BROADCASTWELL RESEARCH · STUDY 3 · SEPTEMBER 2026

Run-to-run noise: what one AI answer tells you and what three do

Three scheduled runs of 140 buyer questions on five AI engines, and how often a single run would have told you something the other two did not.

22.2%

of single-run vendor namings were not repeated by both of the other two runs of the same question on the same engine.

Base: 8,055 namings in 14,192 question × engine × vendor cells with three valid runs. Strict variant: 9.1% of namings were contradicted by both other runs. Cell variant: 37.6% of cells that named a vendor at all did not do so in all three runs.

Author Broadcastwell Research · **Version** 1.0 · **Published** 26 Sep 2026 · **License** CC BY 4.0 for this report and its dataset

Data The Absence Index, release 2026-09. DOI 10.5281/zenodo.22907695. Answers captured 22 to 23 Sep 2026 (UTC). Engines: ChatGPT, Claude, Perplexity, Google AI Overviews, Google AI Mode. The source release is CC BY 4.0 with attribution to Broadcastwell.

This study Figures recomputed from the raw answers by a Python implementation of the published matching rule, shipped with the dataset, and cross-checked against the Index site's own matching code: 0 of 14,300 cells differ. Recompute (scripts/recompute_index.py with vendor_rules_2026-09.json) and analysis (scripts/study3_run_noise.py) scripts are kept with the dataset.

Contact sairam@broadcastwell.com

Summary

1. 22.2% of single-run vendor namings were not repeated by both of the other two runs of the same question on the same engine (1,785 of 8,055 namings, across 14,192 cells with three valid runs).

2. 9.1% of single-run namings were contradicted by both other runs: the vendor was named in one run and in neither of the other two (733 of 8,055 namings).

3. The three runs agreed on 91.1% of all cells (12,933 of 14,192), but on only 62.4% of the cells where the vendor was named at all (2,090 of 3,349 active cells).

4. Google AI Overviews and Google AI Mode disagreed with themselves on 50.4% and 48.9% of their active cells; ChatGPT on 27.4%, Claude on 31.2% and Perplexity on 30.6% (bases 599, 705, 580, 812 and 653 active cells).

5. A single run called 57 to 59 of the 286 checked vendors never named across all 140 questions and five engines; three runs called 46 never named (base: 286 vendors).

6. Per engine, one run called on average 16 to 31 more vendors absent than three runs did (base: 286 vendors per engine).

7. 28 vendors (9.8% of 286) were absent in at least one run and present in at least one other; ranked from a single run, a vendor's rank within its category differed from its three-run rank 49.1% of the time (421 of 858 vendor-runs, competition ranking; with single-run ties broken in the three-run order the floor is 28.3%), and the top vendor changed in 3 of 42 single-run category rankings.

8. The mean Wilson 95% interval around a vendor's mention rate was 16.9 points wide with one run (about 50 answers per vendor) and 9.45 points with three (about 150 answers; median 9.8 points; base 286 vendors).

The headline number, explained

A **cell** is one question on one engine for one checked vendor. Each cell has three runs. A **single-run naming** is one run of a cell in which the engine named that vendor under the release's matching rule. There are 8,055 such namings in the 14,192 cells with three valid runs.

For each naming we asked: did both of the other two runs of the same cell also name the vendor? **6,270** namings (77.8%) were named again by both. **1,052** (13.1%) were named again by one of the two. **733** (9.1%) were named by neither.

The headline is the share not repeated by both: (1,052 + 733) / 8,055 = **22.2%**. Read it as: if you look at one AI answer and see a vendor in it, about one time in five the same engine would have left that vendor out of at least one of two more runs of the same question on the same day.

Two companion figures bound it. The strict figure counts only namings contradicted by both other runs: 9.1%. The cell figure asks a different question, how many cells with any naming at all failed to agree across three runs: 37.6% (1,259 of 3,349 active cells).

Why it matters

A single screenshot is a sample of one. A buyer, a vendor or a reporter who runs a question once and reads the answer as the engine's view is reading one draw. In this data, 22.2% of the vendor namings in a single run were not repeated by both of the other two runs of the same question on the same engine, inside the same capture window. The answer was not wrong; it was one of several answers the engine gives. Any claim of the form "engine X recommends vendor Y" or "engine X does not mention vendor Y" should say how many runs it rests on.

Absence needs repeats more than presence does. A vendor that is named in one run is usually named again: 77.8% of namings were repeated by both other runs. A vendor that is missing from one run is a weaker signal. One run called 57 to 59 of 286 vendors never named across all 140 questions and five engines; three runs brought that to 46. Per engine, one run called on average 16 to 31 more vendors absent than three runs did. A vendor that checks its own standing with one query and finds nothing has learned less than it thinks.

Google's two surfaces vary most in this data. Google AI Overviews and Google AI Mode disagreed with themselves on 50.4% and 48.9% of active cells, against 27.4% to 31.2% for ChatGPT, Claude and Perplexity. Any two runs of the same question agreed on 65.4% to 71.1% of active cells on the Google surfaces and on 78.1% to 82.6% on the other three. This study does not say why. It says only that a one-run reading of a Google AI answer needs more caution than a one-run reading of the other three, on this question set and on these dates.

Data and method

Source dataset

The Absence Index, release 2026-09, published by Broadcastwell on 23 Sep 2026. Version DOI 10.5281/zenodo.22907695 (concept DOI 10.5281/zenodo.22150321). License CC BY 4.0 with attribution to Broadcastwell. The release covers 14 categories of business software, 286 checked vendors drawn from buyer directories, and 140 target questions (10 per category) in five question shapes: best option, specific workflow, top tools, alternative, buyer evaluation. Five engines were asked every question: ChatGPT, Claude, Perplexity, Google AI Overviews, Google AI Mode. Each engine and question pair was run three times as scheduled repeats.

