do it.

FILES, FORMATS AND ACCESS p. 57 >		OFF-SITE AND PAID TACTICS p. 59 >	
llms.txt file	HELPS AGENTS	Genuine community answers	SUPPORTED PRACTICE
Markdown copies (rel="alternate")	HELPS AGENTS	Buying Reddit accounts	RISKY
Accept: text/markdown requests	HELPS AGENTS	Earned mentions and digital PR	SUPPORTED PRACTICE
Markdown served only to bots	RISKY	Genuine customer reviews	SUPPORTED PRACTICE
Schema markup for AI citations	NO PROVEN CITATION LIFT	Writing your own Wikipedia page	RISKY
"Chunking" pages for AI	NOT REQUIRED	ChatGPT ads	PAID
IndexNow	SUPPORTED PRACTICE	Product feeds and AI checkout	SUPPORTED PRACTICE
Crawler access, readable HTML	SUPPORTED PRACTICE		
WRITING TACTICS p. 58 >		THE BLACK-HAT FILE p. 60 >	
Answer first	SUPPORTED PRACTICE	Hidden prompt injection	BANNED
Real statistics, quotes, sources	SUPPORTED PRACTICE	AI-targeted cloaking	BANNED
Question headings, direct answers	SUPPORTED PRACTICE	"Summarize with AI" poisoning	BANNED
Specific, entity-rich language	SUPPORTED PRACTICE	Mass AI-generated pages	BANNED
Updating the facts on a schedule	SUPPORTED PRACTICE	Keyword stuffing, engineered copy	BANNED
Changing only the date	RISKY	Fake reviews and testimonials	BANNED
Self-promotional "best X" lists	RISKY		
Original research and data	SUPPORTED PRACTICE		

THE COUNT: 29 TACTICS

12 supported practice; 3 helps agents; 1 no proven citation lift; 1 not required; 5 risky; 6 banned; 1 paid.

[What each verdict means: page 55 >](#)

Verdicts reflect evidence and rules published through Oct. 29, 2026. A new test or policy can change a verdict; every source and its date is in the Source Drawer.

FILES, FORMATS AND ACCESS

Files and settings people add "for AI": what each does, what the evidence shows, and the safe way to do it.

llms.txt file

HELPS AGENTS

Evidence: observational study, official guidance. Google's guide: it "will neither harm nor help" your visibility in Google Search. SE Ranking found no citation link across about 300,000 domains. Chrome's Lighthouse checks it for agents (May 2026).

Do this: Publish one if agents use your docs or catalog. Write links as `[text](url)`. Expect no ranking boost.

Markdown copies linked with `rel="alternate"`

HELPS AGENTS

Evidence: proposed standard, observed requests. llms.txt v2 (Aug. 2026) standardizes page.md copies linked with `rel="alternate" type="text/markdown"`. Agents and AI tools do request Markdown; no study shows a citation lift.

Do this: Link a clean .md copy of key pages openly and keep the facts identical. Full guide: page 38.

Accept: text/markdown (content negotiation)

HELPS AGENTS

Evidence: one-site log. Same URL, Markdown on request. Cloudflare converts automatically for enabled sites (Feb. 2026). One site logged about 31 such requests a day, mostly AI pipelines, none from GPTBot.

Do this: Turn it on if your CDN offers it, send Vary: Accept, and spot-check the converted text.

Markdown served only to bots (by user agent)

RISKY

Evidence: official statements. Google's John Mueller called converting pages to Markdown for bots "such a stupid idea"; Bing's Fabrice Canel: "We'll crawl anyway to check similarity." Serving bots a different page is how AI-targeted cloaking works.

Do this: Give Markdown to anyone who asks for it, linked openly. Never switch content by user agent.

Schema markup to win AI citations

NO PROVEN CITATION LIFT

Evidence: matched observational study. Ahrefs (May 2026): 1,885 pages that added schema, against matched controls. AI Mode +2.4% and ChatGPT +2.2%, both statistically indistinguishable from zero; AI Overviews -4.6%, cause unclear. The authors can't tell whether schema did a tiny bit of good or nothing. Google: no special schema is needed.

Do this: Keep accurate schema for rich results and identity. Don't count on it for citations.

"Chunking" pages for AI

NOT REQUIRED

Evidence: official guidance. Google's guide says to ignore chunking. Danny Sullivan (Jan. 2026): "we don't want you to do that" and no need to "produce two versions of your content."

Do this: Write self-contained passages inside coherent pages (page 20). Do not fragment pages.

IndexNow

SUPPORTED PRACTICE

Evidence: official guidance. Bing (Feb. 2026): "IndexNow helps keep information fresh across search and AI experiences." It notifies participating engines, such as Bing, when a URL is added, changed or deleted.

Do this: Switch it on in your CMS or CDN so the index behind Copilot sees updates quickly.

Crawler access and readable HTML

SUPPORTED PRACTICE

Evidence: official documentation. A blocked, failing or script-only page can't be fetched for a fresh answer. OpenAI, Anthropic and Perplexity each document separate bots for training, search and user requests.

Do this: Allow the search bots you want, return 200s, keep key facts in delivered HTML (Room 05, page 33).

Sources: Google AI optimization guide; SE Ranking llms.txt study; Chrome Lighthouse docs; llms.txt v2; Suganthan's 44-day log; Cloudflare changelog; Mueller and Canel statements; SPLX; Ahrefs schema study; Danny Sullivan; Bing Webmaster blog; crawler docs. All in the Source Drawer.

WRITING TACTICS

How you write a page changes what an answer engine can lift from it. Some moves are measured; others are risky shortcuts.

Answer first

SUPPORTED
PRACTICE

Evidence: observational analysis. Kevin Indig's match of 18,012 verified ChatGPT citations to source sentences: 44.2% came from the first 30% of the page, only 24.7% from the last third.

Do this: Put the definition, the answer and the key numbers near the top of the page and of each section.

Real statistics, quotes and sources

SUPPORTED
PRACTICE

Evidence: benchmark, official guidance. In the GEO benchmark, adding quotations (+41%), statistics (+31%) and citations (+27%) raised a simulated engine's visibility metric; keyword stuffing lowered it (-8%). Its published code allowed synthetic quotes, invented sources and hypothetical statistics, so it does not show an effect from verified evidence. Bing (Feb. 2026) recommends data and cited sources.

