

Where AI engines get their buying recommendations

By Derrick Mwitj, AppearsIn · 2026-10-08 · [Download the dataset \(CSV\)](#).

The numbers

1200

answers collected

150

buying questions

8

AI engines

928

answers that cited
sources

What we found

The most cited site was forbes.com, which appeared in 107 answers, followed by techradar.com (97). Of the 928 answers that cited any source, 10.1% cited Reddit and 9.1% cited YouTube.

That is not a general pattern across engines. It is concentrated in Google's own: Gemini cited Reddit in 52.4% of its answers with sources and Google AI Overview cited Reddit in 44.3% of its answers with sources, and between them they account for 89% of all the answers that cited Reddit and 74% of those that cited YouTube. The other engines cited Reddit far less: Grok 5.3%, Perplexity 1.3%, ChatGPT 0%, Claude Sonnet 5.5 0%, Claude Opus 5.5 0%, Meta Muse Spark 0%.

A zero deserves care, so we also asked every engine two questions that explicitly ask what Reddit users say (outside the 1,200 answers). ChatGPT, Google AI Overview, Gemini, Perplexity and Grok returned Reddit links, so the method can see Reddit citations when an engine gives them. Claude Sonnet 5.5, Claude Opus 5.5 and Meta Muse Spark returned none, although their answers discuss Reddit and listed other sources (Claude Sonnet 5.5 3 and 4, Claude Opus 5.5 2 and 4 and Meta Muse Spark 2 and 3 sources). For those engines, a 0% means no Reddit link appeared among the sources their interface returned, not that Reddit never shaped the answer.

The pattern is consistent with, but does not prove, how Reddit has dealt with AI companies: it has data-licensing agreements with Google and OpenAI, and it sued Anthropic in June 2025 and Perplexity in October 2025. We did not test whether those relationships explain the differences.

The mix of sources changes with what is being bought. For business software, 22.6% of answers with sources cited a review or comparison site such as G2, against 1.6% for everyday products. For everyday products, 51.9% cited a tech or business publication (against 30.3%) and 14.3% cited YouTube (against 6.4%). Reddit was cited in 10.9% of business-software answers and 8.6% of consumer answers.

Almost every answer that cited sources (92.8%) cited at least one blog, list or other site that is not a major publication or review platform, the kind of page that ranks for "best X for Y". 37.6% cited a tech or business publication, and 61.3% cited a site belonging to a brand the answer names.

Engines also differ in whether they cite at all. Gemini included sources in only 28% of its answers, while Grok and Perplexity did so in every one.

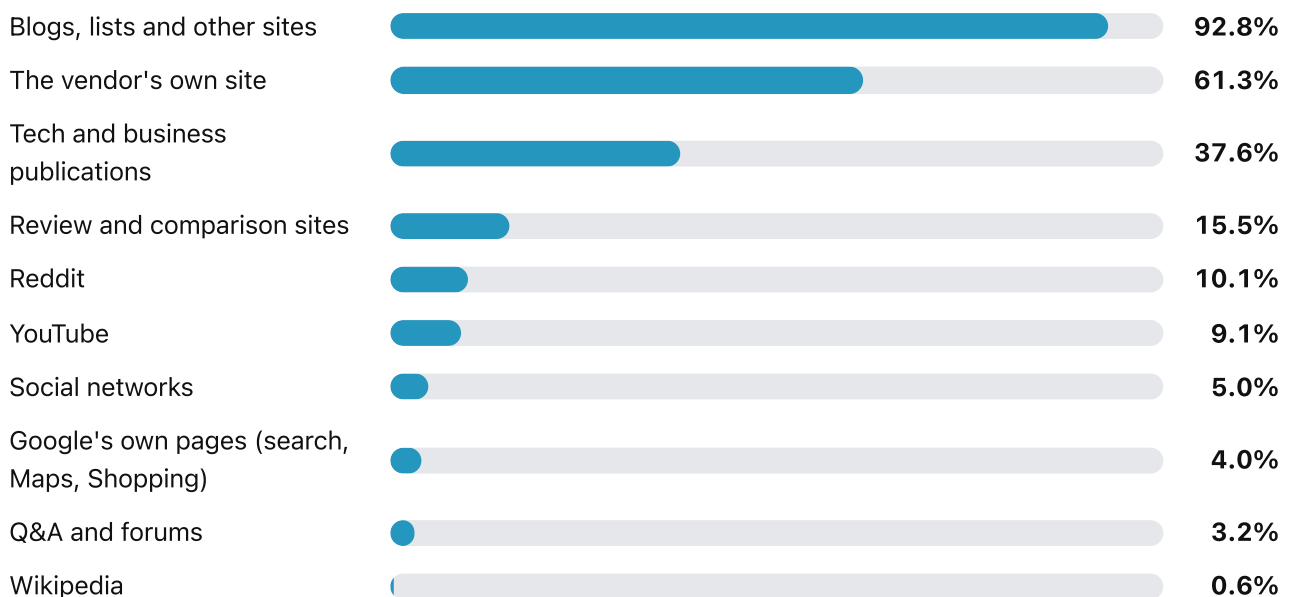
The engines disagree about who to recommend. On 60.8% of questions a majority of them named the same brand first, and the overlap between their top three brands averaged 0.441 (1 would be identical lists).