The release reports 2,250 observations: 2,100 target answers and 150 control answers. Controls were run only for CRM and construction project management and are not used here. Five target answers were excluded by the release with a stated reason, leaving 2,095 valid target answers. Answers were captured from 2026-09-22T19:36:03Z through 2026-09-23T00:17:54Z, that is 22 to 23 Sep 2026 (UTC). Broadcastwell's published method (broadcastwell.com/methodology, version 1.1, 8 September 2026) states: "The measurement is single turn, API based, logged out, US English, gl=us and hl=en" and "It is a controlled instrument, not a replica of any one buyer's screen."

Matching rule, as published

Names and listed aliases match case-insensitive word boundaries. Plain-word vendors require a listed qualifier within five words in the same sentence, or their canonical domain in that answer's cited URLs. Rates are percentages with Wilson 95% intervals rounded to one decimal. A zero count is distinct from an excluded observation. This study applies the same rule, unchanged, to every answer.

Recompute

Every figure in this report was recomputed from the raw answers (answers-2026-09.jsonl) by a Python implementation of the published matching rule (method v1.1), shipped with the dataset as scripts/recompute_index.py with vendor_rules_2026-09.json, and cross-checked against the Index site's own matching code: 0 of 14,300 published question × engine × vendor cells differ and no vendor count differs. scripts/study3_run_noise.py, also shipped, computed every number here from the per-answer matrix the recompute writes.

What this study measures

The unit is a **cell**: one question × one engine × one checked vendor. Each cell has up to three runs, and each run either named the vendor (1) or did not (0). 14,300 cells exist in the release. 108 of them lost one run to an excluded answer, all in Google AI Overviews, and are left out of this study because a two-run cell cannot be compared with a three-run cell on the same terms. The remaining **14,192** cells have three valid runs each. Definitions used throughout:

- **Unanimous cell**: named in all three runs, or in none. **Flip cell**: named in one or two runs.
- **Active cell**: named in at least one run. Flip rates are reported on active cells, because a cell where no run named the vendor cannot disagree.
- **Single-run naming**: one run of one cell in which the vendor was named. Each naming is classed by whether both, one or neither of the other two runs of the same cell also named the vendor.
- **Pairwise agreement**: for two runs of the same engine, the share of active cells where both runs gave the same 0 or 1.
- **Vendor mention rate**: for one vendor in one category, the share of valid target answers in which the vendor was named. One run gives a base of about 50 answers (10 questions × 5 engines); three runs give about 150. Wilson 95% intervals as in the release.
- **Never named**: a vendor with zero namings across the base in question (one run, or three runs).
- **Rank**: a vendor's rank within its category when vendors are ordered by namings, from one run or from three. The headline figure uses competition ranking, where tied vendors share a rank (1, 2, 2, 4); the floor figure breaks single-run ties in the three-run order.

Repeats are scheduled runs of the same question on the same engine on the same day. This study therefore measures **within-day repeat variation**, not change over time, and asserts no change between releases.

Finding 1. Three runs agree on most cells, but not on the cells that matter

91.1%
of all 14,192 cells were unanimous (named in all three runs, or in none)

62.4%
of the 3,349 active cells were unanimous; 37.6% flipped

8.9%
of all cells flipped (1,259 cells named in one or two runs, not three)

CELL OUTCOME ACROSS THREE RUNS	CELLS	SHARE OF ALL CELLS	SHARE OF ACTIVE CELLS
Named in no run	10,843	76.4%	not active
Named in exactly one run	733	5.2%	21.9%
Named in exactly two runs	526	3.7%	15.7%
Named in all three runs	2,090	14.7%	62.4%
All cells with three valid runs	14,192	100%	3,349 active

Base: 14,192 question × engine × vendor cells with three valid runs; 3,349 of them active. Cells: 10,843 + 733 + 526 + 2,090 = 14,192.

Most cells are quiet: in 76.4% of cells no run named the vendor, and those cells cannot disagree. That is why the all-cell unanimity of 91.1% flatters the engines. The cells a reader cares about are the active ones, where the vendor showed up at least once. Of those, 62.4% were named in all three runs, 15.7% in two and 21.9% in only one.

Put the other way: 733 cells contain a vendor that appeared once and then vanished in both other runs, and 526 cells contain a vendor that appeared twice and was missing once. Together those 1,259 flip cells are 37.6% of active cells. Each flip cell also contains exactly one single-run result that both other runs contradict, so 1,259 of the 42,576 single-run results in the data (2.96%), or 12.5% of the 10,047 results in active cells, are the odd one out.

Finding 2. One naming in five is not repeated by both other runs

Was a single run's vendor naming repeated by the other two runs?

Share of single-run namings, per engine. Base: 8,055 namings in 14,192 cells with three valid runs, 22 to 23 Sep 2026 (UTC).

