Gracker.ai

RESEARCH REPORT

AI Search Visibility in Enterprise Third-Party Risk Management

A multi-engine, multi-market benchmark study of how six AI search engines — ChatGPT, Perplexity, Gemini, Microsoft Copilot, Google AI Overview and Google AI Mode — answer buyer-intent questions about enterprise third-party risk management, vendor risk monitoring and questionnaire automation across the United States, Canada, Germany, India and Australia.

Table 1. Study scale by AI engine — responses recorded and sources cited

AI engine	Responses	Cited sources	Markets
ChatGPT	50,000	567,000	5
Microsoft Copilot	50,000	213,000	5
Gemini	50,000	119,000	5
Google AI Mode	50,000	170,000	5
Google AI Overview	50,000	449,000	5
Perplexity	50,000	33,000	5
Total	**300,000**	**1,551,000**	**5**

Published: August 2026

Research period: August 2026

Category: Enterprise third-party risk management (TPRM, VRM)

Scope: 10,000 buyer-intent queries · 300,000 AI engine responses · 6 AI engines · 5 markets

Evidence base: 1,551,000 cited sources · 127 distinct product entries

Prepared by: GrackerAI · AI Search Visibility Benchmark Series

Abstract

A buyer assembling a third-party risk management shortlist now begins inside a generated response, so the vendor set an AI engine returns forms the consideration set before any vendor is contacted. This study measures how six commercial AI search engines answer buyer-intent questions about enterprise third-party risk management. Each engine received 10,000 enterprise buyer-intent queries in each of five markets during a single collection window in August 2026, producing 300,000 responses that contained 1,551,000 cited sources and covered 127 distinct product entries. For every response the study recorded length, cited pages and their source type, page reachability, hedging, recency anchoring, truncation, refusal, and each brand mention with its ordinal position. The engines converge on a broad core: 9 of 72 brand strings were named by all six engines, led by OneTrust at 205,000 mentions, UpGuard at 165,000 and Bitsight at 161,000. The brand named most often is the brand named first in none of the six engines; Bitsight opens three, and OneTrust averages the third or fourth position in every engine. The engines cite largely separate domains, with a mean pairwise overlap of 0.19, and no two engines share a most-cited domain, yet 52.2% of everything they cite sits on a vendor's own site. Dead links are rare, at 0.2% of cited pages. OneTrust leads every one of the five markets; only the second and third positions move.

Contents

1. Introduction

1.1 Background

An enterprise buyer evaluating third-party risk management now begins inside a generated response. The buyer asks an AI search engine which platforms monitor suppliers continuously, automate the security questionnaire, and connect to the ticketing and procurement systems already in place. The engine returns a short ordered list of named vendors with a paragraph of justification for each. That list is not a ranking of the market. It is the output of one system reading one collection of web sources at one moment, and it arrives before the buyer has opened a single vendor site.

The structural consequence is narrow. A vendor absent from the returned list is not ranked low; it is not present. A generated response has no second page, so the buyer has no mechanism equivalent to scrolling further. Where a search results page offered ten positions and a set of related queries, a generated response typically names between three and seven vendors and closes. Visibility in this setting is closer to inclusion than to ranking, and the difference between being inside and outside the returned set is discontinuous.

A second consequence follows for vendors that are inside the set. A generated response is read in order, and the first name it gives carries a weight the fourth does not. Being named is one result; being named first is another; being named early on average is a third. This study records all three separately because, as §4.7 shows, the category's most-mentioned vendor holds the first of them and neither of the others.

1.2 Why this category

Enterprise third-party risk management is a suitable instrument for this measurement. The vendor set is mature and well documented, so the engines have abundant published material to draw on, and 9 of the 72 brand strings recorded were named by every engine. A category with a settled core isolates engine behaviour from genuine market churn.

The category also has a stable and publicly articulated capability vocabulary. Across the corpus the engines returned the same selection factors: continuous external monitoring through security ratings, automated questionnaire and evidence workflows, and integration with service-management and procurement systems. These criteria are specific enough that responses can be compared to each other rather than to a general description of governance software.

Buying behaviour in the category is procurement-led and the cycles are long, so the cost of omission from an early shortlist is high and difficult to reverse. A vendor omitted from the generated shortlist loses the opportunity at the earliest stage, before any conversation in which the omission could be corrected.

1.3 Contribution

This study provides four things a single-engine measurement does not. It establishes how far six engines agree on the vendor set in one category, separating the shared core from the tail each engine holds alone. It measures the overlap between the engines' evidence bases directly, as a pairwise coefficient over cited domains, and classifies every cited page by source type, so that agreement on vendors can be set against what the engines actually read. It reports mention volume, first-mention position and mean ordinal position as three separate quantities, which allows them to be shown moving independently. And it tests the vendor set across five markets in the same window, which distinguishes market-specific results from a set the engines return everywhere. All four rest on a single corpus of 300,000 responses collected under one design.

2. Research Questions

Six questions are addressed, each answerable from the recorded data and each answered explicitly in §5.

1. **RQ1.** Do the six engines converge on the same set of named vendors?
2. **RQ2.** How concentrated is brand visibility across the recorded vendor set?
3. **RQ3.** Do the engines draw on a shared evidence base?
4. **RQ4.** Does ordinal position track mention volume?
5. **RQ5.** Does the vendor set vary between the five markets?
6. **RQ6.** Do the engines cite vendors' own sites or third-party sources?

3. Methodology

3.1 Study design

The study is observational and cross-sectional. Six AI search engines were queried with a fixed query set in five markets during a single collection window in August 2026. No engine was manipulated, no result was solicited, and no intervention was applied between conditions. The design records what the engines returned; it does not test a hypothesis about why they returned it. The collected responses were processed under the instrumentation's current citation, source-type and entity rules.

3.2 Engines under test

Six engines were tested: ChatGPT, Microsoft Copilot, Gemini, Google AI Mode, Google AI Overview and Perplexity. All six are commercial services. None is version-pinned, none exposes a stable model identifier to the caller, and all are subject to continuous change by their operators. A result reported here describes the service as it behaved during the August 2026 window, not a fixed artefact that can be re-queried later in the same state. Three of the six — Gemini, Google AI Mode and Google AI Overview — are operated by the same parent, which is relevant to the overlap results in §4.6.

3.3 Markets

Five markets were tested: the United States, Canada, Germany, India and Australia. Market variance was tested because engines localise results, and a vendor set returned in one market cannot be assumed to hold in another. The five combine three large English-language markets on three continents, one large non-English market, and one market with a substantial domestic buyer base and a separate regulatory environment. Each query was issued in every market, so market is a within-design condition rather than a separate sample.

3.4 Query construction

The query set contains 10,000 distinct buyer-intent queries. It combines core category questions of the form a buyer would ask when assembling a shortlist with paraphrase variants of those questions, so that the stability of a returned vendor set under rewording can be observed rather than assumed. Queries were written to elicit vendor recommendations rather than definitions, and they name the buyer's scale and integration constraints as a buyer would.

Every query was issued to every engine in every market. The corpus identity follows directly: 10,000 queries in five markets gives 50,000 responses per engine, and 10,000 queries across 6 engines and 5 markets gives 300,000 responses in total.

3.5 Corpus and unit of analysis

The unit of analysis is one engine response to one query in one market. Metrics are computed over responses and aggregated to the engine, the brand, the product entry or the market as the metric requires. The corpus contains 300,000 responses, distributed evenly across the six engines at 50,000 responses each and across the five markets at 60,000 responses each. Every per-engine comparison in §4 therefore rests on the same denominator.

3.6 Metric definitions

Every metric used anywhere in this report, including the appendices, is defined here.

Table 2. Metric definitions — all metrics used in this report

Metric	Definition	Unit
Response	One engine output returned for one query in one market.	count
Mean response length	Arithmetic mean word count of an engine's responses.	words
Hedging rate	Share of an engine's responses containing explicit qualification of a recommendation.	% of responses
Recency-anchored rate	Share of an engine's responses that anchor a claim to an explicit calendar date.	% of responses
Truncation rate	Share of an engine's responses ending mid-sentence, inferred from the terminal characters of the response text.	% of responses
Refusal rate	Share of an engine's responses declining to answer the query.	% of responses
First-mention vendor	The brand most often named before any other brand in an engine's responses.	brand
Brand mention	One occurrence of a vendor brand name in a response. Repeat occurrences within one response are counted once. Spelling variants of one brand are merged.	count
Mean ordinal position	Arithmetic mean of the position at which a brand is named within a response, counting named vendors from one.	position
Mean sentiment	Arithmetic mean of the score the analysis model assigned to how favourably the response describes the brand, on a scale from minus one to one, under a fixed rubric.	score
Cited source	One distinct page a response cites, in its source list or as an in-text link, after removal of page assets, trackers and advertising links. Variants of one page differing only in tracking parameters or text fragments count once.	count
Distinct domains	The number of different registrable domains cited by that engine across its responses. Reported per engine and never summed across engines.	count

Metric	Definition	Unit
Source type	The class of the cited page's domain: vendor site (a domain of a vendor named in the response), review aggregator, news, forum, social, academic, government, or other.	class
Dead-link rate	Share of an engine's cited pages that returned not-found, gone or a server error, or whose host could not be reached, when tested. A page that refused the request is not a dead link.	% of cited sources
Self-citation rate	Share of an engine's cited sources hosted on a domain operated by the engine's own parent.	% of cited sources
Jaccard coefficient	Size of the intersection of two engines' cited-domain sets divided by the size of their union. Zero is disjoint, one is identical.	ratio
Product entry	One product recorded against a brand, after plan, edition and deployment spellings of one product are merged.	count
Mean rank	Arithmetic mean of the ordered position of a product entry within the response that named it.	position

3.7 Data collection and processing

Responses were collected through each engine's public interface and normalised to plain text. Word counts, terminal-character inspection for truncation, and detection of qualifying and date-anchoring language were computed over that normalised text.

