Gracker.ai

RESEARCH REPORT

AI Search Visibility in Enterprise Password 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 password management, credential vaulting and workforce secrets across the United States, Canada, India and Germany.

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

AI engine	Responses	Cited sources	Markets
ChatGPT	40,000	2,730,000	4
Microsoft Copilot	40,000	171,000	4
Gemini	40,000	147,000	4
Google AI Mode	40,000	691,000	4
Google AI Overview	40,000	471,000	4
Perplexity	40,000	382,000	4
Total	**240,000**	**4,592,000**	**4**

Published: August 2026

Research period: August 2026

Category: Enterprise password management, credential vaulting and workforce secrets

Scope: 10,000 buyer-intent queries · 240,000 AI engine responses · 6 AI engines · 4 markets

Evidence base: 4,592,000 cited sources · 31 distinct product entries

Prepared by: GrackerAI · AI Search Visibility Benchmark Series

Abstract

Enterprise buyers now begin a password-management evaluation 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 password management. Each engine received 10,000 enterprise buyer-intent queries in each of four markets during a single collection window in August 2026, producing 240,000 responses that contained 4,592,000 cited sources and covered 31 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 narrow vendor set: 7 of 33 brand strings were named by all six engines. The three leading brands took 61.5% of all mentions, led by Bitwarden at 235,000, Keeper at 218,000 and 1Password at 214,000. The brand named most often is not the brand named first: 1Password holds the first-mention position in 4 of the six engines, and Keeper holds it in none. The engines cite largely separate domains, with a mean pairwise overlap of 0.14, yet 58.9% of everything they cite sits on a vendor's own site. Dead links are rare, at 0.9% of cited pages, and the same three brands lead all four markets in the same order.

Contents

1. Introduction
- 1.1 Background
- 1.2 Why this category
- 1.3 Contribution

2. Research Questions

3. Methodology
- 3.1 Study design
- 3.2 Engines under test
- 3.3 Markets
- 3.4 Query construction
- 3.5 Corpus and unit of analysis
- 3.6 Metric definitions
- 3.7 Data collection and processing
- 3.8 Scope exclusions

4. Results
- 4.1 Response characteristics by engine
- 4.2 Brand visibility
- 4.3 Product-level results
- 4.4 Citation volume and source type
- 4.5 Source quality and link decay
- 4.6 Cross-engine source overlap
- 4.7 First-mention position
- 4.8 Geographic variance
- 4.9 Inter-engine disagreement

5. Discussion

6. Threats to Validity

7. Limitations

8. Practical Implications

9. Conclusion

10. Reproducibility and Data Availability

Disclosure

How to cite this report

Appendix A. Brand visibility by AI engine

Appendix B. Most-cited domains and source types

Appendix C. Product entries

Appendix D. Study parameters

Appendix E. Metric computation

1. Introduction

1.1 Background

An enterprise buyer selecting a password manager now begins inside a generated response. The buyer asks an AI search engine which products support single sign-on, directory provisioning and audit reporting for a workforce of several hundred people. 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 six 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.

The measurement problem that follows is that no single engine describes the market. Engines differ in the sources they read, the length at which they answer, and the order in which they name vendors. A vendor that measures its position in one engine learns about that engine. This study measures six at once, in four markets, over one collection window, so that agreement and divergence between them can be separated from the behaviour of any one system.

1.2 Why this category

Enterprise password management is a suitable instrument for this measurement. The vendor set is mature and bounded: the engines in this study named a small core of established vaulting vendors, and 7 of the 33 brand strings recorded were named by every engine. A category with a settled vendor population isolates engine behaviour from genuine market churn.

The evaluation criteria are stable and technical. Across the corpus the engines returned the same selection factors: SOC 2 attestation, HIPAA alignment, role-based access control, SCIM provisioning, SAML single sign-on, audit logging and passkey support. To those they add the choice between a self-hosted deployment and a vendor-hosted one. These criteria are specific enough that responses can be compared to each other rather than to a general description of security software.

Buying behaviour in the category is research-heavy and the cost of invisibility is high. The product is chosen by security and IT teams who shortlist before contacting vendors, and the chosen vault holds every other credential the organisation owns. A vendor omitted from the generated shortlist loses the opportunity at the earliest stage of the cycle, 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 four 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 240,000 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 the brand named most often hold the first-mention position?
5. **RQ5.** Does the vendor set vary between the four 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 four 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.