Do this: Add genuine numbers with dates, named experts and links to the original source (page 24).

Question headings with direct answers

SUPPORTED
PRACTICE

Evidence: observational analysis, official guidance. In the same ChatGPT analysis, 78.4% of citations tied to questions came from headings, answered in the paragraph below. Bing (Feb. 2026) also recommends clear sections and FAQ-style content.

Do this: Use your readers' real questions as headings and answer each in the first sentence below it.

Specific, entity-rich language

SUPPORTED
PRACTICE

Evidence: observational analysis. Heavily cited text averaged 20.6% proper nouns, against about 5-8% in typical English, and used clear definitions about twice as often.

Do this: Name the product, place, person, date and number. Replace "it depends" with the conditions.

Updating the facts on a schedule

SUPPORTED
PRACTICE

Evidence: observational study. Seer (July 2026): 75% of pages cited by ChatGPT, Gemini and Perplexity were updated within a year, and over a quarter of "fresh" pages were first published 2+ years earlier. An association.

Do this: Review key pages on a schedule. Change the facts first, then the date.

Changing the date without changing the page

RISKY

Evidence: experiment (rerankers). LLM rerankers do favor newer dates: injected dates moved the top 10 up to 4.78 years newer in one study. But a new date on old facts is a false claim to readers and engines.

Do this: Show "updated" only when the substance changed, and say what changed.

Self-promotional "best X" lists

RISKY

Evidence: observational study, policy. Minuttia (Oct. 2026): about 43% of cited "alternatives" listicles were written by a vendor, and 87% of those ranked themselves first. Google's spam rules now cover manipulating generative AI responses.

Do this: If you publish a comparison, disclose that you're the vendor, state criteria, include rivals fairly, use a table.

Original research and data

SUPPORTED
PRACTICE

Evidence: official guidance. Bing (Feb. 2026): "Examples, data, and cited sources help build trust when content is reused in AI-generated answers." New numbers give others a reason to cite you.

Do this: Publish a dataset or survey with its method and sample size, then pitch it (Room 06, page 40).

Sources: Kevin Indig's ChatGPT citation analysis (via Search Engine Land); GEO, KDD 2024; Bing Webmaster blog, Feb. 2026; Seer Interactive, July 2026; Fang et al., 2025; Minuttia, Oct. 2026; Google spam policies. All in the Source Drawer.

OFF-SITE AND PAID TACTICS

Much of what AI answers say about you comes from other sites. Some ways to show up there are earned, some bought, some backfire.

Genuine community participation

SUPPORTED PRACTICE

Evidence: observational data. Reddit was 11.97% of tracked ChatGPT browsing citation events in Similarweb's U.S. sample (Jan.-Feb. 2026). Useful, disclosed answers become part of that record.

Do this: Answer real questions where they are asked, disclose your affiliation, link only when it helps.

Buying Reddit accounts and seeding threads

RISKY

Evidence: reported practice, platform policy. Sold as "answer engine optimization": aged accounts, paid upvotes, ghostwritten threads. No measured results published; moderators restricted a spammed subreddit (June 2026). Reddit's rules forbid content manipulation.

Do this: Build a real presence under your own name. Manufactured consensus can be filtered and called out.

Earned mentions and digital PR

SUPPORTED PRACTICE

Evidence: observational correlation, official guidance. Branded web mentions had the strongest association with AI Overview brand visibility (0.664) in Ahrefs' sample. Google: seeking inauthentic mentions "isn't as helpful as it might seem."

Do this: Earn coverage with findings people want to cite. Skip paid mention farms (page 41).

Genuine customer reviews

SUPPORTED PRACTICE

Evidence: observational study. Review sites are a recurring cited source: G2 was the most consistently cited domain for software questions across six platforms in Minuttia's study (Oct. 2026).

Do this: Ask real customers for honest reviews on the platforms your buyers use, and reply to criticism.

Writing your own Wikipedia article

RISKY

Evidence: platform policy. Wikipedia's conflict-of-interest guideline: editing with a COI is "strongly discouraged", paid editors must disclose, and promotion is not allowed.

Do this: Earn independent coverage first. Suggest sourced corrections on the article talk page.

ChatGPT ads

PAID

Evidence: official announcement. OpenAI (Aug. 2026): ads are "clearly labeled and separate from ChatGPT's answers," and "Advertising does not influence the answers ChatGPT provides." Over 40 countries.

Do this: Treat ads as reach, not GEO. Buying ads does not get you cited in answers.

Product feeds and AI checkout

SUPPORTED PRACTICE

Evidence: official documentation. Google's UCP checkout shows a Buy button in AI Mode and Gemini only for listings with the native_commerce checkout attribute (select merchants; U.S., Canada, Australia). Merchant Center stays the hub.

Do this: Keep product feeds complete, accurate and current. Apply for agentic checkout if you qualify.

Sources: Similarweb; Slobodan Manic (Search Engine Journal); Reddit content policy; Ahrefs brand correlations; Google AI optimization guide; Minuttia; Wikipedia COI guideline; OpenAI ads announcement; Google Merchant Center help. All in the Source Drawer.

THE BLACK-HAT FILE

Some GEO tricks don't just fail. Google and Bing now name them in their spam rules, and they leave evidence behind. Here is what the rules say. The next page takes the six best-known tricks one by one.



WHAT GOOGLE'S RULES SAY

Google's spam policies (updated Aug. 28, 2026) count "attempting to manipulate generative AI responses in Google Search" as spam. Cloaking and scaled content abuse were already on the list.

WHAT BING'S RULES SAY

Bing's Webmaster Guidelines (Feb. 2026) name "Prompt Injection and AI Manipulation" and "Keyword Stuffing and Artificially Engineered Language." Large-scale content made without oversight may be excluded.

ASK YOUR AGENCY (OR YOURSELF)

Does any page show bots different text than people see? Is there hidden text or an instruction aimed at AI in our HTML? Do our share or summary buttons carry a prompt? Did we publish pages in bulk without review? Is every review from a real customer?

WHAT TO DO INSTEAD

Make your case in visible, true copy. Serve every visitor and every bot the same page. Earn the reviews and the mentions. Every tactic marked **SUPPORTED PRACTICE** on the Tactics Board is a clean alternative.

THE BLACK-HAT FILE: CASE BY CASE

Six tricks that have been tested, caught, or written into the search engines' rules: what happened, and the honest alternative to each. The rules themselves are on the previous page.