Cited sources were taken as the union of the source list each engine attached to a response and the pages the response linked to in its text. Page assets, tracking and advertising links were removed, and variants of one page that differ only in tracking parameters or highlighted-text fragments were counted once. Where an engine wrapped a citation in a redirect, the redirect was followed to the cited page. Each page was then reduced to its registrable domain for the distinct-domain, source-type and overlap measures. Source type was assigned by rule from the domain. A domain belonging to a vendor the response names is a vendor site. Forums, social and video sites, news publishers, academic hosts and review aggregators were matched against lists and host-name patterns, and the remainder is other.

Reachability was tested once per page. A page was recorded as a dead link when it returned not-found, gone or a server error, or when its host could not be reached after a retry. A page that refused the request, as sites behind bot protection do, was recorded as reachable, since the page exists.

Brand mentions were detected against the vendor names each response used and recorded with the ordinal position at which the brand appeared among the named vendors, counting from one. Spelling variants of one brand, such as a vendor's name with and without a corporate suffix, were merged, and plan, edition and category-noun variants of one product were merged into one product entry. The first-mention vendor for an engine is the brand that most often held position one. Leading brands by market are the brands ranked by mention count within the responses collected for that market. Sentiment was assigned by the analysis model to each brand and product mention, on a scale from minus one to one. The rubric scores how favourably the response describes the entity and is documented with the instrumentation.

One implementation detail is not documented by the instrumentation: the lexicon used to detect qualifying and date-anchoring language. §6.1 states the consequence for how those two rates should be read. Source age was recorded where a page declared a publication date, but fewer than a fifth of cited pages did, so it is not reported.

3.8 Scope exclusions

The study did not paraphrase, summarise or edit engine output before analysis. It did not check any claim an engine made against an external source, and it did not score responses for accuracy. It did not test, benchmark or evaluate any vendor product. No placement was paid for, solicited or arranged. The object of study is how the engines described the category, not whether their descriptions were correct.

4. Results

4.1 Response characteristics by engine

The six engines construct responses differently. Table 3 gives the response-construction measures for each engine over an identical number of responses.

Table 3. Response construction by AI engine — n = 50,000 responses per engine, 300,000 in total

AI engine	Responses	Mean words	Hedging	Recency-anchored	Truncation	Refusal	First-mention vendor
ChatGPT	50,000	520.0	17%	23%	91%	0%	Whistic
Microsoft Copilot	50,000	475.8	8%	67%	0%	0%	Bitsight
Gemini	50,000	348.1	0%	0%	31%	0%	ServiceNow
Google AI Mode	50,000	292.4	0%	2%	2%	0%	UpGuard
Google AI Overview	50,000	211.0	4%	18%	0%	0%	Bitsight
Perplexity	50,000	318.4	0%	20%	20%	0%	Bitsight

Mean response length spans a factor of 2.5, from 211.0 words for Google AI Overview to 520.0 for ChatGPT. Figure 1 shows the distribution. One engine exceeds 500 words — ChatGPT — and Google AI Mode and Google AI Overview fall below 300.

Truncation was recorded in four of the six engines: 91% for ChatGPT, 31% for Gemini, 20% for Perplexity and 2% for Google AI Mode, with ChatGPT's the highest value in the study. Microsoft Copilot and Google AI Overview recorded none. Recency anchoring is highest for Microsoft Copilot at 67%, and Gemini recorded none. Hedging was recorded in three engines — ChatGPT at 17%, Microsoft Copilot at 8%, Google AI Overview at 4% — and Gemini, Google AI Mode and Perplexity recorded none. No engine refused a query: the refusal rate is 0% for all six. The first-mention column names four different vendors across the six engines, and OneTrust, the most-mentioned brand in the corpus, is not among them; §4.7 reports the position data behind that column.

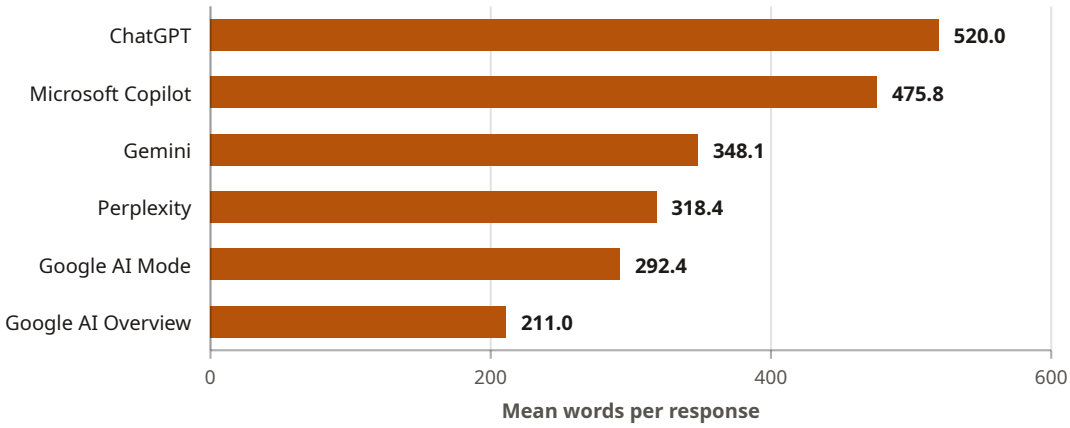

Figure 1. Mean response length by AI engine. Arithmetic mean word count per response; n = 50,000 responses per engine. Source: GrackerAI AI Search Visibility Benchmark Series, August 2026.

4.2 Brand visibility

Brand visibility was recorded for 72 brand strings, listed in full in Appendix A. Table 4 gives the twelve most-mentioned, with the number of engines that named each and its mean sentiment. Figure 2 shows the same twelve.

Table 4. Brand visibility, twelve most-mentioned brands — of 72 brand strings recorded; mean sentiment where recorded, a dash otherwise; n = 300,000 responses

Brand	Total mentions	Engines recording	Mean sentiment
OneTrust	205,000	6	0.77
UpGuard	165,000	6	0.80
Bitsight	161,000	6	0.80
SecurityScorecard	146,000	6	0.74
ProcessUnity	132,000	6	0.80
Panorays	104,000	6	0.78
ServiceNow	102,000	6	0.78
Vanta	59,000	6	0.79
Whistic	51,000	5	0.80
MetricStream	34,000	5	0.74
Drata	29,000	5	—
Riskonnect	27,000	6	—

Visibility is broad at the top and long-tailed below it. The most-mentioned brand, OneTrust, records 205,000 mentions, 40,000 ahead of UpGuard at 165,000; Bitsight follows at 161,000, 4,000 behind. The three together take 35.5% of all mentions recorded across the 72 brand strings, and the five leading brands 54.1%. Below ProcessUnity at 132,000 the counts fall away: Panorays records 104,000, ServiceNow 102,000, and no brand outside the seven most-mentioned exceeds 59,000.

9 of the 72 brand strings were named by all six engines: OneTrust, UpGuard, Bitsight, SecurityScorecard, ProcessUnity, Panorays, ServiceNow, Vanta and Riskonnect. 8 more were named by five, among them Whistic and MetricStream. 38 were named by exactly one engine, none of them above 10,000 mentions. The strings recorded include adjacent categories — compliance automation, governance platforms, procurement suites and ticketing tools — each named in a small number of responses as the engines wrote them. Mean sentiment for the 10 brands in Table 4 with a recorded score occupies a band from 0.74 for SecurityScorecard to 0.80 for UpGuard, a spread of 0.06.