What it means if you sell something

- Reddit and YouTube matter most for Google's engines. If your buyers use Gemini or Google's AI Overview, being present in the threads and videos that answer your buyers' questions is worth the effort. For the others, it showed up rarely in this sample.
- Being on the ordinary "best X" lists and in the publications that write them matters for every engine, because that is where most citations went.
- Do not measure one engine and assume it represents the others. Their sources and their picks differ, so track each engine your customers use.
- Your own site gets cited when an answer names you, so give it clear, specific pages that state what you do and what it costs.

Which kinds of source get cited

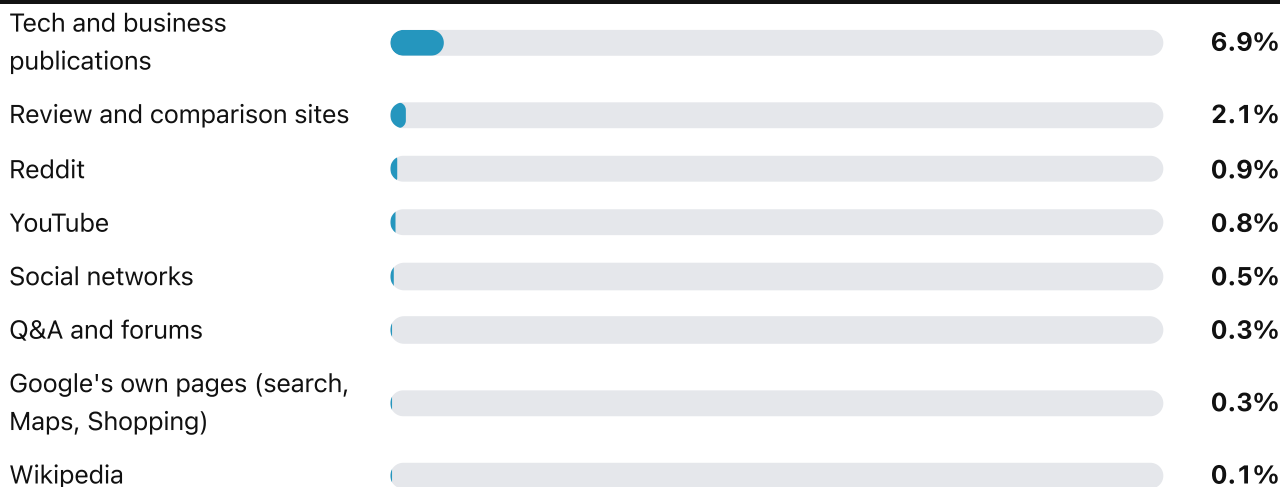
Of the answers that cited any source, this is the share that cited at least one site of each type. An answer can cite several types, so the bars add up to more than 100%.



Share of all citations

Counting every cited site once per answer, this is where the citations went.





Business software versus consumer products

We asked 100 questions about business software and 50 about everyday products. This is the share of answers with sources that cited at least one site of each type, in each group.

Source type	Business software	Consumer products
Blogs, lists and other sites	91.4%	95.5%
The vendor's own site	66.9%	50.3%
Tech and business publications	30.3%	51.9%
Review and comparison sites	22.6%	1.6%
YouTube	6.4%	14.3%
Reddit	10.9%	8.6%
Google's own pages (search, Maps, Shopping)	2.0%	8.0%
Social networks	4.7%	5.4%
Q&A and forums	2.9%	3.8%
Wikipedia	1.0%	0%

Reddit's share of the answers with sources, by engine:

Engine	Business software	Consumer products
ChatGPT	0.0%	0.0%
Claude Sonnet 5.5	0.0%	0.0%
Claude Opus 5.5	0.0%	0.0%
Gemini	47.8%	57.9%

Engine	Business software	Consumer products
Google AI Overview	49.0%	34.1%
Grok	8.0%	0.0%
Meta Muse Spark	0.0%	0.0%
Perplexity	1.0%	2.0%

Each engine behaves differently

Engines differ in how often they cite anything and in what they cite. Percentages are the share of that engine's answers with sources that cited each type.

