

AI CITATIONS SERIES · NO. 5 · THE DRIFT REPORT

The Reddit collapse.

Fifteen days ago, Reddit appeared in 87% of Perplexity's answers to our 39 buyer questions — the most inevitable source we had ever measured. We re-ran the **identical questions** through the **identical pipeline**. It's 18%.

87% → 18%

answers citing Reddit · byte-identical 39 questions · July 1 → July 16, 2026 · Perplexity

39 identical questions · 15 days apart · same engines, prompts, gateway, extraction · **source-level figures are Perplexity only**

79% of questions changed their single top-cited source 31 of 39 · Perplexity	25% Perplexity's brand-list overlap with its own answers from 15 days ago ChatGPT 37% · Gemini 38%
59% all-3-engine agreement — structurally unchanged was 56% · all 3 engines	2.0% Reddit's share of all citations now was 10.4%

How controlled is this comparison? The 39 questions are byte-identical (mechanically enforced), same three engines and model ids, same system prompt and locale, same gateway, same counting method as report No. 2. One honest caveat: both runs pass through the same API gateway, so we measure the identical *pipeline's* behavior — we

can't fully apportion the change between the engine and its citation passthrough. Either way, nothing on our side changed.

FINDING 01

The collapse — with its controls

Only seven questions cited Reddit both times. And the two controls say this is Reddit-specific, not a citation drought: answers actually got **richer** (8.4 → 9.1 cited domains per answer), and YouTube barely moved.

SET A — IDENTICAL 39 QUESTIONS	JULY 1	JULY 16
Answers citing Reddit	34 / 39 = 87%	7 / 39 = 18%
Reddit's share of all citations	10.4%	2.0%
Answers citing YouTube (control)	15 / 39 = 38%	14 / 39 = 36%
Cited domains per answer (control)	8.4	9.1

FINDING 02

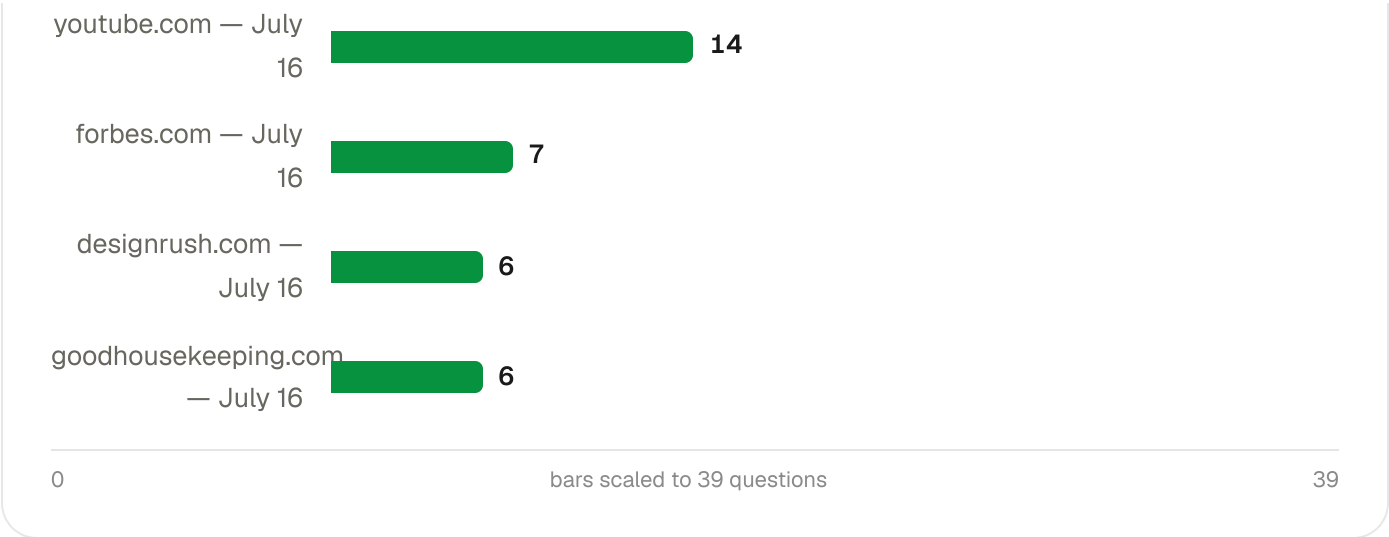
Where the citations went

The hole Reddit left was filled — largely by editorial and listicle publishers. Community corroboration was swapped for edited, publisher-shaped pages.

Most-cited domains on the identical 39 questions

citation counts, Set A · Perplexity only





If that direction holds, the pool got *more* winnable by published pages — which is most of what a brand can actually control — and less dependent on being talked about in threads.

FINDING 03

It isn't just Reddit. Everything churns.

The single top-cited source changed on **31 of 39 questions (79%)** in fifteen days. And at the brand level, each engine barely agrees with *itself* across the fortnight:

ENGINE	OVERLAP WITH ITS OWN JULY 1 BRAND LIST
Perplexity	25%
ChatGPT	37%
Gemini	38%
For scale: two different engines, same day	7.5%

"Your brand ranks on ChatGPT" is a statement with a half-life measured in days.