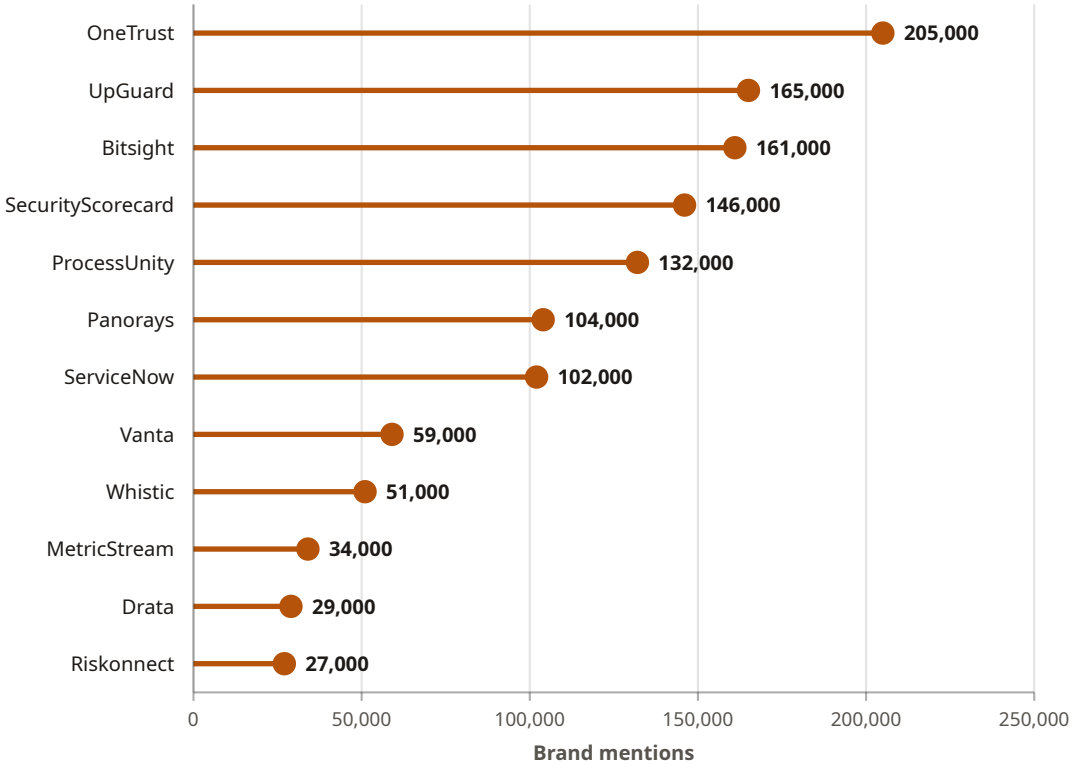

Figure 2. Brand mentions, twelve most-mentioned brands. The 12 most-mentioned of 72 brand strings recorded; n = 300,000 responses. Source: GrackerAI AI Search Visibility Benchmark Series, August 2026.

4.3 Product-level results

127 distinct product entries were recorded after plan, edition and category-noun spellings of one product were merged. An entry is one product as an engine named it, so a brand with several products, or several names for one product, holds several entries: OneTrust holds 9, ServiceNow 8 and Bitsight 6. Table 5 gives the ten most-mentioned entries.

Table 5. Product entries, ten most-mentioned — of 127 entries; n = 300,000 responses

Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
Bitsight	Bitsight	146,000	2.8	0.82	6	5
UpGuard	UpGuard	134,000	2.8	0.82	6	5
OneTrust	OneTrust	127,000	3.5	0.78	6	5
SecurityScorecard	SecurityScorecard	124,000	3.8	0.79	6	5
ProcessUnity	ProcessUnity	105,000	4.5	0.82	6	5

Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
Panorays	Panorays	95,000	3.8	0.81	6	5
Prevalent	Prevalent	64,000	4.5	0.77	6	5
Vanta	Vanta	50,000	3.9	0.82	6	5
Whistic	Whistic	44,000	3.1	0.81	5	5
ServiceNow TPRM	ServiceNow	37,000	3.5	0.81	5	5

The product table does not reproduce the brand table's order. The leading entry is Bitsight at 146,000 mentions, and the bare OneTrust entry is third at 127,000, because OneTrust's mentions are spread across 9 entries, of which the bare name is one. Bitsight and UpGuard were mostly named by their bare brand, which is why they lead the product table while OneTrust leads the brand table.

Product entries carry their own brand attributions, and Appendix C lists entries attributed to brands outside the twelve in Table 4. The distribution is long-tailed. 102 of the 127 entries record fewer than 10,000 mentions, 61 record exactly 1,000, and 81 were named by exactly one engine. 9 entries were named by all six engines, 27 were recorded in all five markets, and 64 were recorded in a single market. The full table is in Appendix C.

Mean rank among the ten most-mentioned entries runs from 2.8 to 4.5. The leading entry, Bitsight, records 146,000 mentions at a mean rank of 2.8, and the second, UpGuard, records 134,000 at 2.8.

4.4 Citation volume and source type

The corpus contains 1,551,000 cited sources, distributed unevenly across the six engines, as Figure 3 shows. Table 6 gives volume and domain breadth; Table 7 gives the source-type mix.

Table 6. Cited sources and domain breadth by AI engine — n = 1,551,000 cited sources

AI engine	Cited sources	Share of corpus	Distinct domains	Most-cited domain
ChatGPT	567,000	36.6%	65	whistic.com (69,000)
Microsoft Copilot	213,000	13.7%	23	zipdo.co (37,000)
Gemini	119,000	7.7%	34	spycloud.com (29,000)
Google AI Mode	170,000	11.0%	52	gartner.com (17,000)
Google AI Overview	449,000	28.9%	65	bitsight.com (49,000)
Perplexity	33,000	2.1%	23	upguard.com (4,000)
Total	**1,551,000**	**—**	**—**	**—**

ChatGPT and Google AI Overview together account for 65.5% of all cited sources in the corpus, at 567,000 and 449,000. ChatGPT and Google AI Overview cite from the widest domain population at 65 distinct domains each, and Microsoft Copilot and Perplexity from the narrowest at 23. Distinct-domain counts are reported per engine and are not summed across engines.

No two engines share a most-cited domain. ChatGPT's is whistic.com, a questionnaire-automation vendor; Google AI Overview's is bitsight.com and Perplexity's is upguard.com, two of the leading vendors; Google AI Mode's is gartner.com, an analyst firm; Gemini's is spycloud.com, a security vendor outside the leading group; and Microsoft

Copilot's is zipdo.co, a statistics aggregator. Appendix B gives the three most-cited domains for each engine.

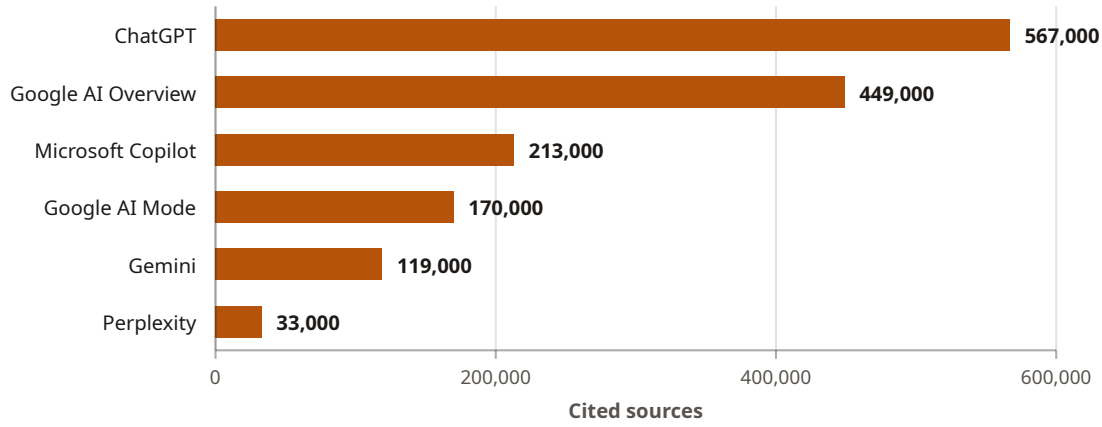

Figure 3. Cited sources by AI engine. Sources cited per engine; n = 1,551,000 cited sources. Source: GrackerAI AI Search Visibility Benchmark Series, August 2026.

Table 7. Source type by AI engine — share of each engine's cited sources by the class of the cited domain; n = 1,551,000 cited sources

AI engine	Vendor site	Other	Review aggregator	Social	Forum	Academic
ChatGPT	72.3%	15.7%	11.6%	0.0%	0.0%	0.4%
Microsoft Copilot	39.9%	55.4%	4.7%	0.0%	0.0%	0.0%
Gemini	42.0%	46.2%	11.8%	0.0%	0.0%	0.0%
Google AI Mode	36.5%	44.7%	10.6%	8.2%	0.0%	0.0%
Google AI Overview	41.0%	49.9%	5.1%	3.6%	0.4%	0.0%
Perplexity	54.5%	39.4%	6.1%	0.0%	0.0%	0.0%
All engines	**52.2%**	**37.1%**	**8.6%**	**1.9%**	**0.1%**	**0.1%**

Across the corpus, 52.2% of cited sources sit on a vendor site, 809,000 of 1,551,000, and 37.1% fall outside every named class. Review aggregators, which here include analyst firms, account for 8.6%, and the remaining classes — social, forum and academic — for 2.2% together. The mix differs by engine. Vendor sites are the largest class for two of the six engines, ChatGPT and Perplexity, at 72.3% and 54.5%; for the other four the residual class is largest, with vendor sites between 36.5% for Google AI Mode and 42.0% for Gemini. Review aggregators reach 11.8% of Gemini's citations, the highest share. No news or government domain was recorded by any engine. Figure 4 shows the vendor-site share by engine, and Appendix B gives the counts behind Table 7.