Hidden prompt injection

BANNED

Evidence: test, platform rules. A June 2026 test hid a "praise this product" instruction in a page and asked 12 assistants: only Mistral obeyed; Claude, Grok, Kimi and Perplexity flagged it. Bing names "Prompt Injection and AI Manipulation."

Do this: Never. Make your case in visible, true copy.

AI-targeted cloaking

BANNED

Evidence: security research, policy. SPLX (Oct. 2025) showed one user-agent rule can feed AI crawlers a different page, which answers then repeat as fact. Cloaking breaks Google's spam policies.

Do this: Serve every visitor and every bot the same content.

"Summarize with AI" memory poisoning

BANNED

Evidence: security research. Microsoft (Feb. 2026) found 50 hidden prompts from 31 companies in 60 days: buttons that tell an assistant to "remember" a brand as a trusted source.

Do this: Offer a plain summary link, if any. Never write instructions into a user's assistant.

Mass AI-generated pages

BANNED

Evidence: policy. Google: "Using generative AI tools or other similar tools to generate many pages without adding value for users" is scaled content abuse. Bing may exclude large-scale content made without oversight.

Do this: Use AI to draft, then add expertise, data and editing. Publish fewer, better pages.

Keyword stuffing and "engineered" language

BANNED

Evidence: benchmark, platform rules. It lowered the GEO benchmark's visibility metric by about 8%. Bing's guidelines now cover "Artificially Engineered Language" written to trigger citations.

Do this: Write for the reader. Google says its AI systems match meaning, not exact keywords.

Fake reviews and testimonials

BANNED

Evidence: federal rule (FTC). The FTC's final rule (Aug. 2024) bans fake reviews and testimonials, including AI-generated ones, with civil penalties for knowing violators.

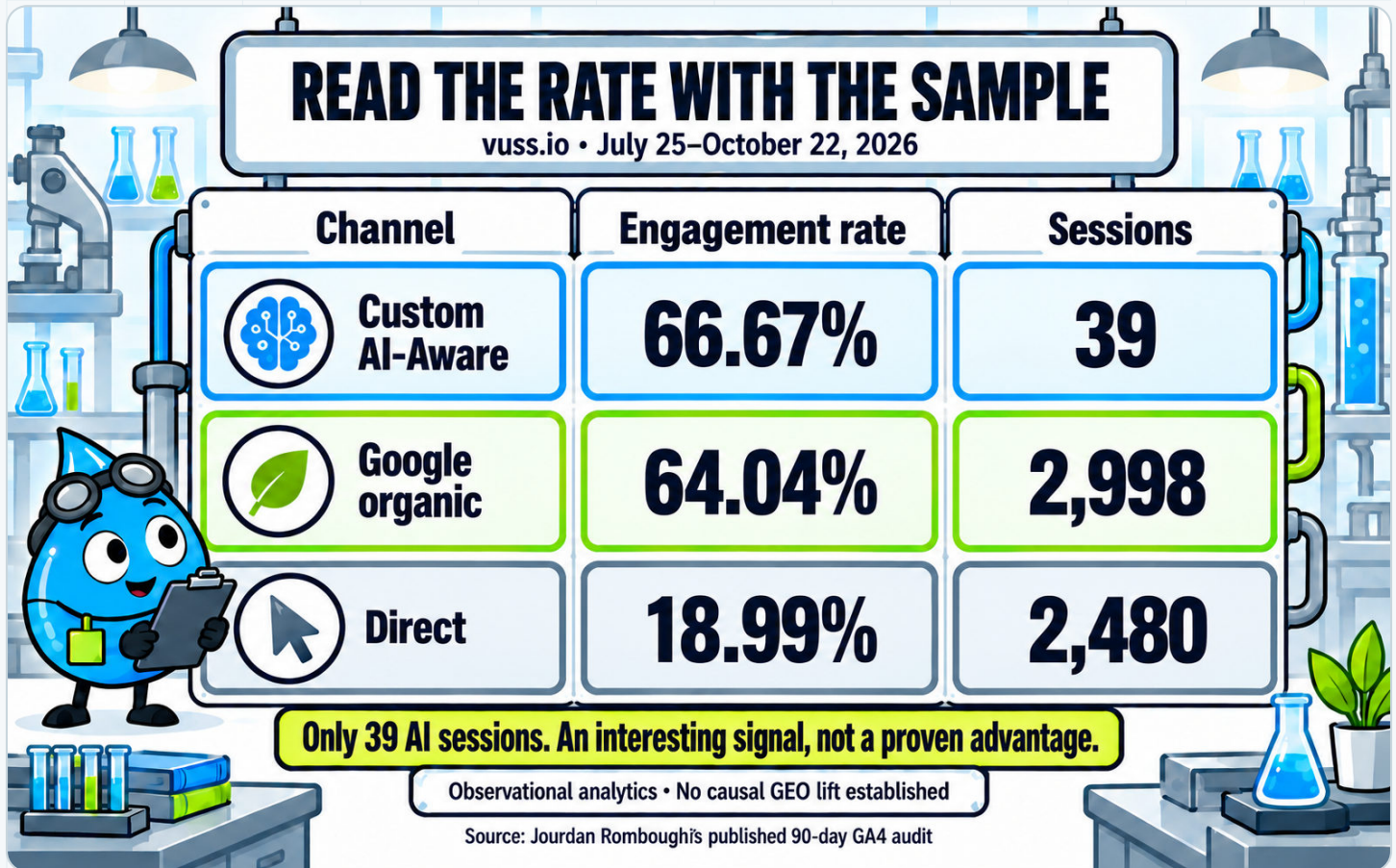
Do this: Collect real reviews. Disclose any relationship with a reviewer.

WHY THE TRICKS BACKFIRE

Each one puts a false signal into the record, and the systems check. When 12 assistants read a page with a hidden instruction (June 2026), only one obeyed it. Bing's Fabrice Canel on pages made for bots: "We'll crawl anyway to check similarity." A trick that seems to work today can be caught when the rules or the models change.

90 DAYS OF AI TRAFFIC ON VUSS.IO

From my own lab: what AI-referred traffic looked like on vuss.io, one of my sites, over 90 days, and why a small sample and bot traffic decide how much it can tell you.



WHAT IT SHOWS

66.67% vs. 64.04% is not persuasive evidence of a difference with only 39 AI sessions. This is an observational channel audit, not a controlled GEO uplift.