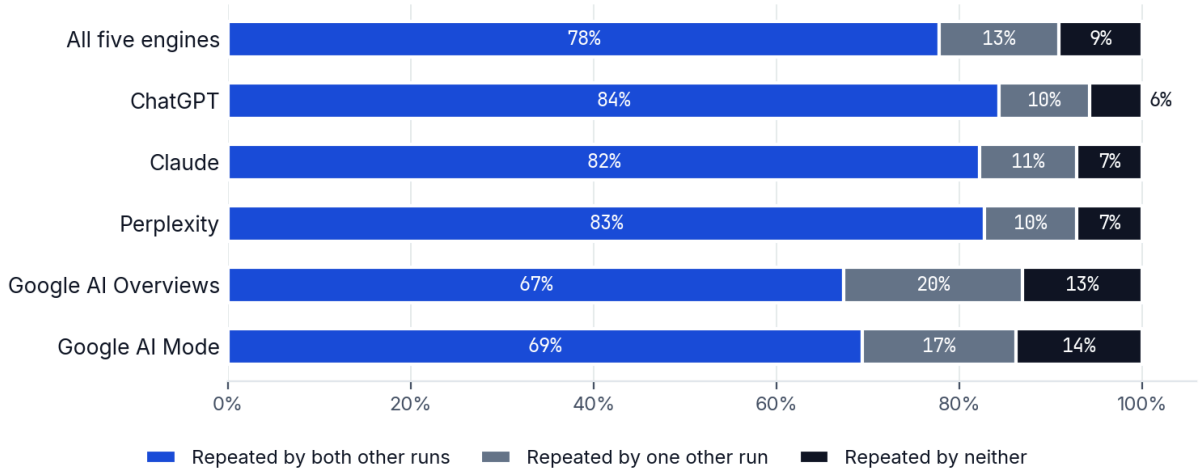


Chart 1. Each single-run vendor naming, classed by whether both, one or neither of the other two runs of the same cell also named the vendor. Base: 8,055 namings.

ENGINE	NAMINGS	BY BOTH	%	BY ONE	%	BY NEITHER	%	NOT BY BOTH
All five engines	8,055	6,270	77.8%	1,052	13.1%	733	9.1%	22.2%
ChatGPT	1,496	1,263	84.4%	148	9.9%	85	5.7%	15.6%
Claude	2,039	1,677	82.2%	218	10.7%	144	7.1%	17.8%
Perplexity	1,642	1,359	82.8%	166	10.1%	117	7.1%	17.2%
Google AI Overviews	1,322	891	67.4%	258	19.5%	173	13.1%	32.6%
Google AI Mode	1,556	1,080	69.4%	262	16.8%	214	13.8%	30.6%

Base: single-run namings per engine (column 2). "Not by both" is the headline measure: the share of namings that at least one of the other two runs did not repeat.

On ChatGPT, 15.6% of namings were not repeated by both other runs; on Claude 17.8%; on Perplexity 17.2%. On Google AI Overviews the figure was 32.6% and on Google AI Mode 30.6%: about one naming in three. The share named by neither other run, the strict reading, ran from 5.7% on ChatGPT to 13.8% on Google AI Mode.

The engines also differ in how much they name at all. Claude produced 2,039 namings across its 2,860 cells, ChatGPT 1,496, Perplexity 1,642, Google AI Mode 1,556 and Google AI Overviews 1,322 across 2,752 cells. A higher naming count does not by itself make an engine more or less stable:

Google AI Overviews named the fewest vendors and repeated them least, while ChatGPT named fewer than Claude or Perplexity and repeated them most.

Finding 3. The engine picture: the two Google surfaces flip about half their active cells

Where the three runs disagreed, by engine

Share of active cells with a disagreement. Bases: 580, 812, 653, 599, 705 active cells, in engine order.

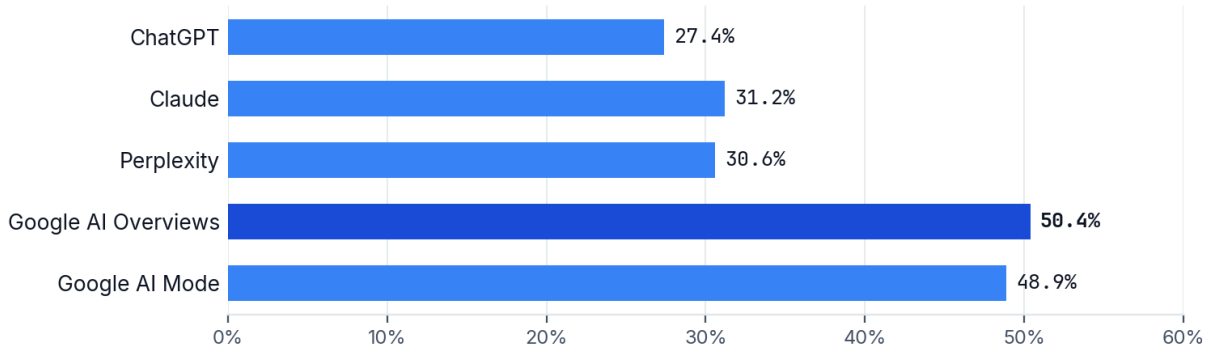


Figure. Share of active cells where the three runs did not agree, per engine. Bases in the table below.

ENGINE	CELLS	ACTIVE CELLS	FLIP CELLS	FLIP % OF ACTIVE	FLIP % OF ALL	UNANIMOUS %	ODD ONE OUT %
ChatGPT	2,860	580	159	27.4%	5.6%	94.4%	9.1%
Claude	2,860	812	253	31.2%	8.8%	91.2%	10.4%
Perplexity	2,860	653	200	30.6%	7.0%	93.0%	10.2%
Google AI Overviews	2,752	599	302	50.4%	11.0%	89.0%	16.8%
Google AI Mode	2,860	705	345	48.9%	12.1%	87.9%	16.3%
All five engines	14,192	3,349	1,259	37.6%	8.9%	91.1%	12.5%

Base: cells with three valid runs per engine (column 2) and active cells (column 3). "Odd-one-out" is the share of single-run results in active cells that both other runs contradict.

ChatGPT was the steadiest engine on this question set: 27.4% of its 580 active cells flipped. Perplexity (30.6%) and Claude (31.2%) sit close behind. Google AI Overviews flipped 50.4% of 599 active cells and Google AI Mode 48.9% of 705. On those two surfaces a vendor that appeared at all had roughly even odds of appearing in every run.

