



CiteHawk

Research study No. 1

— THE AI INDEX RESEARCH SERIES · AUGUST 2026

The coverage illusion.

Track one AI model and the number one brand you see is the wrong one in 61.4% of categories. Add a second engine, then a third, and the error rate barely moves. This study measures exactly what a partial tracker gets wrong, using the frozen August 2026 edition of the CiteHawk AI Index: 184 categories, 1,467 archived answers, every number reproducible from the public record.

One corpus, four partial trackers

Most AI visibility tools watch one, two, or three engines. The question this study asks is simple: if the full record is what all eight AI models say, how wrong is the partial view? We took the frozen August 2026 edition corpus, rebuilt each category's all-eight baseline ranking with the Index's published aggregation, then recomputed every ranking as if only a subset of engines had been collected. Nothing else changes: same answers, same scoring, same tie rules.

TRACKER CONFIGURATION	WRONG NO. 1	TOP-10 BRANDS MISSED ENTIRELY	TOP-10 OVERLAP
ChatGPT only	61.4%	30.7%	69.4%
ChatGPT + Perplexity	54.9%	53.3%	46.8%
ChatGPT + Perplexity + Gemini	54.9%	35.0%	65.0%
ChatGPT + Claude + Gemini	51.6%	32.7%	67.4%

August 2026 edition. Wrong No. 1: share of categories where the subset's number one differs from the all-eight number one. Missed entirely: share of all-eight top-10 brands appearing nowhere in the subset's ranking, pooled across categories. Overlap: mean per-category share of the all-eight top 10 that the subset ranks anywhere.

61.4%

of categories crown the wrong No. 1 when only ChatGPT is measured. August 2026 edition.

113 of 184

categories where the ChatGPT-only number one differs from the eight-model record.

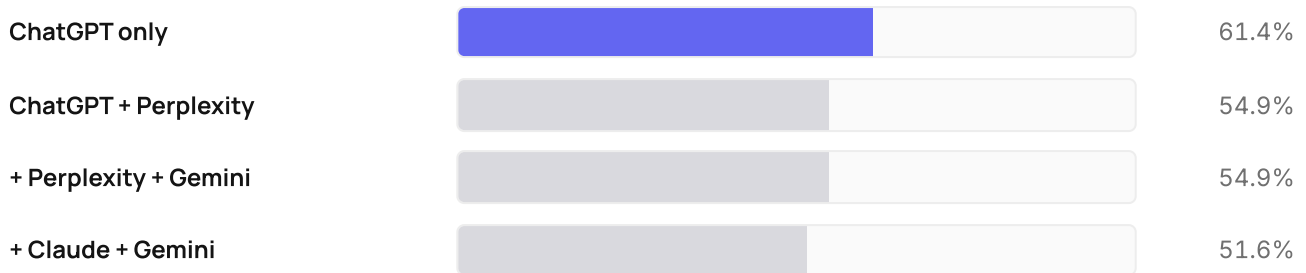
53.3%

average wrong-No.-1 rate for the three-engine mixes. Three engines are not three times safer.

The headline is not that partial data is imperfect. It is that the partial view is confidently, specifically wrong: it names a number one, and in a majority of categories it is a different brand than the one the full record supports.

Three engines are not three times safer

Wrong-No.-1 rate by configuration, August 2026 edition:



Doubling and tripling the engine count buys roughly ten points of protection on the number one call. The solo error rate is also stable across months: 62.2% in the July 2026 edition, 61.4% in August. This is not one bad month. It is the structure of the answers themselves: the engines genuinely disagree, so any subset inherits blind spots no dashboard can display.

The month-to-month record makes a second point. In July 2026 the three-engine mixes called the wrong number one in roughly 38.9% of categories; in August, 53.3%. When the field moves, the movement is not evenly distributed across engines, so whatever safety a third engine buys is unstable. The solo error rate stays above sixty either way.