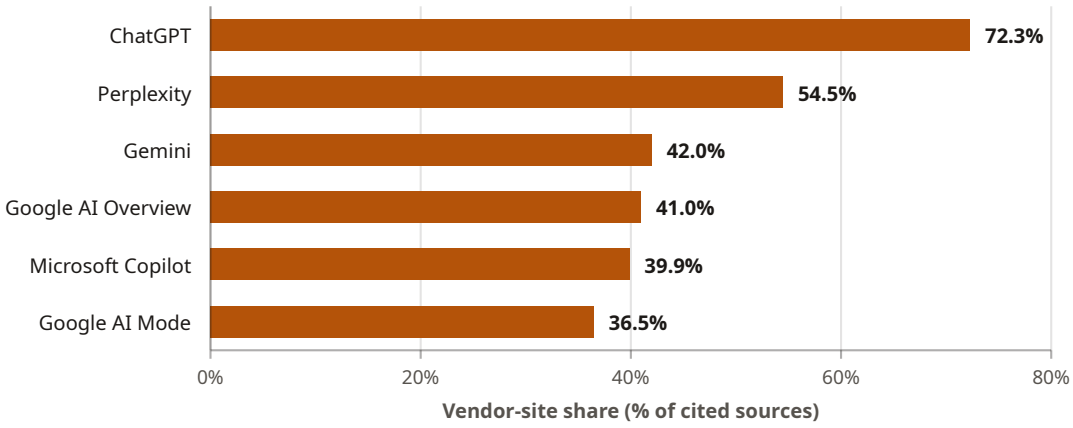

Figure 4. Share of cited sources on vendor sites by AI engine. Cited sources whose domain belongs to a vendor named in the response, as a share of that engine's cited sources; n = 1,551,000 cited sources. Source: GrackerAI AI Search Visibility Benchmark Series, August 2026.

4.5 Source quality and link decay

A small share of the pages the engines cited did not resolve when tested. Table 8 gives the rate for each engine.

Table 8. Dead links and self-citation by AI engine — n = 1,551,000 cited sources

AI engine	Cited sources	Dead-link rate	Dead links	Self-citation rate
ChatGPT	567,000	0.4%	2,270	0%
Microsoft Copilot	213,000	0%	0	0%
Gemini	119,000	0%	0	0%
Google AI Mode	170,000	0.6%	1,020	0%
Google AI Overview	449,000	0%	0	0%
Perplexity	33,000	0%	0	0%
Total	**1,551,000**	**0.2%**	**3,290**	**—**

The overall dead-link rate is 0.2%, 3,290 of 1,551,000 cited pages. Four engines — Microsoft Copilot, Gemini, Google AI Overview and Perplexity — record no dead links at all, and no engine exceeds 0.6%, the rate for Google AI Mode. ChatGPT accounts for 2,270 of the 3,290 dead links, 69.0% of the total, on a rate of 0.4% applied to the largest citation volume.

The self-citation rate is 0% for all six engines.

4.6 Cross-engine source overlap

The Jaccard coefficient over each pair of engines' cited-domain sets measures how far the engines read the same web. Table 9 gives the full matrix.

Table 9. Cross-engine cited-domain overlap — Jaccard coefficient over cited-domain sets; 0 is disjoint, 1 is identical; n = 1,551,000 cited sources

	ChatGPT	Copilot	Gemini	Google AI Mode	Google AI Overview	Perplexity
ChatGPT	**1.00**	0.09	0.22	0.21	0.21	0.16
Copilot	0.09	**1.00**	0.10	0.10	0.09	0.15
Gemini	0.22	0.10	**1.00**	0.28	0.30	0.20
Google AI Mode	0.21	0.10	0.28	**1.00**	0.40	0.14
Google AI Overview	0.21	0.09	0.30	0.40	**1.00**	0.13
Perplexity	0.16	0.15	0.20	0.14	0.13	**1.00**

Shading increases with the coefficient: < 0.10 · 0.10–0.17 · 0.18–0.25 · 0.26–0.33 · ≥ 0.34

The mean coefficient across the 15 distinct engine pairs is 0.19. The highest value is 0.40, between Google AI Mode and Google AI Overview. The lowest is 0.09, recorded twice, for Microsoft Copilot with ChatGPT and with Google AI Overview. The three engines operated by the same parent — Gemini, Google AI Mode and Google AI Overview — record coefficients of 0.28, 0.30 and 0.40 with each other, the three highest values in the matrix.

Microsoft Copilot is the lowest-overlap partner for four of the five other engines; its five values run from 0.09 to 0.15, the latter with Perplexity, which overlaps least with Google AI Overview instead. No pair of engines outside the same-parent group exceeds 0.22, the value for ChatGPT with Gemini.

4.7 First-mention position

Mention volume records how often a brand appears. Mean ordinal position records how early it appears among the vendors a response names. The first-mention vendor is the brand that most often opens a response. Table 10 gives all three for the three most-mentioned brands, and Figure 5 shows OneTrust against Bitsight.

Table 10. Mean ordinal position of the three leading vendors — lower is earlier in the response; n = 300,000 responses

AI engine	OneTrust	UpGuard	Bitsight	First-mention vendor
ChatGPT	3.2	3.1	2.8	Whistic
Microsoft Copilot	3.2	3.2	3.3	Bitsight
Gemini	3.6	2.5	2.5	ServiceNow
Google AI Mode	3.4	1.7	2.5	UpGuard
Google AI Overview	3.3	2.6	2.7	Bitsight
Perplexity	2.8	4.8	3.3	Bitsight

OneTrust holds the first-mention position in none of the six engines, on 205,000 mentions, the highest total in the corpus. Bitsight holds it in three, Microsoft Copilot, Google AI Overview and Perplexity, on 161,000. Whistic holds it in ChatGPT; ServiceNow holds it in Gemini; UpGuard holds it in Google AI Mode. In ChatGPT the opening vendor, Whistic, records 20,000 mentions against OneTrust's 43,000 in the same engine, and opens the response at a mean position of 1.9.

Mean ordinal position tells the same story from the other side. OneTrust's mean position runs from 2.8 to 3.6: on average the third or fourth vendor named, in every engine. UpGuard records the earlier mean position in four engines, Bitsight in four, and OneTrust is earlier than UpGuard only in Perplexity, with a tie in Microsoft Copilot. The gap is widest in Google AI Mode, where UpGuard averages 1.7 against OneTrust's 3.4. The first-mention vendor is a mode, not a mean, and the two can part: ServiceNow opens Gemini most often while averaging position 4.8 there, which is consistent with an engine that names a vendor first in some responses and near the end in others.

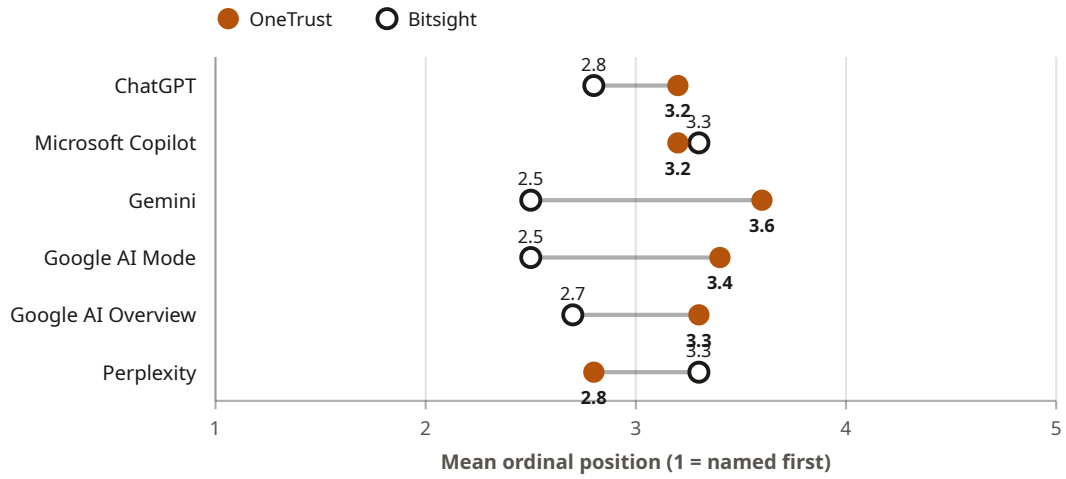

Figure 5. Mean ordinal position of OneTrust and Bitsight by AI engine. Lower is earlier in the response; n = 300,000 responses. Source: GrackerAI AI Search Visibility Benchmark Series, August 2026.

4.8 Geographic variance

Table 11 gives the three leading brands in each market, ranked by mentions within that market.

Table 11. Leading brands by market — ranked by brand mentions within the market; n = 60,000 responses per market

Market	First	Second	Third
United States	OneTrust	UpGuard	Bitsight
Canada	OneTrust	Bitsight	UpGuard
Germany	OneTrust	Bitsight	UpGuard
India	OneTrust	Bitsight	UpGuard
Australia	OneTrust	UpGuard	SecurityScorecard

OneTrust leads all five markets. The second position is UpGuard in the United States and Australia and Bitsight in Canada, Germany and India. The third position is Bitsight in the United States, UpGuard in Canada, Germany and India and SecurityScorecard in Australia. Every market's three leading brands are drawn from the same four brands, and no market returns a vendor absent from the others in its leading positions.

At product level, 27 of the 127 product entries were recorded in all five markets, and every entry in Table 5 was recorded in all five.

4.9 Inter-engine disagreement

The engines diverge on three specific points in the category.

The first concerns the primary mechanism for reducing questionnaire workload. ChatGPT returns AI-assisted evidence extraction, in the form Whistic offers, as the answer. Gemini more often returns native workflow orchestration inside ServiceNow. The engines therefore return different kinds of product for the same question, and each names the vendor it opens with.