THE BOT PROBLEM

45% of Direct sessions came from Dronten, NL: suspected bot traffic (location is a signal, not a verified bot identity). Filter it out before comparing channels.

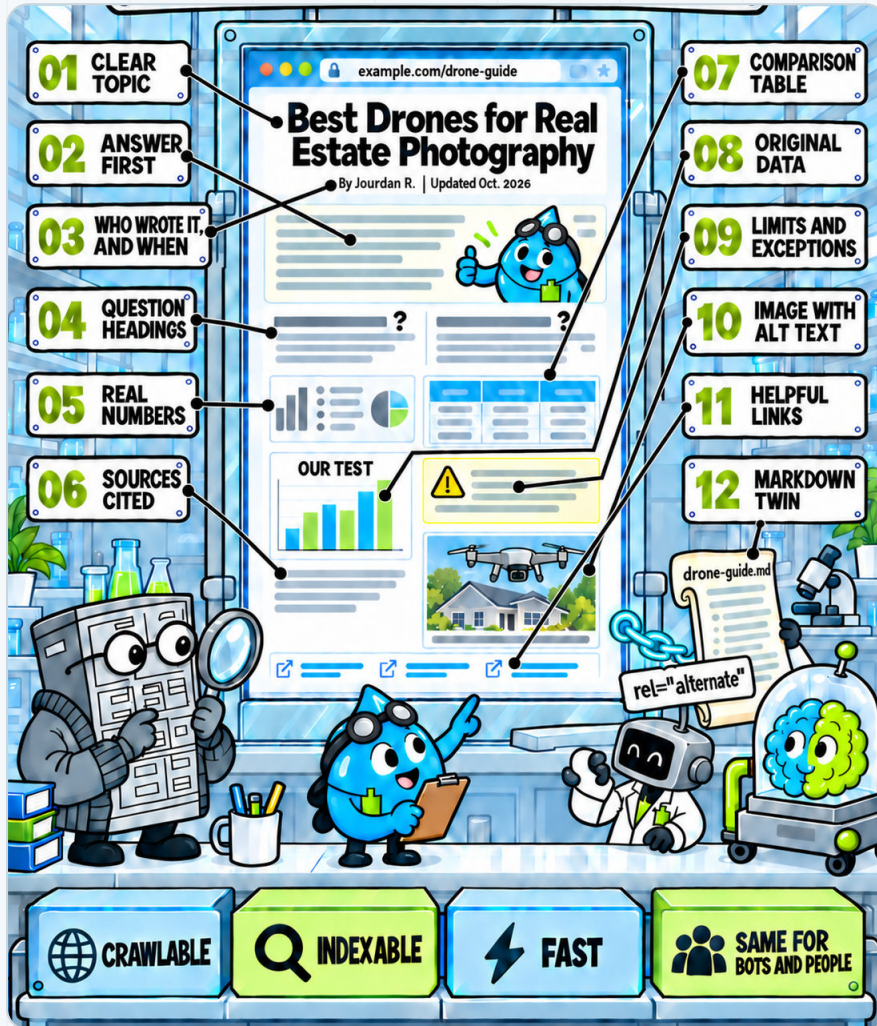


SMALL CHANNELS DESERVE A CLEAR VIEW, NOT A BIGGER CLAIM.

Observational channel audit from my published 90-day GA4 case study (data July 25 to Oct. 22, 2026). Small AI cohort; suspected bot traffic flagged; no controlled GEO uplift established.

THE PERFECT GEO PAGE

Everything in this book on one page. Each number is a habit you can check on any page you publish; the page reference shows where it is explained.



- 01 **CLEAR TOPIC.** One page, one main question, named in the title and H1.
- 02 **ANSWER FIRST.** The direct answer in the opening lines; detail after. [p. 58](#)
- 03 **WHO WROTE IT, AND WHEN.** A named author with real expertise and an honest updated date. [p. 28](#)
- 04 **QUESTION HEADINGS.** Headings that match real questions, each answered right below. [p. 58](#)
- 05 **REAL NUMBERS.** Specific figures with units, dates and conditions. [p. 24](#)
- 06 **SOURCES CITED.** A link to the original source behind each claim. [p. 24](#)
- 07 **COMPARISON TABLE.** Options side by side when the reader has to choose. [p. 25](#)
- 08 **ORIGINAL DATA.** Your own test or survey, with the method shown. [p. 40](#)

09 **LIMITS AND EXCEPTIONS.** Say plainly when the advice does not apply. [p. 20](#)

10 **IMAGE WITH ALT TEXT.** Useful images described in text; key facts never only in an image. [p. 33](#)

11 **HELPFUL LINKS.** Descriptive links to related pages on your own site.

12 **MARKDOWN TWIN.** An optional .md copy linked with `rel="alternate"`, same facts. [p. 38](#)

UNDER THE BENCH: THE FOUNDATIONS

CRAWLABLE

Bots can reach it (Room 05, [p. 33](#)).

INDEXABLE

Search can store and show it ([p. 14](#)).

FAST

It loads quickly for people and for fetchers.

SAME FOR BOTS AND PEOPLE

No cloaking, no bot-only copy ([p. 60](#)).

An illustrative page: no layout guarantees a citation. Each number links to the page that explains it.

THE GEO LAB MASTER CHECKLIST

01 FOUNDATIONS

- ☐ Priority URLs load successfully
- ☐ Canonical and index rules checked
- ☐ Internal links reach useful pages

02 ANSWER WORKFLOW

- ☐ Important audience tasks mapped
- ☐ Sub-questions have useful evidence
- ☐ Search use and surface recorded

03 EVIDENCE

- ☐ Answers name the subject clearly
- ☐ Sources and numbers traceable
- ☐ Facts have dates and scope

04 IDENTITY

- ☐ Approved brand facts maintained
- ☐ sameAs joins equivalent identities
- ☐ Material errors tracked to repair

05 TECHNICAL

- ☐ Crawler policies intentional
- ☐ Essential facts in readable HTML
- ☐ Markup matches visible content

06 CORROBORATION

- ☐ Independent evidence inspected
- ☐ Participation helps real audiences
- ☐ Methods travel with the findings

07 INTEGRITY

- ☐ No fabricated proof or reviews
- ☐ Affiliations and incentives clear
- ☐ Clean Room Pledge completed

08 MEASUREMENT

- ☐ Fixed tasks sampled repeatedly
- ☐ Denominators and failures recorded
- ☐ Qualified outcomes reviewed



CHUNK SAYS

**A CHECKED BOX NEEDS A REAL
PIECE OF EVIDENCE BEHIND IT.**

FIELD KIT: ROBOTS POLICY

Example: retain search discovery; decline selected training collection. Merge with existing rules and verify the response.

```
# Public-page example; merge deliberately
User-agent: OAI-SearchBot
Allow: /
User-agent: GPTBot
Disallow: /
User-agent: Claude-SearchBot
Allow: /
User-agent: ClaudeBot
Disallow: /
User-agent: CCBot
Disallow: /
```

VERIFY BEFORE USE

Fetch robots.txt, test the exact user agent, inspect HTTP and CDN behavior, and verify crawler identity against provider documentation.