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 and the first-mention results in §4.7.

3.3 Markets

Four markets were tested: United States, Canada, India and Germany. Market variance was tested because engines localise results, and a vendor set returned in one market cannot be assumed to hold in another. The four combine two large English-language markets, 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.

Every query was issued to every engine in every market. The corpus identity follows directly: 10,000 queries in 4 markets gives 40,000 responses per engine, and 10,000 queries across 6 engines and 4 markets gives 240,000 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 240,000 responses, distributed evenly across the six engines at 40,000 responses each and across the four 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

Metric	Definition	Unit
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
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, 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 and review aggregators were matched against lists and host-name patterns, and the remainder is other.

Reachability was tested once per page during the collection window. 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 deployment 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. 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 = 40,000 responses per engine, 240,000 in total

AI engine	Responses	Mean words	Hedging	Recency-anchored	Truncation	Refusal	First-mention vendor
ChatGPT	40,000	492.9	17%	45%	23%	0%	1Password
Microsoft Copilot	40,000	448.9	3%	63%	0%	0%	1Password
Gemini	40,000	419.7	3%	0%	38%	0%	1Password
Google AI Mode	40,000	292.3	5%	0%	7%	0%	Bitwarden
Google AI Overview	40,000	200.9	0%	13%	0%	0%	Bitwarden
Perplexity	40,000	97.5	28%	7%	42%	0%	1Password

Mean response length spans a factor of 5.1, from 97.5 words for Perplexity to 492.9 for ChatGPT. Figure 1 shows the distribution. Three engines exceed 400 words — ChatGPT, Microsoft Copilot and Gemini — and the two Google search surfaces and Perplexity fall below 300.

Truncation was recorded in 4 of the six engines, ranging from 7% for Google AI Mode to 42% for Perplexity, with Gemini at 38%. Microsoft Copilot and Google AI Overview recorded none. Recency anchoring is highest for Microsoft Copilot at 63%, and Gemini and Google AI Mode recorded none. Hedging is highest for Perplexity at 28%, followed by ChatGPT at 17%; Google AI Overview recorded none. No engine refused a query: the refusal rate is 0% for all six. Four engines record 1Password as the first-mention vendor and the two Google search surfaces record Bitwarden; §4.7 reports the position data behind that column.

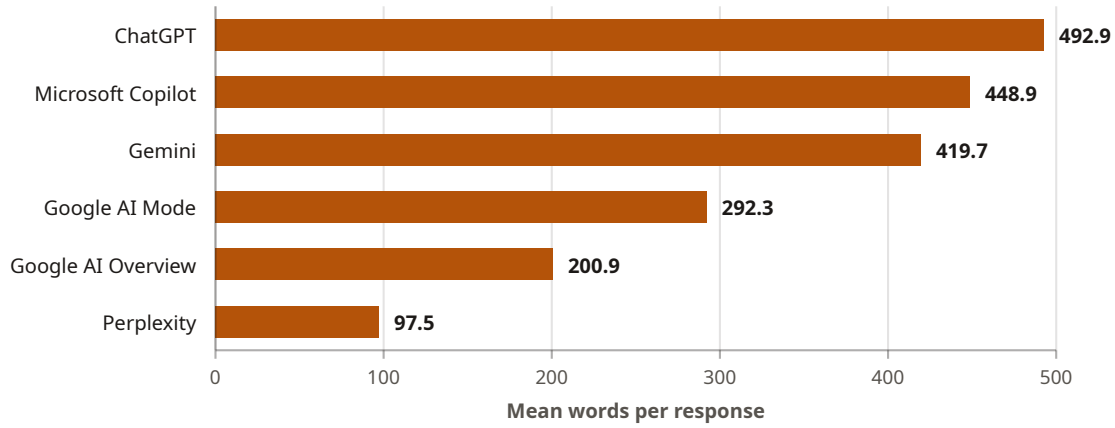

Figure 1. Mean response length by AI engine. Arithmetic mean word count per response; n = 40,000 responses per engine. Source: GrackerAI AI Search Visibility Benchmark Series, August 2026.