Engine	Model	Answers	Cited sources	Reddit	YouTube	Reviews	Vendor	Lists and blogs
ChatGPT	gpt-6.1-sol	150	98.0%	0.0%	0.0%	0.0%	80.3%	68.0%
Claude Sonnet 5.5	claude-sonnet-5-5	150	67.3%	0.0%	0.0%	7.9%	73.3%	98.0%
Claude Opus 5.5	claude-opus-5-5	150	54.0%	0.0%	0.0%	8.6%	63.0%	97.5%
Gemini	gemini-3.8-flash	150	28.0%	52.4%	40.5%	2.4%	38.1%	97.6%
Google AI Overview	google_ai_overview	150	93.3%	44.3%	32.1%	6.4%	67.9%	96.4%
Grok	grok-4.7	150	100.0%	5.3%	14.7%	49.3%	82.7%	100.0%
Meta Muse Spark	meta/muse-spark-1.3	150	78.0%	0.0%	0.0%	3.4%	30.8%	91.5%
Perplexity	sonar-pro	150	100.0%	1.3%	0.0%	27.3%	36.7%	100.0%

The most cited sites

Site	Answers citing it	Type
forbes.com	107	publication
techradar.com	97	publication
toolradar.com	96	other
tomsguide.com	95	publication

Site	Answers citing it	Type
reddit.com	94	reddit
rtings.com	94	other
zapier.com	89	other
itechguides.com	85	other
youtube.com	84	youtube
pcmag.com	80	publication
capterra.com	75	review
g2.com	66	review
zoho.com	53	other
dupple.com	53	other
zdnet.com	52	publication

How much the engines agree

Across the 143 questions where we could compare, the engines shared the same top pick on 60.8% of questions, and the overlap between their top three picks averaged 0.441 (1 means identical lists, 0 means nothing in common).

How we did it

- We wrote 150 buying questions in plain buyer language: 10 questions for each of 15 categories. Ten are business software (project management, CRM, email marketing, invoicing and accounting, customer support, website builders, appointment scheduling, webinars, password managers and applicant tracking) and five are everyday products (VPNs, mattresses, wireless headphones, laptops and running shoes). Every category has the same 10 question types, such as best for a beginner, cheapest, cheap versus premium, how to choose, alternatives, and downsides. No question names a product from the category it asks about. The ten integration questions name the platforms a tool must connect to, such as Slack or Stripe, and never a tool in the same category.
- We asked every question to seven AI engines through their programming interfaces, with web search on, and stored the full answer and every cited address. ChatGPT, Gemini, Claude, Perplexity and Google AI Overview were asked through DataForSEO, Grok through xAI, and Meta Muse Spark through OpenRouter. Claude was asked through two models, Sonnet 5.5 and Opus 5.5, so the table has eight rows. The models used are listed in the table below.
- We sorted each cited address by its site into a source type using fixed lists (Reddit, YouTube, review and comparison sites, Q&A forums, social networks, Wikipedia, tech and business

publications, Google's own pages). A site counts as a brand's own when the brand's name appears in the answer text, written like a name, outside any link. Everything else is a blog, list or other site. Gemini cites through redirect links, so we followed each one to the page it points at before sorting it.

- A small language model read each answer and listed the products it recommends, in order. We used that list to measure how often the engines agree.

The full list of questions and every answer's sources are in the [dataset](#).

What this study cannot tell you

- A 0% for an engine means no link of that type was among the sources its interface returned. It does not prove the type never shaped the answer, and for Claude and Meta Muse Spark we found that even a direct request for Reddit opinions returned no Reddit link.
- Some engines were asked through a smaller or cheaper model than the one in their consumer app, as the models in the table show, so a chat app may behave differently.
- Answers fetched through programming interfaces are not identical to what a person sees in a chat app, which can personalise answers and use different settings.
- The questions are in English, as asked from the United States, about business software. The pattern may be different for consumer products, other languages or other countries.
- It is a snapshot taken on the dates shown. AI answers change from run to run and week to week.
- The source-type lists are ours and the vendor test is a rule of thumb, so some sites will be sorted wrongly. The list of recommended products was read by a model and can contain mistakes. We read a sample of rows by hand to check both, and corrected two flaws that way before publishing, but we did not check every row.
- Disclosure: AppearsIn sells software that measures AI visibility, and we ran this study with it. We have no commercial relationship with any brand named in the data.

Further reading

- [Built In: Why Reddit is suing Perplexity and other data scrapers](#)

Questions

▼ Can I use the data?

Yes. The dataset is free to use under the Creative Commons Attribution 4.0 licence. Please credit AppearsIn and link to this page.

▼ Why not name the best brands in each category?

The study is about where recommendations come from, not who wins. The dataset lists the products each answer recommended if you want them.

▼ **Does this mean Reddit does not matter?**

No. It describes software questions in this sample on these dates. Reddit can matter a lot in other categories, and for the engines and questions where it is cited. Read the table by engine.