Pairwise agreement between runs

ENGINE	RUN 1 VS RUN 2	RUN 1 VS RUN 3	RUN 2 VS RUN 3	ACTIVE CELLS
ChatGPT	80.3%	82.2%	82.6%	580
Claude	78.1%	78.7%	80.9%	812
Perplexity	80.6%	79.8%	78.4%	653
Google AI Overviews	65.8%	66.8%	66.6%	599
Google AI Mode	65.7%	65.4%	71.1%	705

Share of active cells where the two runs gave the same result (both named, or both did not). Base: active cells per engine.

No single run stands out: on every engine the three pairwise agreements sit within a few points of each other, from 80.3% to 82.6% on ChatGPT and from 65.4% to 71.1% on Google AI Mode.

Finding 4. By question shape: workflow and best-option questions flip most

Where the three runs disagreed, by question shape

Share of active cells (at least one run named the vendor) with a disagreement. Base: 3,349 active cells.

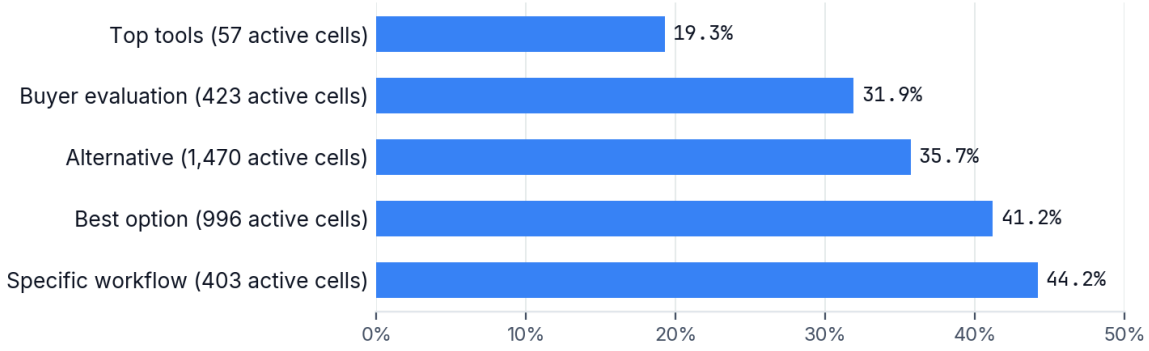


Figure. Share of active cells where the three runs did not agree, by question shape. Bases in the table below.

QUESTION SHAPE	CELLS	ACTIVE CELLS	FLIP % OF ACTIVE	UNANIMOUS %	NAMINGS	NOT BY BOTH	BY NEITHER
Alternative	4,090	1,470	35.7%	87.2%	3,591	21.1%	8.2%
Best option	4,151	996	41.2%	90.1%	2,350	25.2%	9.7%
Buyer evaluation	4,246	423	31.9%	96.8%	1,040	16.9%	9.0%
Specific workflow	1,605	403	44.2%	88.9%	919	26.6%	12.2%
Top tools	100	57	19.3%	89.0%	155	11.0%	3.2%

Base: cells and active cells per question shape (columns 2 and 3); namings per shape (column 6) for the last two columns.

"Specific workflow" questions ("which tool handles X for a business like mine") flipped 44.2% of their 403 active cells, and "best option" questions 41.2% of 996. "Alternative" questions ("alternatives to vendor X") were steadier at 35.7%, and "buyer evaluation" questions at 31.9%. The "top tools" shape flipped least (19.3%), but it rests on only 100 cells from two questions and should be read as a hint, not a result.

On the naming side the order is the same. 26.6% of namings in workflow questions and 25.2% in best-option questions were not repeated by both other runs, against 21.1% for alternatives and 16.9% for buyer-evaluation questions. On this question set, the open-ended shapes flipped more often than the shapes anchored on a named vendor.

Finding 5. By category: every category flips, from about a quarter to about half of active cells

Where the three runs disagreed, by category

Share of active cells with a disagreement. Base: 3,349 active cells across 14 categories.