The second concerns the platform for a programme onboarding more than two hundred vendors a year. ChatGPT most often names Whistic or ProcessUnity first. Google AI Overview more often leads with UpGuard or Bitsight, on the strength of combining security ratings with questionnaire workflows.

The third concerns vendors that only one engine associates with the category. Microsoft Copilot alone returns 28 product entries, among them Exalate and ComplyScore. Gemini alone returns 19, Google AI Overview 14, ChatGPT 12, Google AI Mode 6 and Perplexity 2. The divergences occur on use-case recommendations and at the edge of the vendor set rather than on its core.

5. Discussion

5.1 RQ1 — Do the six engines converge on the same set of named vendors?

They converge on a broad core and diverge in a long tail. §4.2 shows 9 of the 72 recorded brand strings named by all six engines, and §4.3 shows 9 of 127 product entries reaching the same coverage. The core is wider than in other categories in this series: 9 brands in every engine, against a typical six or seven.

The tail is where the engines part. 38 of the 72 brand strings and 81 of the 127 product entries were named by exactly one engine, and §4.9 shows each engine holding a tail the others do not, with Microsoft Copilot's the longest at 28 entries. A plausible explanation is that the core set is reachable from almost any part of the category's web presence. The tail depends on which sources a particular engine happens to read and on where that engine draws the category boundary. §4.6 is consistent with that: the engines reach the same core from evidence bases that overlap by a mean Jaccard coefficient of 0.19.

5.2 RQ2 — How concentrated is brand visibility across the recorded vendor set?

Less concentrated at the top than in other categories in this series, and fragmented below it. §4.2 shows the two leading brands taking 24.7% of all mentions and the five leading brands 54.1%, against 72 brand strings recorded. §4.3 shows 102 of 127 product entries below 10,000 mentions and 61 at exactly 1,000.

The shape is a core of 9 vendors the engines reuse, then a tail of dozens each engine names once or twice. The data is consistent with a threshold effect rather than a gradient: a vendor is either inside the group the engines reuse across queries and markets, or it is in a tail where mention counts fall away quickly. The measurement does not establish what places a vendor on one side of that boundary.

5.3 RQ3 — Do the engines draw on a shared evidence base?

No. §4.6 gives a mean pairwise Jaccard coefficient of 0.19 over cited domains, with a maximum of 0.40 and a minimum of 0.09. The three highest values are the three same-parent pairs, and no pair outside that group exceeds 0.22. §4.4 adds that no two engines share a most-cited domain. Six engines returned substantially the same core vendors while reading substantially different sets of domains.

Volume compounds the divergence. ChatGPT and Google AI Overview together contribute 65.5% of all cited sources, so a source list that serves one of them describes little of what the other four read. A practical consequence follows for measurement rather than for the vendors: a visibility programme built against one engine's source list transfers poorly to the others.

5.4 RQ4 — Does ordinal position track mention volume?

No. §4.7 shows OneTrust recording the highest mention total and the first-mention position in none of the six engines, while Bitsight, third on mentions, opens three. UpGuard records the earlier mean position than OneTrust in four engines and Bitsight in four. In Google AI Mode the brand named most often across the corpus averages 3.4 while UpGuard averages 1.7.

The three measures answer different questions. Mention volume records how often a vendor enters the response; first-mention position records which vendor opens it; mean ordinal position records where a vendor lands on average once named. A vendor can lead the first and hold neither of the others, which is OneTrust's position in this corpus. §6.1 sets out why ordinal position should be read as a proxy for prominence rather than as a measure of it. A visibility measurement that reports only one of the three is incomplete.

5.5 RQ5 — Does the vendor set vary between the five markets?

The set does not, and neither does the leader; the second and third positions do. §4.8 shows OneTrust leading all five markets, UpGuard in the United States and Australia and Bitsight in Canada, Germany and India in second place, and a third position that differs by market within the same leading group. 27 of 127 product entries were recorded in all five markets, and every entry in Table 5 was.

The result is a negative one for membership and is useful as such. It removes market localisation as an explanation for a vendor's absence from the returned set: a vendor missing in one market is, in this corpus, missing in all of them. The study covered five markets in one window and cannot establish that the same holds for markets or languages it did not test.

5.6 RQ6 — Do the engines cite vendors' own sites or third-party sources?

Vendors' own sites carry the largest single share, but not a majority for most engines. §4.4 shows 52.2% of all cited sources on a vendor site, with vendor sites the largest class for only two of the six engines. ChatGPT draws 72.3% of its citations from vendor domains; Google AI Mode draws 36.5%, the least. Analyst and review sites are the third class, at 8.6% of the corpus. The residual class is large for four engines, and Microsoft Copilot's most-cited domain sits in it.

The data is consistent with two evidence strategies behind one vendor list. ChatGPT and Perplexity read mostly the vendors' own documentation and product pages; the other four read a broader and less classifiable web, supplemented by analyst pages. That the six arrive at the same core set from such different mixes is the strongest indication in the study that the core set is not an artefact of any one kind of source. What the measurement does not establish is whether a vendor's own site earns citations because the vendor is already in the set, or the reverse.

5.7 What these results do not resolve

The study measures outputs, not mechanisms. It cannot say why a vendor is inside the core set or what would move one into it, because it observes no vendor changing position. It records that the engines cite largely separate domains but not whether a citation influenced the vendor list, since citation and recommendation were recorded from the

same response and cannot be ordered causally. It classifies cited domains but does not read the cited pages, so it cannot say what on a vendor site was cited.

Two further limits apply to specific metrics. Truncation is inferred from the terminal characters of a response and cannot distinguish a response that ended mid-sentence from one whose final sentence was recorded without terminal punctuation; ChatGPT's 91% should be read with that in mind. Sentiment is a model's reading of the response under a rubric, and the 0.06 spread across the leading brands is too narrow to support a distinction between any two of them.

6. Threats to Validity

6.1 Construct validity

Mean ordinal position is used as a proxy for prominence. It records where a brand name falls in the sequence of vendors a response names, which is not the same as how prominently a reader perceives it. A vendor named first in a list and then qualified, and a vendor named third and then recommended, receive positions that do not reflect that difference. The metric is defensible for comparison between engines over the same corpus and should not be read as a measure of persuasive weight. The first-mention vendor is a mode, not a mean, and §4.7 shows the two disagreeing in one engine.

Sentiment is assigned by the analysis model under a documented rubric (§3.7). The values in §4.2 occupy a narrow band, which is consistent either with the engines describing these vendors in similar terms or with the rubric having limited resolution over this kind of text. The study cannot distinguish the two, and no claim in this report rests on a difference between two sentiment values.

Source type is assigned from the domain, not from the page. A vendor's blog and its product documentation both count as vendor site; an analyst firm's page counts as a review aggregator. The classes are coarse by design, and the residual other class is large for four engines. Truncation is an inference from the terminal characters of a response, not an observation of a process. Hedging and recency anchoring rest on an undocumented lexicon (§3.7) and are subject to the same limit. The same detectors were applied to all six engines, so the values are comparable within this corpus. They are not comparable to figures produced by another detector.

6.2 Internal validity

All responses were collected in one window in August 2026. The vendor sets and the cited domains describe that window; reachability describes the date on which the pages were tested. A collection window that included a major vendor acquisition or an analyst publication would be expected to produce different results, and the study has no second window against which to test that.

The engines are non-deterministic. The same query issued twice to the same engine can return different text, a different vendor ordering, and a different citation list. The design mitigates this through corpus size and through paraphrase variants within the query set, which spread the measurement over many samples of each engine's behaviour rather than relying on single responses. It does not eliminate it.

Personalisation and session state were not controlled beyond issuing each query in a defined market. Where an engine adapts to prior interaction, that adaptation is not observable from the response text, and its contribution to the results cannot be separated. No engine was version-pinned, because none of the six exposes a version identifier to the

caller (§3.2), so a change made by an operator during the window would be indistinguishable from variation in the responses.

6.3 External validity

The study covers one category, five markets and one collection window. Enterprise third-party risk management has a broad core and a long tail (§1.2, §4.2). The core is what makes it a usable instrument; the tail is what limits generalisation, since a category with a different vendor population would not be expected to produce the same shape.

Query sampling is purposive, not random. The query set was written to elicit buyer-intent vendor recommendations, so it samples a deliberately chosen region of the space of things a buyer might ask rather than a representative draw from it. Combined with the non-determinism in §6.2 and the absence of version pinning, this precludes statistical inference from these results. All figures in this report are descriptive statistics over the recorded corpus, and no significance claim, interval or test is reported anywhere in it.

6.4 Reliability

A repeat run in a later window would be expected to reproduce the structural results. Those are the concentration of visibility into a core with a long tail, the separation between mention volume and ordinal position, the elevated overlap between same-parent engines, the split between vendor-led and residual-led evidence bases, and the stability of the leading group across markets. All rest on large differences that ordinary variation would not close.

Point values would not be expected to reproduce. Individual mention counts, per-engine cited-source totals, the composition of the tail, the identity of each engine's first-mention vendor where the margin is small, and the specific most-cited domain for an engine depend on what each engine read during the window. Dead-link rates would be expected to move, because reachability was tested once (§3.7) and the reachability of the cited web changes independently of the engines.