FINDING 04 · THE COUNTERWEIGHT

The rules are stable. The winners are not.

While the contents shuffled wholesale, the *shape* of the system did not move:

STRUCTURAL METRIC — SET A	JULY 1	JULY 16
All 3 engines shared ≥1 pick	56%	59%
At least 2 engines shared ≥1 pick	90%	87%
Full-list overlap (mean Jaccard)	7.2%	7.5%

Every structural claim this series has published survived the fortnight. Every source-specific fact was perishable. That asymmetry *is* the strategy: **optimize for the structure — be citable, be structured, be in the pool — not for any single incumbent source.**

WHAT REMAINS

Reddit today, across all 250 questions

In the full July 16 scan (250 questions), Reddit appears in **13% of answers** and holds **1.8% of all citations** — 36 thread URLs spread across 30 different subreddits, none appearing more than three times. The subreddit taxonomy we had planned is therefore moot: there is no concentration left to map.

Disclosure. We attempted to fetch the 35 threads' public metadata for age analysis; reddit.com returned HTTP errors for every unauthenticated request. We don't scrape around blocks, so thread ages are not reported — degraded scope over fabricated scope. And a limit worth repeating: **two snapshots make a line, not a law.** This is the first interval of a monthly series; Reddit could be back at 87% next month, and that would be just as publishable.

WHAT TO DO ABOUT IT

Four things this data actually supports

- 01

Date every source-specific stat

"Reddit is 87% of AI answers" was true, measured — and lasted under fifteen days. Ask anyone quoting a citation stat when it was measured, and how often they re-measure.
- 02

Optimize for the structure, not the incumbent

The durable facts are structural: engines disagree, agreement is shallow, the pool is a long tail. Structured, citable pages on your own domain competed in both versions of this landscape; a Reddit-thread strategy competed in exactly one.
- 03

Measure continuously or not at all

Monthly is now the minimum honest cadence. This series re-runs the same 39 questions every month and publishes whatever they say.
- 04

Re-read report No. 2 with today's numbers

Its structural findings held; its Reddit reach figure did not — which is why it carries a dated update banner instead of a quiet edit.

THE DATA

Every number, in one place

METRIC	SCOPE	VALUE
Reddit presence — July 1	Perplexity, Set A n=39	34 / 39 = 87%

METRIC	SCOPE	VALUE
Reddit presence — July 16	Perplexity, Set A n=39	7 / 39 = 18%
Questions citing Reddit both runs	Perplexity, Set A	7
Reddit share of citations — July 1	Perplexity, Set A	10.4%
Reddit share of citations — July 16	Perplexity, Set A	2.0%
YouTube presence, July 1 → 16	Perplexity, Set A	38% → 36%
Cited domains per answer	Perplexity, Set A	8.4 → 9.1
Top-cited domain changed	Perplexity, Set A	31 / 39 = 79%
Self-overlap over 15 days — Perplexity	brand lists, n=39	25%
Self-overlap — ChatGPT / Gemini	brand lists, n=39	37% / 38%
Cross-engine overlap, same day	all 3, July 16	7.5%
All-3 agreement, July 1 → 16	all 3, Set A	56% → 59%
≥2 agreement, July 1 → 16	all 3, Set A	90% → 87%
Reddit presence, all 250 questions	Perplexity, July 16	33 / 250 = 13%
Reddit share of all citations	Perplexity, July 16 n=250	1.8%
Distinct subreddits / thread URLs	Perplexity, July 16	30 / 36
Thread metadata fetches blocked	reddit.com, unauthenticated	35 / 35

Recompute us. Both datasets ship with this report — `multi39.csv` (July 1) and `multi250.csv` (July 16; Set A is the identical re-run) — and `verify-numbers.py` asserts every figure above straight from them. We don't know *why* the collapse happened — licensing, crawl access, retrieval re-ranking are all consistent with the data and none is verifiable from outside — so we report the measurement and decline to speculate.

QUESTIONS THIS REPORT ANSWERS

Quick answers, straight from the data

How fast do AI answers actually change?

In 15 days, on byte-identical questions and an identical pipeline, Reddit fell from 87% of Perplexity's answers to 18%, and the top-cited source changed on 31 of 39 questions (79%).

Did the structure of AI answers change too?

No. All-three-engine agreement went 56% → 59%, at-least-two went 90% → 87%, and full-list overlap went 7.2% → 7.5%. The rules held; the winners churned.

Is the churn just a Perplexity problem?

No engine agrees with its own two-week-old answers: brand-list overlap with July 1 was 25% for Perplexity, 37% for ChatGPT, 38% for Gemini.

Which sources cite you — this week?

Same method, pointed at your domain, re-measurable monthly. We'll ask the engines your buyers' questions and show you what they cite instead of you — with a date on it.

[Run the free audit →](#)

The 2026 AI Citations series · No. 5

By HarperFlow · July 16, 2026 · every figure recomputes from the released dataset

[Download the dataset](#)

[All six reports →](#)