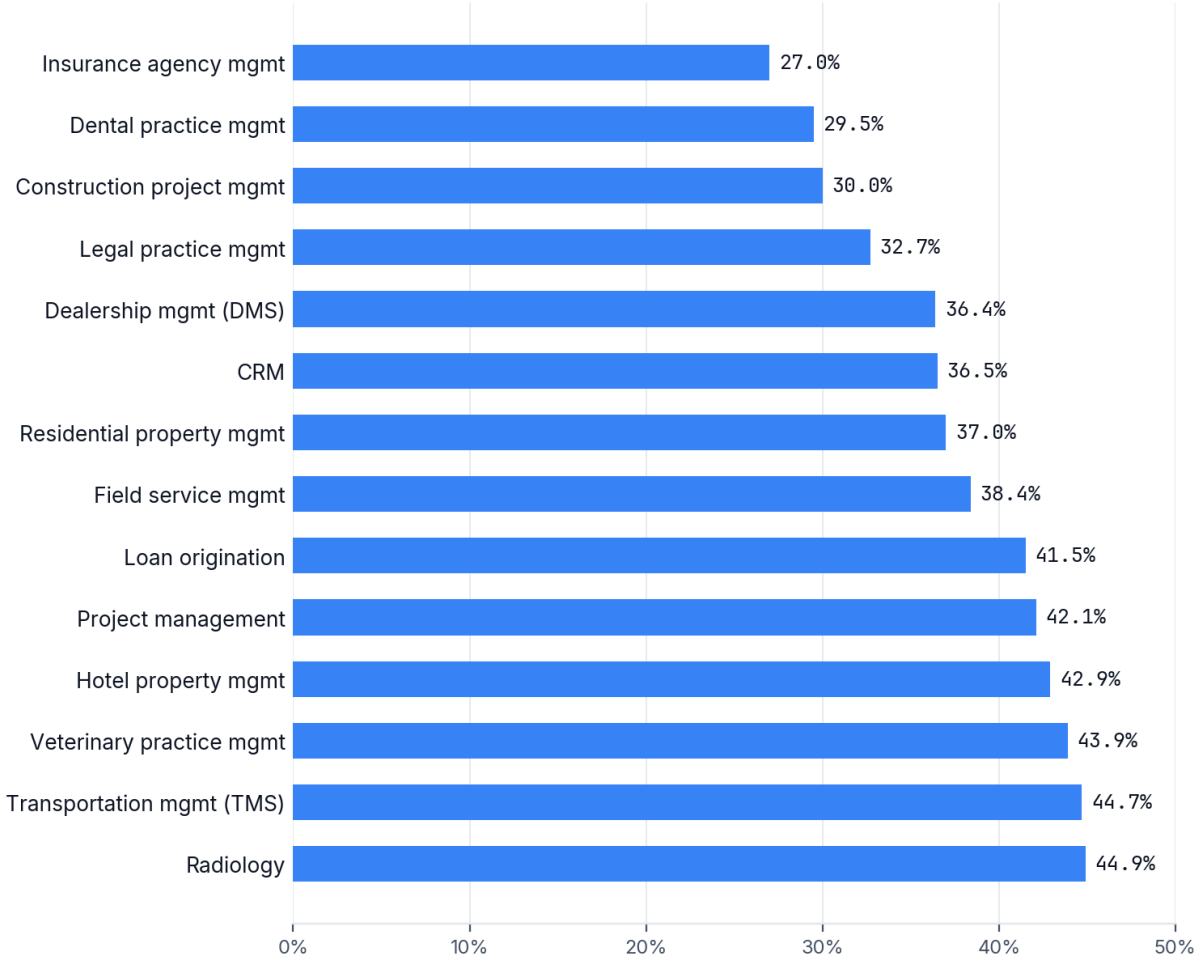


Figure. Share of active cells where the three runs did not agree, by category, sorted. Bases in the table below.

Radiology and medical imaging software flipped most, 44.9% of 247 active cells, and Insurance agency management systems least, 27.0% of 237. No category was stable. The two categories with the smallest vendor lists, CRM and project management (500 cells each), also had the lowest all-cell unanimity (85.6% and 84.0%); with ten vendors each they have fewer never-named cells, so a smaller share of their cells is quiet.

CATEGORY	CELLS	ACTIVE	FLIP CELLS	FLIP % OF ACTIVE	UNANIMOUS %	NAMINGS	NOT BY BOTH
Radiology and medical imaging software	1,200	247	111	44.9%	90.8%	573	28.8%
Transportation management systems	1,250	235	105	44.7%	91.6%	535	27.1%
Veterinary practice management software	931	262	115	43.9%	87.6%	598	26.3%
Hotel property management software	1,225	240	103	42.9%	91.6%	554	25.8%
Project management	500	190	80	42.1%	84.0%	443	25.5%
Loan origination and lending software	1,250	241	100	41.5%	92.0%	559	24.3%
Field service management software	1,100	237	91	38.4%	91.7%	578	24.2%
Residential property management software	1,056	246	91	37.0%	91.4%	588	20.9%
CRM	500	197	72	36.5%	85.6%	476	21.2%
Dealership management systems	1,100	261	95	36.4%	91.4%	633	21.3%
Legal practice management software	1,050	251	82	32.7%	92.2%	625	18.9%
Construction project management software	1,100	217	65	30.0%	94.1%	555	17.8%
Dental practice management software	980	288	85	29.5%	91.3%	732	16.8%
Insurance agency management systems	950	237	64	27.0%	93.3%	606	14.4%

Sorted by flip share of active cells. Base: cells and active cells per category (columns 2 and 3); namings per category (column 7) for the last column.

Finding 6. What a single run would have said: more vendors absent, and a different order

Vendors an engine never named: one run versus three runs

A single run calls on average 16 to 31 more vendors absent per engine than three runs do (base: 286 checked vendors, 140 questions).

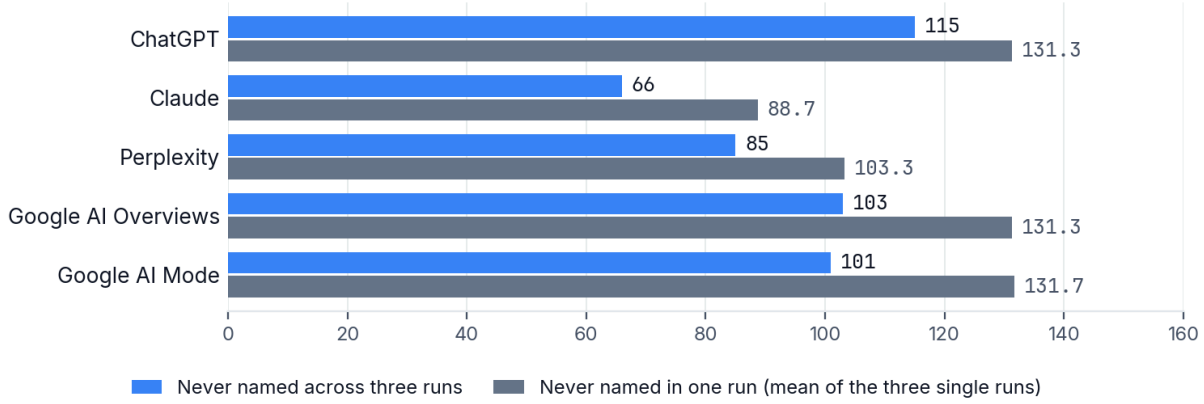


Chart 2. Vendors an engine never named in any of its 140 questions: across three runs, and in one run (mean of the three single runs). Base: 286 checked vendors per engine.