7. Limitations

- One category. The results describe enterprise third-party risk management and do not transfer to other security categories without measurement.
- Five markets, all tested in one window in August 2026. Markets and languages outside the five were not tested.
- No claim an engine made was adjudicated. Where two engines returned incompatible statements (§4.9), the study records the disagreement and does not determine which is correct.
- Source type was assigned from the domain and the residual other class is large for four engines; the study says which kinds of site the engines cite, not what on those sites was cited.
- Source age is not reported, because fewer than a fifth of cited pages declared a publication date.

8. Practical Implications

This section is derived from the results rather than measured by them.

Measure who opens the response, not only who is in it. §4.7 shows the most-mentioned brand opening the response in no engine and averaging the third or fourth position in every one, while the third-placed brand opens three. Track mention volume, first-mention share and mean position as three numbers, and decide from them whether the problem is inclusion, opening position or average placement. The three do not respond to the same work.

Treat the vendor's own site as the primary citation surface, and expect it to carry less than half of what most engines read. §4.4 shows 52.2% of all citations on vendor sites, a majority only in ChatGPT and Perplexity, with 37.1% on domains outside every class the study tracks. Documentation, product and integration pages on the vendor's own domain are what the two most vendor-led engines read; the other four read a broader web that the vendor does not control.

Instrument every engine separately, and at least three from different vendor families. §4.6 gives a mean pairwise cited-domain overlap of 0.19, no cross-family pair above 0.22, and no two engines sharing a most-cited domain. A source list that serves ChatGPT describes little of what Microsoft Copilot or the Google surfaces read.

Name products consistently, and expect several entries per vendor regardless. §4.3 records 127 product entries, with OneTrust holding 9 and ServiceNow 8, and it shows the most-mentioned brand falling to third at product level for that reason. Some of that is real product breadth and some is naming drift; only the vendor can tell which, and only consistent naming across published material lets a measurement consolidate it.

Do not build a market-by-market visibility programme for these five markets. §4.8 shows the same brand leading all five, with only the second and third positions changing. Effort that would go into market-specific work is better spent on the inclusion problem, which is the same problem in every market tested.

Do not treat link decay as a lever in this category. §4.5 records an overall dead-link rate of 0.2% and no engine above 0.6%. Do not treat sentiment as one either: the leading brands with a recorded score sit within a 0.06 band (§4.2).

9. Conclusion

This study put 10,000 enterprise buyer-intent queries to six AI search engines in five markets, producing 300,000 responses containing 1,551,000 cited sources and covering 127 distinct product entries in enterprise third-party risk management.

The engines agree about a broad core and disagree about evidence. 9 of 72 brand strings were named by all six engines, led by OneTrust at 205,000, UpGuard at 165,000 and Bitsight at 161,000, with 38 strings named by one engine only. Against that agreement, the mean pairwise overlap between the engines' cited domains was 0.19, the highest value anywhere in the matrix was 0.40, between two engines operated by the same parent, and no two engines shared a most-cited domain. Vendor sites carry 52.2% of everything the engines cite, a majority in two engines and a minority in four.

Two further results qualify that list. Ordinal position does not track mention volume: OneTrust takes the first-mention position in no engine, while Bitsight takes it in three, and OneTrust averages the third or fourth position everywhere. Source reliability is high: 0.2% of cited pages were unreachable, and no engine exceeded 0.6%. OneTrust led every market, with only the positions behind it changing.

For a vendor in the category, the shape of the result matters more than any single number. Visibility here is concentrated into a core the engines reuse across markets, phrasings and engines, reached through six largely separate evidence bases. Entering that core is one problem; opening the response once inside it is a different one,

which the most-mentioned vendor in the category has not solved in any engine; and both have to be solved once per engine.

10. Reproducibility and Data Availability

Responses were collected through each engine's public interface using the study's own instrumentation and retained as normalised plain text. Retained alongside them are the extracted citation lists with source type and reachability results, and the brand and product mention records with their ordinal positions. The dataset for this report was generated in August 2026 and processed under the instrumentation's current citation and entity rules.

Reproducing the study would require the query set, the market configuration, the vendor name list used for mention detection and the reachability test, applied to the same six engines within a comparable collection window. Exact reproduction is not achievable, because the engines are commercial services under continuous change and none is version-pinned; §6.4 states what a repeat run would and would not be expected to reproduce.

Requests for the dataset supporting the tables in this report may be directed to GrackerAI.

Disclosure

GrackerAI produces this research and also sells AI search visibility measurement for the category it covers. The study measured how six AI search engines describe vendors in enterprise third-party risk management. It did not test, evaluate or verify any vendor's product, and no statement in this report is an assessment of product quality. No vendor named in this report paid for, requested, reviewed or was given advance sight of any placement, ordering or finding in it.

How to cite this report

GrackerAI. (August 2026). *AI Search Visibility in Enterprise Third-Party Risk Management: A Multi-Engine, Multi-Market Benchmark Study*. GrackerAI AI Search Visibility Benchmark Series. https://gracker.ai/data-and-research-reports

Appendix A. Brand visibility by AI engine

Table A1. Brand mentions and mean ordinal position by AI engine — all 72 brand strings recorded; mean ordinal position in parentheses, a dash where the engine recorded no mention or no position; n = 300,000 responses

Brand	Total	ChatGPT	Microsoft Copilot	Gemini	Google AI Mode	Google AI Overview	Perplexity
OneTrust	205,000	43,000 (3.2)	38,000 (3.2)	42,000 (3.6)	43,000 (3.4)	35,000 (3.3)	4,000 (2.8)
UpGuard	165,000	21,000 (3.1)	35,000 (3.2)	31,000 (2.5)	35,000 (1.7)	38,000 (2.6)	5,000 (4.8)
Bitsight	161,000	18,000 (2.8)	38,000 (3.3)	28,000 (2.5)	36,000 (2.5)	36,000 (2.7)	5,000 (3.3)
SecurityScorecard	146,000	31,000 (4.5)	34,000 (5.0)	22,000 (2.7)	32,000 (3.1)	24,000 (3.5)	3,000 (6.5)
ProcessUnity	132,000	36,000 (2.6)	17,000 (5.3)	24,000 (5.5)	29,000 (4.8)	22,000 (3.9)	4,000 (6.3)

Brand	Total	ChatGPT	Microsoft Copilot	Gemini	Google AI Mode	Google AI Overview	Perplexity
Panorays	104,000	10,000 (5.3)	24,000 (3.7)	29,000 (3.5)	14,000 (4.1)	23,000 (3.8)	4,000 (5.3)
ServiceNow	102,000	18,000 (3.7)	29,000 (3.1)	21,000 (4.8)	17,000 (3.3)	15,000 (2.4)	2,000
Vanta	59,000	15,000 (2.7)	2,000 (4.0)	13,000 (6.3)	9,000 (4.6)	19,000 (2.4)	1,000 (7.0)
Whistic	51,000	20,000 (1.9)	13,000 (4.5)	5,000 (4.0)	4,000 (3.5)	9,000 (3.2)	—
MetricStream	34,000	6,000 (4.0)	6,000 (3.0)	7,000 (3.2)	7,000 (6.4)	8,000 (4.1)	—
Drata	29,000	12,000 (3.3)	2,000 (6.0)	11,000 (7.5)	1,000 (4.0)	3,000 (3.5)	—
Riskonnect	27,000	4,000 (8.0)	12,000 (3.8)	1,000 (7.0)	5,000 (3.6)	4,000 (3.0)	1,000 (4.0)
Venminder	24,000	—	16,000 (6.4)	—	5,000 (7.6)	2,000 (6.5)	1,000 (3.0)
Aravo	22,000	8,000 (4.0)	10,000 (5.7)	1,000 (4.0)	2,000 (5.0)	1,000 (7.0)	—
Black Kite	20,000	3,000	7,000 (4.0)	9,000 (6.5)	—	—	1,000
Diligent	17,000	5,000	—	3,000 (3.0)	8,000 (6.6)	—	1,000 (7.0)
Prevalent	16,000	11,000 (3.4)	—	—	3,000 (6.3)	2,000 (4.8)	—
Hyperproof	14,000	4,000 (5.3)	2,000 (4.5)	1,000 (3.0)	5,000 (7.0)	—	2,000 (8.0)
LogicGate	12,000	5,000 (8.0)	—	4,000 (2.5)	1,000 (3.0)	2,000 (7.5)	—
VISO TRUST	12,000	1,000 (4.0)	—	4,000 (2.8)	2,000 (3.5)	4,000 (3.3)	1,000 (6.0)
SpyCloud	10,000	—	—	10,000 (8.5)	—	—	—
ComplyScore	9,000	—	9,000 (3.2)	—	—	—	—
Jira	9,000	1,000	1,000	3,000	1,000	3,000	—
Nudge Security	9,000	—	9,000 (8.0)	—	—	—	—
Scrut Automation	9,000	—	—	—	—	9,000 (5.0)	—
Coupa	8,000	2,000	3,000 (7.5)	—	2,000 (7.0)	1,000	—
SAP	6,000	1,000	1,000 (7.0)	1,000 (8.0)	2,000 (9.0)	1,000	—
Enlighta	5,000	4,000 (2.0)	—	—	—	—	1,000 (8.0)
Exalate	5,000	—	5,000 (2.0)	—	—	—	—
Prewave	4,000	1,000	2,000 (4.5)	1,000 (7.0)	—	—	—
Conveyor	3,000	1,000	—	1,000 (7.0)	—	—	1,000 (8.0)
Inference Systems	3,000	—	3,000 (4.7)	—	—	—	—
Sphera	3,000	1,000	2,000 (7.5)	—	—	—	—
ZigiOps	3,000	—	3,000 (3.0)	—	—	—	—
Archer	2,000	1,000	1,000 (2.8)	—	—	—	—
Certa	2,000	2,000	—	—	—	—	—
CheckFirst	2,000	—	2,000 (8.0)	—	—	—	—
Coverbase	2,000	—	—	—	—	2,000 (4.0)	—