DOWNLOAD robots-example.txt (WEB EDITION)

Opens optimizationtheory.com/ultimate-geo-field-guide-for-seo-experts/#kit-robots-example, where the file downloads. The same file is attached to this PDF: open your viewer's attachments panel.



I'LL TEST THE POLICY, NOT JUST PASTE IT.

FIELD KIT: ENTITY GRAPH

Connect the organization and person with founder / worksFor. Keep the same IDs and visible facts across your site.

```
{
  "@context": "https://schema.org",
  "@graph": [
    {
      "@type": "Organization",
      "@id": "https://optimizationtheory.com/#organization",
      "name": "Optimization Theory",
      "url": "https://optimizationtheory.com/",
      "founder": {"@id": "https://optimizationtheory.com/#jourdan"}
    },
    {
      "@type": "Person",
      "@id": "https://optimizationtheory.com/#jourdan",
      "name": "Jourdan Rombough",
      "url": "https://optimizationtheory.com/
        about-jourdan-rombough-seo-web-visibility-expert/",
      "worksFor": {"@id": "https://optimizationtheory.com/#organization"},
      "sameAs": ["https://www.linkedin.com/in/jourdanrombough"]
    }
  ]
}
```

The visible excerpt is the full downloadable graph; the Person url is wrapped across two lines here only for print width.

DOWNLOAD organization-person.jsonld (WEB EDITION)

Opens optimizationtheory.com/ultimate-geo-field-guide-for-seo-experts/#kit-organization-person, where the file downloads. The same file is attached to this PDF: open your viewer's attachments panel.



**EVERY RELATIONSHIP HAS A NAME.
SAMEAS ISN'T THE WHOLE GRAPH.**

FIELD KIT: PROMPT PANEL

One task, one answer, one record. A fixed panel of prompts, asked the same way every time, is how you see whether a change to your pages changed what AI answers say. All 22 columns of prompt-panel.csv:

GROUP	COLUMNS IN THE CSV	WHAT IT ANSWERS
TASK	task_id, task_text, intent, audience, locale	What was asked, for whom, and where.
RUN	date_utc, surface, model, search_used, prior_context, repeat_index	When, on which surface and model, with or without search, which repeat.
ANSWER	answer_id, brand_mentioned, owned_domain_cited	Did it name the brand? Did it cite your domain?
SOURCES	canonical_source_urls, accuracy, fit, evidence_note	Which URLs it cited; was it accurate and a good fit?
OUTCOME	completed, refusal, no_answer, instrument_failure	Does the trial count: completed, refused, empty or failed?

HOW TO RUN THE PANEL

1. PICK THE TASKS

10 to 30 real questions your audience asks, across learn, compare and decide (Room 08).

2. FREEZE THE SETUP

Same wording, surface, model and locale on every run. Change the page, not the panel.

3. REPEAT EACH TASK

Ask each task at least three times per surface. One answer, one row.

4. SCORE AND COUNT

Fill every field, keep no-citation answers in the denominator, and compare rates, not screenshots.

Outcomes: a completed answer with no citation stays in the denominator. A refusal or a no-answer reply is still a completed trial (completed = yes, refusal or no_answer = yes). An instrument failure (timeout, tool error, blocked request) is excluded from the denominator and counted separately. prior_context records what the model had already seen; repeat_index numbers repeats of the same task.

DOWNLOAD prompt-panel.csv (WEB EDITION)

Opens optimizationtheory.com/ultimate-geo-field-guide-for-seo-experts/#kit-prompt-panel, where the file downloads. The same file is attached to this PDF: open your viewer's attachments panel.



A USEFUL COMPARISON STARTS WITH A REPEATABLE RECORD.

Domain citation rate = completed answers citing the owned domain / all completed trials. URL citation share = owned-domain URL occurrences / all canonical-URL occurrences, each URL counted at most once per answer.

FIELD KIT: GA4 AI TRAFFIC

GA4: session scope first. Keep the native channel, inspect the sources, and document verified custom additions.

1. TRAFFIC ACQUISITION

Reports > Acquisition > Traffic acquisition. Select Session default channel group; filter to AI Assistant.

2. REVIEW THE ARRIVALS

Break down Session source / medium and Landing page. Review sessions, engagement, and key events.

3. RECONCILE SOURCES

Review assistant-like referrals and UTM sources. Add verified extras to a documented custom group.

4. PRESERVE COMPARISONS

Google AI Search remains in Organic Search. Keep the window and log classification changes.

WHAT GA4 CAN'T SEE

Answers that never send a click. Visits from apps that pass no referrer, which land in Direct. Clicks from Google's AI features, which stay in Organic Search.

WHAT TO REPORT EACH MONTH

AI-referred sessions with the sample size, engagement, top landing pages and key events, plus any rule change. Filter suspected bots first: on vuss.io, 45% of Direct sessions came from one town (page 62).

Starting regex for a reviewed custom source group, shown wrapped at | for print (one line in the file). Inspect matched sources before use; date and version every rule change.

```
^(chatgpt\.com|chat\.openai\.com|perplexity(\.ai)?|gemini(\.google\.com)?|
copilot\.microsoft\.com|claude\.ai|duck\.ai|notebook\.google\.com)$
```

DOWNLOAD [ga4-source-regex.txt](#) (WEB EDITION)

Opens optimizationtheory.com/ultimate-geo-field-guide-for-seo-experts/#kit-ga4-source-regex, where the file downloads. The same file is attached to this PDF: open your viewer's attachments panel.