ENGINE	NEVER NAMED, THREE RUNS	EACH SINGLE RUN	SINGLE-RUN MEAN	EXTRA CALLED ABSENT
ChatGPT	115	126, 134, 134	131.3	16.3
Claude	66	89, 89, 88	88.7	22.7
Perplexity	85	108, 101, 101	103.3	18.3
Google AI Overviews	103	134, 132, 128	131.3	28.3
Google AI Mode	101	134, 133, 128	131.7	30.7

Base: 286 checked vendors per engine, 140 questions per run.

A vendor that no run of an engine ever named is a strong absence. A vendor that one run never named is weaker: on ChatGPT, three runs left 115 vendors unnamed, while a single run left 131.3 on average, 16.3 more. On Google AI Mode the gap was 30.7 vendors (131.7 against 101) and on Google AI Overviews 28.3. Across all five engines together the effect is smaller, because a vendor missed by one engine is often caught by another.

Across all engines

SINGLE RUN	NEVER NAMED IN THAT RUN	NEVER NAMED, THREE RUNS	EXTRA CALLED ABSENT	OVERSTATEMENT
Run 1	57	46	11	23.9%
Run 2	59	46	13	28.3%
Run 3	57	46	11	23.9%

Base: 286 checked vendors; a vendor is "never named" when no valid answer in the base named it. Overstatement is the extra count divided by the three-run count.

Across all five engines and 140 questions, a single run called 57 to 59 vendors never named. Three runs called 46. So one run overstated the count of absent vendors by 11 to 13 vendors, or 23.9% to 28.3%. 28 vendors (9.8% of 286) were absent in at least one run and present in at least one other; the dataset flags each of them.

Finding 7. Order within a category: the leader is usually safe, the middle of the table is not

Ranking vendors by namings within each category from a single run instead of three, a vendor’s single-run rank differed from its three-run rank 49.1% of the time (421 of 858 vendor-runs; base 286 vendors × 3 runs), under competition ranking, where tied vendors share a rank. With single-run ties broken in the three-run order the floor is 28.3% (243 of 858). The table below gives the figure under each tie rule. The top vendor of a category, the name most likely to be quoted, differed from the three-run top vendor in 3 of 42 single-run category rankings (14 categories × 3 runs), not counting ties. All three cases are listed further below, taken from the dataset file vendors_by_run_2026-09.csv.

TIE RULE FOR SINGLE-RUN RANKS	VENDOR-RUNS WHOSE RANK CHANGED	OF	SHARE
Competition ranking (tied vendors share a rank; the tie-neutral standard)	421	858	49.1%
Ties broken in the three-run order (the floor)	243	858	28.3%
Ties broken in the release’s vendor order	387	858	45.1%
Ties broken alphabetically	446	858	52.0%

Base: 858 vendor-runs (286 vendors × 3 runs). A vendor-run counts as changed when the vendor’s rank from that single run differs from its rank over three runs.

CATEGORY	SINGLE RUN	TOP VENDOR OVER THREE RUNS (NAMED OF BASE)	TOP VENDOR IN THAT RUN (NAMED OF BASE, THREE RUNS)	NAMINGS IN THAT RUN
CRM	Run 1	HubSpot CRM (113 of 150)	Salesforce Sales Cloud (109 of 150)	38 against 37 of 50
Dental practice management software	Run 1	Dentrix (120 of 149)	Curve Dental (115 of 149)	41 against 39 of 50
Transportation management systems	Run 1	AscendTMS (64 of 150)	Alvys (58 of 150)	21 against 20 of 50

Base: 286 checked vendors in 14 categories; a single run has about 50 valid answers per vendor, three runs about 150.

In each case the two vendors were close over three runs and one run tipped the order. That is the shape of the noise: it rarely lifts an obscure vendor to the top, and it often reorders neighbors. A reader who quotes a single run's leader in a category will usually be right; a reader who quotes the order below the leader is quoting a position that a single run moved in about half of cases.

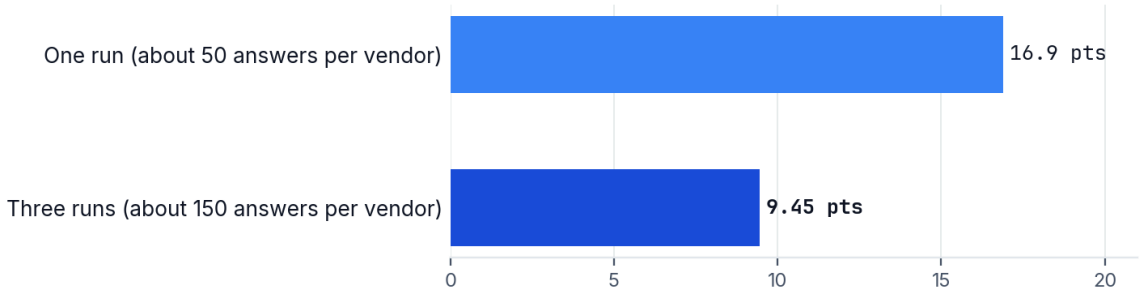
Vendors that come and go

28 vendors (9.8% of 286) had zero namings in at least one run and at least one naming in another. All of them are low-rate vendors: none was named more than 5 times in about 150 answers, and a vendor named that rarely can easily draw a run with none. The dataset flags each of them in the column absent_in_some_runs_present_in_others. For these vendors a single run cannot distinguish "never named" from "rarely named", and a claim of absence should not be made from one run.

Finding 8. Interval widths: three runs cut the uncertainty around a mention rate almost in half

How wide is the 95% interval around a vendor's mention rate?