Brand	Total	ChatGPT	Microsoft Copilot	Gemini	Google AI Mode	Google AI Overview	Perplexity
Craft	2,000	1,000	1,000 (9.0)	—	—	—	—
Cyber Sierra	2,000	—	2,000 (9.0)	—	—	—	—
Fairmarkit	2,000	—	2,000 (4.5)	—	—	—	—
GAN Integrity	2,000	2,000	—	—	—	—	—
IBM OpenPages	2,000	—	2,000 (9.0)	—	—	—	—
Interos	2,000	—	—	—	1,000 (8.0)	—	1,000
LogicManager	2,000	2,000	—	—	—	—	—
LSEG/Avetta	2,000	1,000	1,000 (10.0)	—	—	—	—
NAVEX	2,000	2,000	—	—	—	—	—
Onspring	2,000	2,000	—	—	—	—	—
RiskProfiler	2,000	1,000	—	—	—	—	1,000
SAI360	2,000	2,000	—	—	—	—	—
Secureframe	2,000	—	1,000 (10.0)	—	1,000 (6.0)	—	—
SecurityPal	2,000	—	—	—	—	1,000 (4.0)	1,000 (7.0)
VISOTRUST	2,000	—	—	2,000 (1.0)	—	—	—
VRM GRC	2,000	—	2,000 (5.0)	—	—	—	—
AuditBoard	1,000	—	—	—	—	1,000 (3.0)	—
Ciphrix	1,000	—	—	1,000 (3.0)	—	—	—
ConnectALL	1,000	—	1,000 (4.0)	—	—	—	—
CoreStream	1,000	1,000 (6.0)	—	—	—	—	—
CyberGRX	1,000	—	—	—	1,000	—	—
Exiger	1,000	—	—	1,000 (5.0)	—	—	—
HyperComply	1,000	1,000	—	—	—	—	—
OpsHub	1,000	—	1,000 (4.0)	—	—	—	—
Optro	1,000	1,000	—	—	—	—	—
Recorded Future	1,000	—	—	—	—	—	1,000
Resilinc	1,000	—	—	1,000 (6.0)	—	—	—
Risk Ledger	1,000	—	1,000	—	—	—	—
RiskImmune	1,000	—	—	—	—	—	1,000 (7.0)
Safe Security	1,000	—	—	—	—	1,000	—
Smarsh	1,000	—	—	—	—	—	1,000
TemplarShield	1,000	—	1,000 (2.0)	—	—	—	—
Vendict	1,000	—	1,000 (5.0)	—	—	—	—

Brand	Total	ChatGPT	Microsoft Copilot	Gemini	Google AI Mode	Google AI Overview	Perplexity
Workiva	1,000	—	1,000 (8.0)	—	—	—	—
Total	**1,496,000**	**301,000**	**343,000**	**277,000**	**266,000**	**266,000**	**43,000**

Appendix B. Most-cited domains and source types

Table B1. Three most-cited domains by AI engine — cited-source counts in parentheses; n = 1,551,000 cited sources

AI engine	Cited sources	Distinct domains	Most-cited domains
ChatGPT	567,000	65	whistic.com (69,000), servicenow.com (55,000), upguard.com (53,000)
Microsoft Copilot	213,000	23	zipdo.co (37,000), servicenow.com (30,000), upguard.com (27,000)
Gemini	119,000	34	spycloud.com (29,000), panorays.com (15,000), gartner.com (14,000)
Google AI Mode	170,000	52	gartner.com (17,000), bitsight.com (13,000), servicenow.com (12,000)
Google AI Overview	449,000	65	bitsight.com (49,000), panorays.com (36,000), spycloud.com (24,000)
Perplexity	33,000	23	upguard.com (4,000), bitsight.com (3,000), hyperproof.io (2,000)

Table B2. Cited sources by source type and AI engine — counts; n = 1,551,000 cited sources

AI engine	Vendor site	Other	Review aggregator	Social	Forum	Academic	Total
ChatGPT	410,000	89,000	66,000	0	0	2,000	567,000
Microsoft Copilot	85,000	118,000	10,000	0	0	0	213,000
Gemini	50,000	55,000	14,000	0	0	0	119,000
Google AI Mode	62,000	76,000	18,000	14,000	0	0	170,000
Google AI Overview	184,000	224,000	23,000	16,000	2,000	0	449,000
Perplexity	18,000	13,000	2,000	0	0	0	33,000
Total	**809,000**	**575,000**	**133,000**	**30,000**	**2,000**	**2,000**	**1,551,000**

Appendix C. Product entries

Table C1. All product entries — 127 entries ordered by mentions; a dash where no mean rank was recorded; n = 300,000 responses

#	Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
1	Bitsight	Bitsight	146,000	2.8	0.82	6	5
2	UpGuard	UpGuard	134,000	2.8	0.82	6	5
3	OneTrust	OneTrust	127,000	3.5	0.78	6	5
4	SecurityScorecard	SecurityScorecard	124,000	3.8	0.79	6	5
5	ProcessUnity	ProcessUnity	105,000	4.5	0.82	6	5
6	Panorays	Panorays	95,000	3.8	0.81	6	5
7	Prevalent	Prevalent	64,000	4.5	0.77	6	5
8	Vanta	Vanta	50,000	3.9	0.82	6	5
9	Whistic	Whistic	44,000	3.1	0.81	5	5
10	ServiceNow TPRM	ServiceNow	37,000	3.5	0.81	5	5
11	ServiceNow Third-Party Risk Management	ServiceNow	26,000	2.4	0.85	5	5
12	Drata	Drata	24,000	5.4	0.78	5	5
13	UpGuard Vendor Risk	UpGuard	23,000	1.7	0.84	5	5
14	ServiceNow Vendor Risk Management	ServiceNow	23,000	4.3	0.84	5	5
15	OneTrust Third-Party Risk Management	OneTrust	22,000	2.2	0.82	4	5
16	OneTrust Vendor Risk Management	OneTrust	22,000	3.5	0.80	4	5
17	MetricStream	MetricStream	22,000	4.5	0.81	5	4
18	Riskonnect	Riskonnect	21,000	4.0	0.77	6	5
19	Archer	Archer	20,000	5.3	0.72	5	5
20	Venminder	Venminder	20,000	6.5	0.73	4	5
21	Aravo	Aravo	19,000	5.1	0.75	4	5
22	OneTrust Third-Party Management	OneTrust	18,000	3.6	0.81	4	5
23	ProcessUnity Vendor Risk Management	ProcessUnity	13,000	3.1	0.80	4	5
24	Hyperproof	Hyperproof	12,000	6.1	0.78	5	3
25	Black Kite	Black Kite	11,000	5.8	0.78	2	5
26	VISO TRUST	VISO TRUST	9,000	3.6	0.84	5	5
27	Diligent	Diligent	8,000	6.6	0.81	2	5
28	RiskRecon	Mastercard	7,000	6.3	0.78	3	4
29	ComplyScore	ComplyScore	6,000	3.2	0.77	1	3

#	Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
30	Bitsight Vendor Risk Management	Bitsight	5,000	2.2	0.86	3	3
31	LogicGate Risk Cloud	LogicGate	5,000	2.6	0.82	2	5
32	Enlighta	Enlighta	5,000	3.2	0.86	2	3
33	Scrut Automation	Scrut Automation	5,000	4.6	0.84	1	5
34	OneTrust Third-Party Governance	OneTrust	4,000	2.0	0.85	2	3
35	Exalate	Exalate	4,000	2.0	0.75	1	3
36	Vanta Third-Party Risk Management	Vanta	4,000	2.5	0.82	2	3
37	ProcessUnity Third-Party Risk Management	ProcessUnity	4,000	3.5	0.85	3	3
38	ServiceNow IRM	ServiceNow	4,000	4.8	0.82	3	3
39	Nudge Security	Nudge Security	4,000	8.0	0.65	1	4
40	ZigiOps	ZigiOps	3,000	3.0	0.67	1	2
41	OneTrust Vendorpedia	OneTrust	3,000	4.0	0.83	2	3
42	MetricStream Third-Party Risk Management	MetricStream	3,000	4.3	0.80	1	3
43	Prevalent Third-Party Risk Management Platform	Prevalent	3,000	4.7	0.90	2	3
44	Prewave	Prewave	3,000	5.3	0.83	2	2
45	Scrut Automation Vendor Risk Management	Scrut Automation	3,000	5.7	0.83	1	3
46	LogicGate	LogicGate	3,000	7.7	0.73	2	2
47	SpyCloud Supply Chain Threat Protection	SpyCloud	3,000	8.7	0.87	1	3
48	SAP Ariba Supplier Risk	SAP	3,000	8.7	0.80	2	3
49	Vantavanta.com	Vanta	2,000	1.0	0.85	1	2
50	VISOTRUST	VISOTRUST	2,000	1.0	0.85	1	2
51	Dratadrata.com	Drata	2,000	2.0	0.80	1	2
52	Archer (RSA Archer)	RSA	2,000	2.0	0.60	1	2
53	OneTrust Third-Party Risk Exchange	OneTrust	2,000	2.5	0.90	2	2
54	Bitsight Third-Party Risk Management	Bitsight	2,000	2.5	0.80	1	2
55	MetricStream CyberGRC	MetricStream	2,000	3.0	0.90	2	2
56	Whisticwhistic.com	Whistic	2,000	3.0	0.85	1	2
57	3rdRisk	3rdRisk	2,000	3.5	0.70	2	2
58	OneTrustonetrust.com	OneTrust	2,000	4.0	0.65	1	2
59	Inference Systems Custom AI Workflow	Inference Systems	2,000	4.5	0.75	1	2
60	Fairmarkit	Fairmarkit	2,000	4.5	0.70	1	2