The two-engine paradox

ChatGPT plus Perplexity misses **53.3%** of consensus top-10 brands entirely, more than ChatGPT alone at **30.7%**. That is not a typo. A single-engine tracker ranks everything it sees. A responsible multi-engine tracker keeps the published two-model consensus gate, so a brand seen by only one of its two engines does not rank at all. The extreme case: in the business credit cards category, the two-engine tracker produced **no rankable number one at all** in August 2026. The full record has one.

Missed-brand rates also swing with the month: the solo tracker missed 61.1% of consensus top-10 brands in July 2026 and 30.7% in August, a month whose answers named noticeably more brands per response. The wrong-No.-1 rate barely moved, which is why this study leads with it: it is the stable failure.

A partial tracker does not show you a smaller version of the truth. It shows you a different truth, on the same confident dashboard.

The receipts: who your tracker crowns instead

In 113 of 184 categories, ChatGPT alone crowns a different number one than the eight-model record. Twelve of those flips, all reproducible from the frozen August 2026 edition:

CATEGORY	ALL-EIGHT NO. 1	CHATGPT-ONLY NO. 1
Accounting software	Xero	QuickBooks Online
AI tools	Claude	ChatGPT
AI image generators	Midjourney	OpenAI
AI video generators	HeyGen	Google Veo / Flow
Air fryers	COSORI	Ninja
Air purifiers	Blueair	Coway
Antivirus software	Bitdefender	Microsoft Defender
Applicant tracking systems	JazzHR	Greenhouse
Blenders	Nutribullet	Vitamix
Business bank accounts	Relay	Chase Business Complete Banking
Business credit cards	American Express	Chase
Business intelligence software	Tableau	Microsoft Power BI

The second row deserves a pause. Ask only ChatGPT who leads in AI tools and the answer is ChatGPT. The eight-model consensus says Claude. A one-engine tracker cannot tell you when the engine is grading itself.

A pattern worth noticing in these examples: the solo view tends to promote the biggest household default over the consensus pick, Microsoft Defender over Bitdefender, Chase over American Express, Power BI over Tableau. A brand competing on merit against a giant has the most to lose from a tracker that only asks one AI model.

What this means in practice: if your tool watches one engine and says you are number one, the full record disagrees in a majority of categories. The reverse cuts too. A brand that looks invisible on one engine may hold the consensus crown, and a partial tracker will never tell it so.

Method, stated fully

Corpus. The archived answers behind the latest global snapshot of each category in the frozen August 2026 edition of the CiteHawk AI Index: 184 published categories, 1,467 answers. Categories excluded from publication are excluded here too. July 2026 figures come from the same analysis published against the July edition.

Baseline. Each category's all-eight ranking is rebuilt from the corpus with the Index's published aggregation: brands keyed by name, the two-model consensus gate, and the published 60/40 blend of share of voice and mention rate with its ceilings. Ties keep score-then-mentions order.

Subsets. Each partial tracker recomputes the same ranking from only its engines' answers, with one honest change: a single-engine tracker keeps everything it sees (a gate of one), while multi-engine subsets keep the two-model consensus gate, matching how each kind of tool behaves. An empty subset ranking counts as a wrong No. 1.

Limits, stated plainly

Brands are keyed by exact name, so alias spellings count separately. Answer order breaks exact score ties for single-engine trackers. The missed-brands metric pools across categories while overlap averages per category; they answer different questions. Every figure is from the August 2026 edition unless labeled July 2026.

Reproduce it: the editions are frozen and public at

citehawk.com/leaderboards/editions, and the full methodology, including scoring and the consensus gate, is published at **citehawk.com/leaderboards/methodology**. Rankings are never for sale.

See what all eight AI models say about your brand.

The coverage illusion ends where the measurement is complete. Your free CiteHawk report asks the eight AI models the Index tracks about your brand and shows every answer with receipts. No credit card.

Start free → citehawk.com/signup