Mean width of the Wilson 95% interval, in percentage points. Base: 286 checked vendors, 14 categories.



Source: Broadcastwell Research, Study 3, from the Absence Index release 2026-09, DOI 10.5281/zenodo.22907695. CC BY 4.0.

Chart 3. Mean width of the Wilson 95% interval around a vendor's mention rate, one run against three. Base: 286 checked vendors.

16.9 pts
mean interval width with one run (about 50 answers per vendor)

9.45 pts
mean interval width with three runs (about 150 answers); median 9.8 pts

3.8 pts
mean spread between a vendor’s highest and lowest single-run mention rate

A mention rate is a share of answers, and a share carries an interval. With one run a vendor’s rate rests on about 50 answers and its Wilson 95% interval is 16.9 points wide on average. With three runs the base is about 150 answers and the mean width is 9.45 points (median 9.8). That is the plain arithmetic of sample size, which is why three runs, not one, are the Index’s unit.

The run-to-run spread is smaller than the interval, which is what should happen if the runs are draws from the same underlying answer. The mean gap between a vendor’s highest and lowest single-run rate was 3.8 points. 10 of 286 vendors had a spread above 10 points and 0 had a spread above 20.

The widest case: Jira in project management

RUNS	NAMED	VALID ANSWERS	MENTION RATE (WILSON 95%)
Run 1	16	50	32.0%
Run 2	22	50	44.0%
Run 3	25	50	50.0%
Three runs	63	150	42.0% (34.4 to 50.0)

Base: valid target answers for the project management category (10 questions × 5 engines per run).

Jira was named in 16, 22 and 25 of 50 answers across the three runs: a single-run mention rate anywhere from 32.0% to 50.0%, a spread of 18.0 points. Three runs put it at 42.0% with an interval of 34.4 to 50.0. Someone quoting run 1 would call Jira a minority mention in this category; someone quoting run 3 would call it a coin flip. Both would be quoting real answers.

Per-segment table

Every segment in the study on one page: per engine, per category and per question shape. Flip % is the share of active cells where the three runs did not agree. The last two columns class single-run namings: not repeated by both other runs (the headline measure) and named by neither other run (the strict measure).

SEGMENT	NAME	CELLS	ACTIVE	FLIP %	NAMINGS	NOT BY BOTH	BY NEITHER
All	All five engines	14,192	3,349	37.6%	8,055	22.2%	9.1%
Engine	ChatGPT	2,860	580	27.4%	1,496	15.6%	5.7%
Engine	Claude	2,860	812	31.2%	2,039	17.8%	7.1%
Engine	Perplexity	2,860	653	30.6%	1,642	17.2%	7.1%
Engine	Google AI Overviews	2,752	599	50.4%	1,322	32.6%	13.1%
Engine	Google AI Mode	2,860	705	48.9%	1,556	30.6%	13.8%
Category	Construction project management software	1,100	217	30.0%	555	17.8%	5.6%
Category	Dealership management systems	1,100	261	36.4%	633	21.3%	8.7%
Category	Dental practice management software	980	288	29.5%	732	16.8%	6.4%
Category	Field service management software	1,100	237	38.4%	578	24.2%	7.3%
Category	Hotel property management software	1,225	240	42.9%	554	25.8%	11.4%
Category	Insurance agency management systems	950	237	27.0%	606	14.4%	6.8%
Category	Legal practice management software	1,050	251	32.7%	625	18.9%	7.4%
Category	Loan origination and lending software	1,250	241	41.5%	559	24.3%	11.4%
Category	Radiology and medical imaging software	1,200	247	44.9%	573	28.8%	9.9%
Category	Residential property management software	1,056	246	37.0%	588	20.9%	10.0%
Category	Transportation management systems	1,250	235	44.7%	535	27.1%	12.1%
Category	Veterinary practice management software	931	262	43.9%	598	26.3%	12.2%
Category	CRM	500	197	36.5%	476	21.2%	9.0%
Category	Project management	500	190	42.1%	443	25.5%	10.6%
Shape	Alternative	4,090	1,470	35.7%	3,591	21.1%	8.2%
Shape	Best option	4,151	996	41.2%	2,350	25.2%	9.7%
Shape	Buyer evaluation	4,246	423	31.9%	1,040	16.9%	9.0%
Shape	Specific workflow	1,605	403	44.2%	919	26.6%	12.2%
Shape	Top tools	100	57	19.3%	155	11.0%	3.2%

Base: cells with three valid runs (column 3), active cells (column 4) and single-run namings (column 6) per segment. Categories in release order. 22 to 23 Sep 2026 (UTC).

Limitations

Three runs is a small sample per cell. With three draws, "repeated by both", "repeated by one" and "named by neither" are coarse classes. A vendor named in two of three runs might be named in seven of ten; a vendor named once might never appear again. The cell-level figures describe three runs and should not be read as long-run probabilities.

Same-day runs understate longer-horizon variation. All runs were captured inside one window on 22 to 23 Sep 2026 (UTC). Variation between days, weeks or releases is likely larger, and this study says nothing about it. It measures within-day repeat variation only.

API based and logged out, not a consumer interface. Broadcastwell's published method (broadcastwell.com/methodology, version 1.1, 8 September 2026) states: "The measurement is single turn, API based, logged out, US English, gl=us and hl=en" and "It is a controlled instrument, not a replica of any one buyer's screen." A logged-in user with history, location or personalization may see more or less variation than this instrument did.

The question set is fixed. 140 questions, 10 per category, written for the Index in five shapes. Other phrasings, or follow-up turns, may vary more or less. The "top tools" shape rests on 100 cells and is not a stable estimate on its own.

Checked vendor lists come from buyer directories. A vendor not on a category's list is not measured, and a naming of a vendor outside the list does not count anywhere. The lists have 10 to 25 vendors per category.