#	Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
61	Prevalent / Mitratech	Prevalent	2,000	4.5	0.60	1	1
62	SecurityPal	SecurityPal	2,000	5.5	0.75	2	1
63	Conveyor	Conveyor	2,000	7.5	0.85	2	2
64	Sphera	Sphera	2,000	7.5	0.75	1	2
65	Secureframe	Secureframe	2,000	8.0	0.70	2	2
66	Cyber Sierra	Cyber Sierra	2,000	9.0	0.80	1	1
67	VISO TRUST Third-Party Risk Management	VISO TRUST	1,000	1.0	0.90	1	1
68	ServiceNow Third-Party Risk Management (TPRM)	ServiceNow	1,000	1.0	0.80	1	1
69	MetricStream Third-Party Cyber Risk	MetricStream	1,000	2.0	0.85	1	1
70	TemplarShield Third-Party Risk	TemplarShield	1,000	2.0	0.80	1	1
71	Whistic AI-First TPRM	Whistic	1,000	2.0	0.80	1	1
72	BitSight Cyber Risk Intelligence	Bitsight	1,000	2.0	0.80	1	1
73	Bitsight Integrated Third-Party Risk Management	Bitsight	1,000	2.0	0.80	1	1
74	Riskonnect Third-Party Risk Management	Riskonnect	1,000	2.0	0.80	1	1
75	AuditBoard	AuditBoard	1,000	3.0	1.00	1	1
76	UpGuard Questionnaire Module	UpGuard	1,000	3.0	0.90	1	1
77	UpGuard Risk Automations	UpGuard	1,000	3.0	0.80	1	1
78	Bitsight Framework Intelligence	Bitsight	1,000	3.0	0.80	1	1
79	Ciphrix	Ciphrix	1,000	3.0	0.80	1	1
80	UpGuard Risk Automation	UpGuard	1,000	3.0	0.80	1	1
81	ProcessUnity Vendor Risk Assessments	ProcessUnity	1,000	3.0	0.70	1	1
82	SecurityScorecard Vendor Risk	SecurityScorecard	1,000	3.0	0.70	1	1
83	Diligent HighBond	Diligent	1,000	4.0	0.90	1	1
84	SecurityScorecard Atlas	SecurityScorecard	1,000	4.0	0.90	1	1
85	ServiceNow Supplier Lifecycle & Procurement Operations	ServiceNow	1,000	4.0	0.80	1	1
86	Archer Third-Party Risk Management	Archer	1,000	4.0	0.80	1	1
87	Viso Trust TPRM Automation	Viso Trust	1,000	4.0	0.80	1	1
88	VISO TRUST Third-Party Risk Management Automation	VISO TRUST	1,000	4.0	0.80	1	1
89	Whistic Vendor Risk Management	Whistic	1,000	4.0	0.75	1	1
90	Aravo Vendor Risk Management	Aravo	1,000	4.0	0.75	1	1

#	Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
91	Coupa Risk Assess	Coupa	1,000	4.0	0.70	1	1
92	OpsHub	OpsHub	1,000	4.0	0.70	1	1
93	ServiceNow Risk Assessments Integration for Procurement	ServiceNow	1,000	4.0	0.70	1	1
94	ConnectALL	ConnectALL	1,000	4.0	0.60	1	1
95	Coverbase	Coverbase	1,000	4.0	0.60	1	1
96	Exiger	Exiger	1,000	5.0	0.90	1	1
97	Drata Third-Party Risk Management	Drata	1,000	5.0	0.90	1	1
98	Inference Systems	Inference Systems	1,000	5.0	0.80	1	1
99	ServiceNow IRM / Vendor Risk Management	ServiceNow	1,000	5.0	0.80	1	1
100	Scrut	Scrut	1,000	5.0	0.80	1	1
101	VRM GRC	VRM GRC	1,000	5.0	0.80	1	1
102	Third-Party Risk Management for Jira	Atlassian	1,000	5.0	0.70	1	1
103	Whistic Questionnaire Exchange	Whistic	1,000	5.0	0.70	1	1
104	SecurityScorecardsecurityscorecard.com	SecurityScorecard	1,000	5.0	0.60	1	1
105	Vendict	Vendict	1,000	5.0	0.60	1	1
106	Coupa Supplier Risk	Coupa	1,000	6.0	0.80	1	1
107	Resilinc	Resilinc	1,000	6.0	0.80	1	1
108	Third Party Risk Management for Jira	Atlassian	1,000	6.0	0.80	1	1
109	Hyperproof Risk Management	Hyperproof	1,000	6.0	0.80	1	1
110	Panorays Third-Party Security	Panorays	1,000	6.0	0.70	1	1
111	Prevalentprevalent.ai	Prevalent	1,000	6.0	0.60	1	1
112	CoreStream GRC	CoreStream	1,000	6.0	0.60	1	1
113	Diligent Third-Party Risk Management	Diligent	1,000	7.0	0.80	1	1
114	SAP Business Network	SAP	1,000	7.0	0.70	1	1
115	RiskImmune	RiskImmune	1,000	7.0	0.70	1	1
116	OneTrust Third-Party Due Diligence	OneTrust	1,000	7.0	0.70	1	1
117	Coupa	Coupa	1,000	8.0	0.90	1	1
118	SpyCloud	SpyCloud	1,000	8.0	0.90	1	1
119	CheckFirst	CheckFirst	1,000	8.0	0.80	1	1
120	Workiva Supplier Risk Management	Workiva	1,000	8.0	0.80	1	1
121	Interos	Interos	1,000	8.0	0.80	1	1

#	Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
122	VRM GRC (TrustArc)	TrustArc	1,000	8.0	0.70	1	1
123	Craft	Craft	1,000	9.0	0.70	1	1
124	RiskProfiler (Knyx Vendor AI)	Knyx	1,000	10.0	0.80	1	1
125	IBM OpenPages	IBM	1,000	9.0	0.60	1	1
126	LSEG/Avetta	LSEG/Avetta	1,000	10.0	0.60	1	1
127	Coupa Risk	Coupa	1,000	11.0	0.70	1	1

Appendix D. Study parameters

Table D1. Study parameters — as configured for this study

Parameter	Value
Category	Enterprise third-party risk management (TPRM, VRM)
AI engines	ChatGPT, Microsoft Copilot, Gemini, Google AI Mode, Google AI Overview, Perplexity
Markets	United States, Canada, Germany, India, Australia
Query set	10,000 distinct buyer-intent queries, core questions plus paraphrase variants
Responses per engine	50,000
Responses per market	60,000
Total responses	300,000
Cited sources	1,551,000
Brand strings recorded	72
Product entries	127
Collection window	August 2026
Unit of analysis	One engine response to one query in one market
Design	Observational, cross-sectional, single collection window
Statistical treatment	Descriptive statistics only; no inference is reported

Appendix E. Metric computation

Table E1. Metric computation — the arithmetic behind each derived metric in this report

Metric	Computation
Responses per engine	Queries × markets.

Metric	Computation
Total responses	Queries × engines × markets.
Responses per market	Queries × engines.
Mean response length	Sum of response word counts for an engine, divided by that engine's responses.
Hedging, recency-anchored, truncation and refusal rates	Responses of an engine carrying the characteristic, divided by that engine's responses.
Total cited sources	Sum of the per-engine cited-source counts.
Share of corpus	That engine's cited sources divided by total cited sources.
Source-type share	That engine's cited sources of the type, divided by that engine's cited sources.
Dead links per engine	That engine's cited sources × its dead-link rate.
Total dead links	Sum of the per-engine dead-link counts.
Overall dead-link rate	Total dead links divided by total cited sources.
Total brand mentions	Sum of a brand's per-engine mention counts, summed over brands for the table total.
Engines recording	Count of engines whose mention count for that brand is greater than zero.
Mean ordinal position	Sum of the positions at which a brand was named, divided by the mentions carrying a position.
Mean sentiment	Sum of the per-mention sentiment scores for a brand, divided by that brand's mentions.
Jaccard coefficient	Domains cited by both engines, divided by domains cited by either engine.
Mean rank	Sum of a product entry's ordered positions, divided by the mentions carrying a position.
Leading brands by market	Brands ranked by mention count within the responses collected for that market.