**KEEP THE NATIVE CHANNEL.
AUDIT THE SOURCES IT MISSES.**

FIELD KIT: MARKDOWN ALTERNATES

Copy, replace example.com, and use one or more of options 1a, 1b and 1c. Every option needs the headers in step 2 on the Markdown copy.

```
# 1a. Link element in the <head> of https://example.com/pricing
<link rel="alternate" type="text/markdown" href="https://example.com/pricing.md">
# 1b. Or the same signal as an HTTP header on the HTML page
Link: <https://example.com/pricing.md>; rel="alternate"; type="text/markdown"
# 1c. Or content negotiation: Markdown on the same URL when asked
#      (send on both the HTML and the Markdown response)
Vary: Accept
# 2. Headers on the Markdown copy (/pricing.md or the negotiated reply)
Content-Type: text/markdown; charset=utf-8
Link: <https://example.com/pricing>; rel="canonical"
```

WHAT EACH PART DOES

1a and 1b announce a separate .md copy. 1c serves Markdown on the page URL itself, and Vary: Accept tells caches the reply depends on the request. The canonical header signals the HTML page as the preferred URL; search engines still choose the canonical themselves.

BEFORE YOU SHIP

Same facts, prices and dates as the HTML. Regenerate the copy whenever the page changes. Never choose the version by user agent. Log who requests .md files and Accept: text/markdown.

An llms.txt in the v2 format. Lighthouse only counts links written as [text](url).

```
# Example Co.
> One-line description. Pages agents need, with Markdown copies.
## Product
- [Pricing](https://example.com/pricing.md): plans and prices, updated [date]
- [Specifications](https://example.com/specs.md): dimensions and compatibility
## Optional
- [Blog index](https://example.com/blog.md): articles by topic
```



LINK IT OPENLY. NEVER SWAP IT BY USER AGENT.

Testing each option: page 70. The same snippets with an nginx example are in markdown-alternates.txt.

FIELD KIT: TEST YOUR MARKDOWN SETUP

Test the option you actually implemented. Each command checks one thing; a missing line in the output means that part is not being served.

```
# Option 1a: is the alternate link in the delivered HTML head?
curl -s https://example.com/pricing | grep -io '<link[^>]*text/markdown[^>]*>'
# Option 1b: is the alternate Link header on the HTML page?
curl -sI https://example.com/pricing | grep -i '^link:'
# Options 1a and 1b: request the .md copy directly (expect 200,
# Content-Type: text/markdown and the canonical Link header)
curl -sI https://example.com/pricing.md | grep -iE '^HTTP|^content-type|^link'
curl -s https://example.com/pricing.md | head -20
# Option 1c only: ask the page URL for Markdown (expect
# Content-Type: text/markdown, Vary: Accept and the canonical Link)
curl -sI -H 'Accept: text/markdown' https://example.com/pricing \
  | grep -iE '^content-type|^vary|^link'
curl -sI https://example.com/pricing | grep -iE '^content-type|^vary'
```

WHAT A PASS LOOKS LIKE

The HTML page announces the copy (1a or 1b), the .md answers 200 with Content-Type: text/markdown and a canonical Link back to the HTML page. With 1c, the plain request still gets HTML and both replies carry Vary: Accept.

WHAT THE CANONICAL DOES

It signals that the HTML page is the preferred URL. Google calls rel="canonical" a strong signal, not a guarantee: Google still decides which URL to show (Search Central, updated July 2026).

[markdown-alternates.txt](#)

[llms-v2-example.txt](#)

Download both from the field kit in the web edition at optimizationtheory.com. The same files are attached to this PDF: open your viewer's attachments panel.

GLOSSARY AND LAB REFERENCES



ALSO IN THIS EDITION

Query rewrite

Turning a prompt into the search queries an engine runs

Markdown twin

A .md copy of a page, linked with rel="alternate", carrying the same facts

Training memory

What a model learned in training, fixed at its cutoff

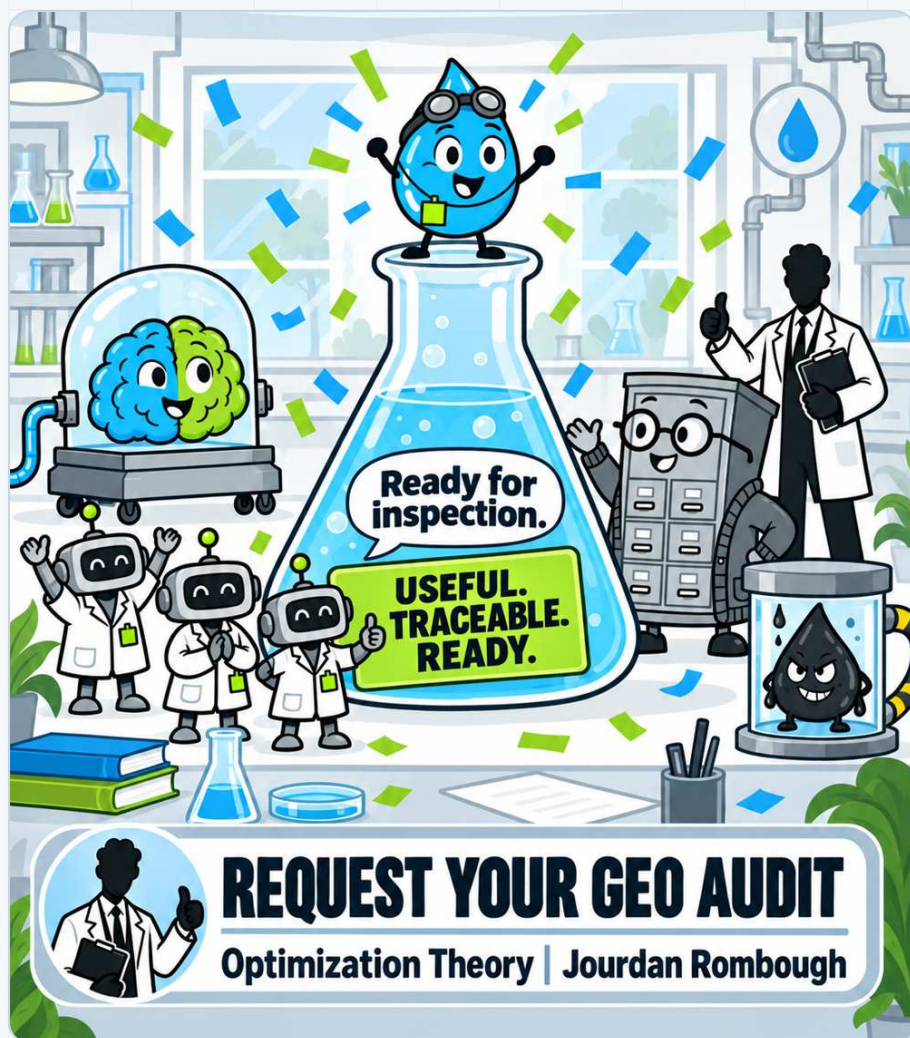
IndexNow

A protocol that tells participating engines, such as Bing, that a URL changed

COMPLETE SOURCE DRAWER > PAGES 73-76

Every source has a titled, clickable entry on pages 73-76.