Plain-word vendor rule. Vendors whose names are ordinary words need a listed qualifier within five words or their canonical domain among the cited URLs to count. A loosely worded mention of such a vendor is scored as not named, which can make a naming look unrepeated when the engine did refer to the vendor in passing.

Excluded answers drop 108 cells. The release excluded five target answers with a stated reason. The cells that lost a run to those exclusions, all in Google AI Overviews, are left out here rather than scored on two runs. The engine's cell base is therefore 2,752 instead of 2,860.

No causal claim. This study measures the outcome, whether a vendor was named, and not the cause. Retrieval differences, sampling, cached results, the set of pages an engine happened to search, and the length of the answer can all move a vendor in or out. None of them is identified here.

Engines change their models and retrieval without notice. The figures describe five engines as they answered on 22 to 23 Sep 2026 (UTC). A later run may be steadier or noisier, and nothing here predicts which.

Single-run ties are common among rarely named vendors. Competition ranking gives tied vendors the same rank; breaking ties in the three-run order lowers the figure to 28.3%, the floor.

How to cite

APA

```
Broadcastwell Research. (2026). Run-to-run noise in AI search answers: three runs of 140 buyer questions on five engines (Version 1.0) [Data set and report]. Broadcastwell. https://doi.org/10.5281/zenodo.22982432
```

BibTeX

```
@misc{broadcastwell2026runnoise,
  author    = {{Broadcastwell Research}},
  title     = {Run-to-run noise in AI search answers: three runs of 140 buyer questions on five engines},
  year      = {2026},
  version   = {1.0},
  publisher = {Broadcastwell},
  doi       = {10.5281/zenodo.22982432},
  url       = {https://doi.org/10.5281/zenodo.22982432},
  note      = {Data set and report. CC BY 4.0. Built on The Absence Index, release 2026-09, DOI 10.5281/zenodo.22907695}
}
```

Source dataset

```
Broadcastwell. (2026). The Absence Index: September 2026 Release (2026-09) [Dataset]. Broadcastwell. https://doi.org/10.5281/zenodo.22907695
```


Appendix

A. Data dictionary: run_cells_2026-09.csv

One row per question × engine × vendor cell with three valid runs; 14,192 rows.

COLUMN	MEANING
category_id	Category identifier as used in the release (see the mapping below)
question_id	Question identifier within the category (q1 to q10); the exact text is in the release
question_type	Question shape: best option, specific workflow, top tools, alternative, buyer evaluation
engine	ChatGPT, Claude, Perplexity, Google AI Overviews, Google AI Mode
vendor	Checked vendor name as listed in the release
run1, run2, run3	1 if that scheduled run's answer named the vendor under the matching rule, 0 if not
named_count	run1 + run2 + run3 (0 to 3)
unanimous	1 if named_count is 0 or 3, else 0

B. Data dictionary: vendors_by_run_2026-09.csv

One row per checked vendor; 286 rows. Bases count valid target answers, so a run base is 50 except where an answer was excluded.

COLUMN	MEANING
category, vendor	Category name and checked vendor name
named_3runs, base_3runs	Answers naming the vendor across three runs, and valid answers in the base (about 150)
rate_3runs, lower, upper, interval_width_pts	Mention rate over three runs in percent, its Wilson 95% interval, and the interval width in points
runN_named, runN_base (N = 1, 2, 3)	Answers naming the vendor in run N and valid answers in that run (about 50)
single_run_rate_min, single_run_rate_max, single_run_spread_pts	Lowest and highest single-run mention rate in percent, and their difference in points
absent_in_some_runs_present_in_others	yes if the vendor had zero namings in at least one run and at least one naming in another

C. Category identifiers

CATEGORY_ID	CATEGORY	CHECKED VENDORS
construction-project-management	Construction project management software	22
dealership-management-systems	Dealership management systems	22
dental-practice-management	Dental practice management software	20
field-service-management	Field service management software	22
hotel-property-management	Hotel property management software	25

CATEGORY_ID	CATEGORY	CHECKED VENDORS
insurance-agency-management	Insurance agency management systems	19
legal-practice-management	Legal practice management software	21
loan-origination-software	Loan origination and lending software	25
radiology-medical-imaging	Radiology and medical imaging software	24
residential-property-management	Residential property management software	22
transportation-management-systems	Transportation management systems	25
veterinary-practice-management	Veterinary practice management software	19
crm	CRM	10
project_management	Project management	10

D. Recompute commands

Run from the study folder, with the release package unpacked under src/INDEXPUBLISH_2026-09-23/ZENODO_UPLOAD_2026-09/:

```
python3 scripts/recompute_index.py
python3 scripts/study3_run_noise.py
python3 scripts/build_study3.py
```

The first command re-scores every raw answer in the release package with the published matching rule, using vendor_rules_2026-09.json, and writes the per-answer matrix (answer_vendor_matrix.csv). The second computes every figure in this report from that matrix and writes the two dataset CSVs and the numbers file. The third builds the charts and this report.

E. SHA256 of the input files

FILE	SHA256
answers-2026-09.jsonl (raw answers, from the release package)	221c94e1ebad2bc87d7cd20837fa670cc40b10dba6f68b75423dc8bee69697b3
release-2026-09.json (published release, used for the recompute diff)	e7829183f6785357787fcf85d3c27999403b0a329f349dad4c7343b11d23f142
vendor_rules_2026-09.json (vendor, alias and qualifier rules used by the matcher, shipped with the dataset)	0736bb7a842b9ef54872ee116b41b5930a7b0fc90d79493c4503b0bd1f86b156

The hashes of the two release files match the SHA256SUMS.txt manifest shipped inside the release package.