



CiteHawk

Methodology v1.1 · current as of July 2026

— THE CITEHAWK AI INDEX

How we measure AI recommendations.

The AI Index ranks the brands that 8 leading AI assistants actually recommend, across 185 categories, refreshed monthly, receipts kept. This paper is the complete method: the questions, the engines, the full scoring formula, and the rules that keep it honest. Nothing is withheld; any position can be recomputed from the published answers.

Maintained as a living document at citehawk.com/leaderboards/methodology. This edition reflects Methodology v1.1.

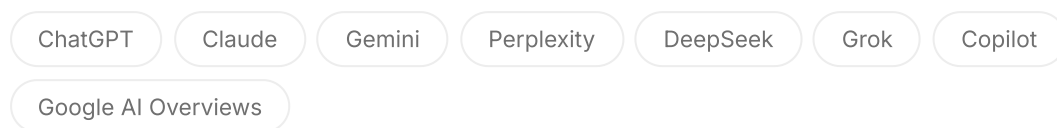
The question we ask

For every category, each assistant is asked the question a real buyer asks: **"What are the best {category}?"**, phrased naturally, with no brand names supplied. An assistant that names a brand in reply is recommending it unprompted.

Each assistant is asked multiple times per refresh, because AI answers vary run to run. A single sample is an anecdote, not a measurement.

The engines

All 8 leading assistants, every refresh. No engine is skipped because it is inconvenient to reach:



Consumer parity

The Index measures what consumers get, so each assistant is queried with the model that matches its consumer default (or the nearest API equivalent), never a cheaper stand-in. Copilot and Google AI Overviews are captured from the consumer web surface directly.

Model transparency

The exact model version behind every captured answer is recorded with it. Model versions change only at refresh boundaries, never mid-month, and every change is listed in the public changelog. When an assistant ships a new model, the Index shows the shift rather than hiding it: a ranking move caused by a model upgrade is attributable, not mysterious.

The score: the full formula

Brands are ranked by their **AI Recommendation Score** (0 to 100). The complete calculation is published so any position can be recomputed from the answers. There is no editorial weighting and no hidden step.

1 • Extraction. Every brand recommendation is extracted from every captured answer, with the position it appeared at. The answers themselves are published: the receipts on every category page.

2 • Consensus gate. A brand must be named by at least two different assistants to be rankable at all. One engine's quirk is not a market position.

3 • Share of voice = the brand's mentions ÷ all mentions of rankable brands in the category.

4 • Mention rate = the share of all captured answers that name the brand.

$$\text{Score} = 60 \times \min(\text{share of voice} \div 40\%, 1) + 40 \times \min(\text{mention rate} \div 80\%, 1)$$

The ceilings mean a brand maxes the share-of-voice component at 40% of all mentions and the mention-rate component at 80% of answers. Dominance beyond that stops adding, which keeps crowded and sparse categories comparable. Brands are ranked by score; ties break toward more total mentions.

Worked example

A brand holding 30% share of voice, named in half of all answers:

$$60 \times (30 \div 40) + 40 \times (50 \div 80) = 45 + 25 = \mathbf{70}$$

Why there is no citation component

Earlier versions added points when AI cited a brand's own site as a source. Our July 2026 audit showed that rewards the wrong thing: review sites and forums that answers cite as *sources* outranked products that six of eight assistants actually *recommended*. Citations measure "is this a source", not "is this recommended", so they were removed from the score. Mistakes in a methodology are fixable; hiding them is not. Every change is logged publicly.

The rules that keep it honest

The gates

Consensus: at least two different assistants must recommend a brand before it can be ranked. **Receipts:** every ranking keeps the verbatim answers that produced it, stored and shown, so any position can be checked against its evidence. **Immutability:** each monthly refresh is captured as a frozen snapshot; movement compares against the previous snapshot, and past editions are never edited.

Regional rankings

Some categories (banks, insurers, agencies, professional services) get genuinely different answers in different countries. Those categories are also asked as a local buyer would ask, for the United States, the United Kingdom, Australia and Canada, using the local term where it differs.

A regional edition is published only when its ranking meaningfully differs from the global one: a different #1, or fewer than three-quarters of the top 10 in common. When a market simply agrees with the global answer, the global page stays the single canonical ranking. Regional editions refresh monthly, offset two weeks from the global refresh.

The refresh rhythm

The Index refreshes monthly across all assistants and categories, and every page shows its refresh date. Between refreshes the ranking does not move: a stable, comparable measurement beats a jittery live one. The same philosophy runs through CiteHawk's brand monitoring, which collects weekly; answers shift too slowly for daily checks to mean anything and too fast for a quarterly review to catch.

"Check it yourself: every monthly record's JSON download contains every verbatim answer and the extracted brand positions behind it."

Anyone can recompute any score from the primary source. That is the difference between a methodology and a marketing claim.

The pledge

Rankings are computed from AI responses only. Claiming a brand cannot change its position, and positions are not for sale.

Brand owners can claim their brand to verify identity details (logo, domain, alternate names) and follow their movement. Identity corrections are human-reviewed and never affect scores. No payment, partnership, or relationship with CiteHawk influences any ranking. Badges render live from current data and degrade automatically when a position lapses, so an embedded badge can never overstate.

Using this method

Researchers, journalists and analysts are welcome to cite the Index and its editions; each monthly edition is an immutable record designed to be referenced. If you republish figures, cite the edition month and link the record so readers can verify against the stored answers.

See the method running.

Browse any of the 185 category rankings and open the receipts behind any position. Or run the same 8-assistant measurement on your own brand: the first 50-prompt report is free, no card, ready within the hour.

citehawk.com/leaderboards

Methodology v1.1, current as of July 2026. The canonical, always-current version lives at citehawk.com/leaderboards/methodology. © 2026 RP & SE Technology Pty Ltd.

Reproduction with attribution welcome.