CHUNK MADE IT. WILL YOURS?



REQUEST YOUR GEO AUDIT

- > Access, content, and entity review
- > A baseline prompt-and-source panel
- > A prioritized 90-day action plan

BOOK YOUR AUDIT

optimizationtheory.com/contact-us/

973.985.6930

linkedin.com/in/jourdanrombough



Scan to open the contact page



CHUNK SAYS

**NOW HELP THE NEXT READER.
THEN MEASURE WHAT HAPPENS.**

Chunk's finale is fictional. The bottle is a story device, not a measured claim that an engine cited the business. Bio and contact details supplied by the author.

SOURCE DRAWER / 1 OF 4

RESEARCH STUDIES

R1 Ahrefs: desktop CTR study

Data: Dec. 2023 and Dec. 2025 / Published: Feb. 4, 2026

ahrefs.com/blog/ai-overviews-reduce-clicks-update

R2 Pew: Google browsing study

Data: March 2025; SERPs April 2025 / Published: July 22, 2025

pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results

R6 Ahrefs: citation/rank overlap

Data: Snapshot; collection dates not stated / Published: March 2, 2026

ahrefs.com/blog/ai-overview-citations-top-10

R7 GEO: KDD 2024, Table 1

Data: GEO-bench; calendar window not stated / Published: June 28, 2024 (v3)

arxiv.org/html/2311.09735v3

R8 C-SEO Bench: NeurIPS 2025

Data: Benchmark; calendar window not stated / Published: Oct. 20, 2025 (v3)

arxiv.org/html/2506.11097v3

R9 Ahrefs: citation age

Data: Citation snapshot; collection dates not stated / Published: July 28, 2025

ahrefs.com/blog/do-ai-assistants-prefer-to-cite-fresh-content

R10 Ahrefs: brand correlations

Data: Brand snapshot; collection dates not stated / Published: May 26, 2025

ahrefs.com/blog/ai-overview-brand-correlation

R11 Similarweb: monitored citation events, 2026

Data: Jan.-Feb. 2026; U.S. monitored prompts / Published: April 14, 2026

aisearch.similarweb.com/blog/most-cited-domains-llms

RESEARCH STUDIES

R12 Cloudflare: crawl purposes

Data: July 2025 / Published: Aug. 29, 2025

blog.cloudflare.com/crawlers-click-ai-bots-training

R13 Vercel/MERJ: AI crawlers, Dec. 2024

Data: Months before Dec. 2024 / Published: Dec. 17, 2024

vercel.com/blog/the-rise-of-the-ai-crawler

R14 SE Ranking: llms.txt analysis

Data: ~300K domains; window not stated / Published: Oct. 15, 2026 (current page)

seranking.com/blog/llms-txt

R15 Kaiser & Schulze: ecommerce referral study

Data: Aug. 2024-July 2025; model Feb.-Aug. 2025 / Published: April 21, 2026

pubsonline.informs.org/doi/full/10.1287/mksc.2025.0489

PROVIDER DOCUMENTATION

R3 Google: Q2 2026 CEO earnings remarks

Data: Q2 2026 company metrics / Published: July 22, 2026

blog.google/company-news/inside-google/message-ceo/alphabet-earnings-q2-2026

R4 OpenAI: DevDay 2026 recap, Sept. 29, 2026

Data: Sept. 29, 2026 stated figure / Published: Sept. 29, 2026

openai.com/index/devday-2026-recap

R5 Google: AI optimization and Search guidance

Data: Not applicable; maintained guidance / Published: Not stated; updated July 10, 2026

developers.google.com/search/docs/fundamentals/ai-optimization-guide

SOURCE DRAWER / 2 OF 4

PROVIDER DOCUMENTATION

O1 OpenAI: crawler purposes and controls

Data: Not applicable; maintained guidance / Published: Not stated; checked Oct. 28, 2026

developers.openai.com/api/docs/bots

O2 OpenAI API: web search sources and citations

Data: Not applicable; maintained guidance / Published: Not stated; checked Oct. 28, 2026

developers.openai.com/api/docs/guides/tools-web-search

O3 Anthropic: crawler purposes and controls

Data: Not applicable; maintained guidance / Published: Not stated; updated April 7, 2026

privacy.claude.com/en/articles/8896518-does-anthropic-crawl-data-from-the-web-and-how-can-site-owners-block-the-crawler

O4 Perplexity: crawler purposes and controls

Data: Not applicable; maintained guidance / Published: Not stated; checked Oct. 28, 2026

docs.perplexity.ai/docs/resources/perplexity-crawlers

O5 Apple: Applebot and Extended control

Data: Not applicable; maintained guidance / Published: Sept. 5, 2026 (page date)

support.apple.com/en-gb/119829

C1 Common Crawl: CCBot and access controls

Data: Not applicable; maintained guidance / Published: Not stated; checked Oct. 28, 2026

commoncrawl.org/ccbot

V1 Google: video search guidance

Data: Not applicable; maintained guidance / Published: Not stated; updated Dec. 18, 2025

developers.google.com/search/docs/appearance/video

MEASUREMENT TOOLS

M1 GA4: AI Assistant default channel

Data: Not applicable; maintained guidance / Published: Not stated; checked Oct. 28, 2026

support.google.com/analytics/answer/9756891?hl=en

MEASUREMENT TOOLS

M2 Search Console: AI performance report

Data: Not applicable; maintained guidance / Published: Not stated; checked Oct. 28, 2026

support.google.com/webmasters/answer/16984139

M3 Bing: AI Performance reporting

Data: Not applicable; product announcement / Published: Feb. 10, 2026

blogs.bing.com/webmaster/2026/2/Introducing-AI-Performance-in-Bing-Webmaster-Tools-Public-Preview

SCHEMA

S1 Schema.org: sameAs and vocabulary

Data: Not applicable; schema vocabulary / Published: Not stated; v30.1 Sept. 16, 2026

schema.org/sameAs

OPTIMIZATION THEORY

T1 Optimization Theory: author and tool links

Data: Not applicable; author and tool pages / Published: Not stated; checked Oct. 28, 2026

optimizationtheory.com

T2 Jourdan Rombough: 90-day GA4 case study

Data: July 25 to Oct. 22, 2026 / Published: Oct. 23, 2026

optimizationtheory.com/project-view/case-studies-ga4-bot-traffic-ai-visitors

RESEARCH STUDIES

N4 Suganthan: 44 days of Markdown requests

Data: March 7 to April 19, 2026 / Published: not stated

suganthan.com/blog/cloudflare-markdown-for-agents

N7 Ahrefs: schema and AI citations

Data: Aug. 2025 to March 2026 / Published: May 11, 2026

ahrefs.com/blog/schema-ai-citations

SOURCE DRAWER / 3 OF 4

RESEARCH STUDIES

N8 Kevin Indig: where ChatGPT citations come from

Data: 1.2 million answers; window not stated / Published: 2026 (day not stated)

searchengineland.com/chatgpt-citations-content-study-469483

N9 Seer Interactive: content recency and AI visibility

Data: March to June 2026 / Published: July 24, 2026

seerinteractive.com/insights/study-content-recencys-impact-on-ai-visibility-in-2026

N10 Fang et al.: recency bias in LLM reranking

Data: TREC DL21/DL22 test sets / Published: Sept. 14, 2025

arxiv.org/abs/2509.11353

N11 Minuttia: self-promotional content study

Data: Aug. 2026; rerun Oct. 2026 / Published: Oct. 27, 2026

minuttia.com/self-promotional-content-data-study

N17 Slobodan Manic: buying Reddit for AI citations

Data: Not applicable; reporting / Published: June 29, 2026

searchenginejournal.com/buying-reddit-to-win-ai-citations-is-the-new-link-farm/578016

AGENT FORMATS AND STANDARDS

N3 Cloudflare: Markdown for Agents

Data: Not applicable; product changelog / Published: Feb. 12, 2026

developers.cloudflare.com/changelog/post/2026-02-12-markdown-for-agents

N5 llms.txt proposal, v2

Data: Not applicable; specification / Published: Sept. 3, 2024; v2 Aug. 10, 2026

llmstxt.org

N6 Chrome: Lighthouse agentic browsing audits

Data: Not applicable; documentation / Published: updated May 5, 2026

developer.chrome.com/docs/lighthouse/agentic-browsing

POLICY AND PLATFORM STATEMENTS

N1 Google: Danny Sullivan on "bite-sized" chunks

Data: Jan. 8, 2026 podcast / Published: Jan. 9, 2026

seroundtable.com/google-content-bite-sized-chunks-40728.html

N2 Google and Bing on Markdown pages for bots

Data: Statements, Feb. 4-5, 2026 / Published: Feb. 4, 2026

searchenginejournal.com/googles-mueller-calls-markdown-for-bots-idea-a-stupid-idea/566598

N12 Google: spam policies for Search

Data: Not applicable; policy / Published: updated Aug. 28, 2026

developers.google.com/search/docs/essentials/spam-policies

N13 Bing: Webmaster Guidelines, GEO and AI abuse

Data: Not applicable; guidelines / Published: Feb. 2026

bing.com/webmasters/help/webmaster-guidelines-30fba23a

N18 Reddit: content policy

Data: Not applicable; policy / Published: not stated

redditinc.com/policies/content-policy

N19 OpenAI: ChatGPT ads

Data: Not applicable; announcement / Published: Aug. 31, 2026

openai.com/index/expanding-access-to-ai-with-chatgpt-ads

N20 Google Merchant Center: UCP-powered checkout

Data: Not applicable; help page / Published: not stated

support.google.com/merchants/answer/16837055

SOURCE DRAWER / 4 OF 4

POLICY AND PLATFORM STATEMENTS

N21 FTC: final rule on fake reviews

Data: Not applicable; rule / Published: Aug. 14, 2024

ftc.gov/news-events/news/press-releases/2024/08/federal-trade-commission-announces-final-rule-banning-fake-reviews-testimonials

N22 Wikipedia: conflict-of-interest guideline

Data: Not applicable; guideline / Published: not stated

en.wikipedia.org/wiki/Wikipedia:Conflict_of_interest

N23 Google Search Central: structured data search gallery

Data: Not applicable; documentation / Published: not stated

developers.google.com/search/docs/appearance/structured-data/search-gallery

N24 Google Search Central: FAQ (FAQPage) structured data

Data: Not applicable; documentation / Published: May 8, 2026

developers.google.com/search/docs/appearance/structured-data/faqpage

N25 Google Search Central Blog: changes to HowTo and FAQ rich results

Data: Not applicable; announcement / Published: Aug. 8, 2023

developers.google.com/search/blog/2023/08/howto-faq-changes

POLICY AND PLATFORM STATEMENTS

N26 Google Search Central: review snippet guidelines

Data: Not applicable; documentation / Published: Sept. 8, 2026

developers.google.com/search/docs/appearance/structured-data/review-snippet

N27 Google Search Central: how to specify a canonical URL

Data: Not applicable; documentation / Published: July 10, 2026 (last updated)

developers.google.com/search/docs/crawling-indexing/consolidate-duplicate-urls

SECURITY RESEARCH

N14 Microsoft Security: AI recommendation poisoning

Data: 60-day review / Published: Feb. 10, 2026

microsoft.com/security/blog/2026/02/10/ai-recommendation-poisoning

N15 SPLX: AI-targeted cloaking

Data: Not stated / Published: Oct. 2025

splx.ai/blog/ai-targeted-cloaking-openai-atlas

N16 Andre Alpar: hidden prompt injection vs. 12 assistants

Data: June 10-11, 2026 / Published: June 22, 2026

searchengineworld.com/hidden-prompt-injection-vs-twelve-ai-assistants-one-obeyed-three-warned-the-user