4.2 Brand visibility

Brand visibility was recorded for 33 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 33 brand strings recorded; mean sentiment for these twelve; n = 240,000 responses

Brand	Total mentions	Engines recording	Mean sentiment
Bitwarden	235,000	6	0.82
Keeper	218,000	6	0.68
1Password	214,000	6	0.73
Dashlane	139,000	6	0.58
NordPass	87,000	6	0.61
LastPass	29,000	5	0.12
Psono	28,000	6	0.62
Passbolt	25,000	6	0.59
CyberArk	16,000	2	0.61
Proton Pass	14,000	3	0.59
Microsoft	12,000	3	0.57
Vaultwarden	11,000	5	-0.23

Visibility is concentrated. The three most-mentioned brands, Bitwarden, Keeper and 1Password, account for 61.5% of all mentions recorded across the 33 brand strings, at 235,000, 218,000 and 214,000 mentions. Keeper and 1Password are separated by 4,000 mentions. Dashlane follows at 139,000 and NordPass at 87,000; below NordPass no brand exceeds 29,000.

7 of the 33 brand strings were named by all six engines: Bitwarden, Keeper, 1Password, Dashlane, NordPass, Psono and Passbolt. 14 were named by exactly one engine, none of them above 4,000 mentions. The strings recorded include vendors from adjacent categories — identity services, secrets managers and consumer credential stores — and one analyst firm, each named in a small number of responses as the engines wrote them. CyberArk was named by two engines, and 15,000 of its 16,000 mentions were recorded by Microsoft Copilot. Mean sentiment for the twelve brands in Table 4 runs from -0.23 for Vaultwarden and 0.12 for LastPass to 0.82 for Bitwarden; the other ten fall between 0.57 and 0.73.

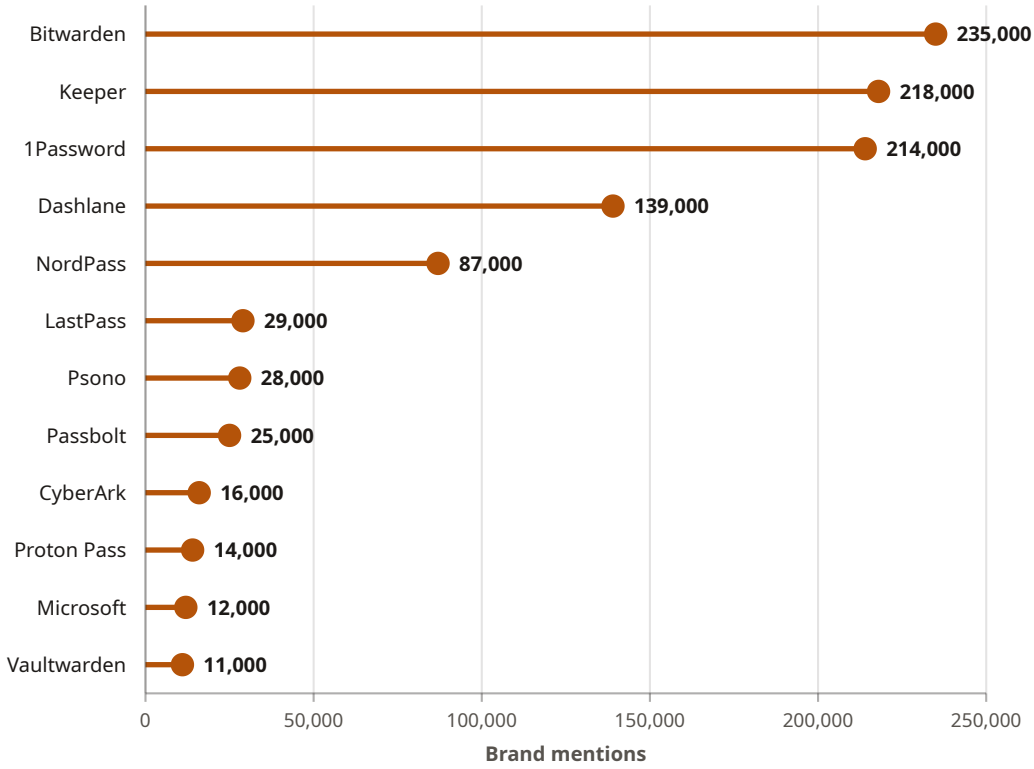

Figure 2. Brand mentions, twelve most-mentioned brands. The 12 most-mentioned of 33 brand strings recorded; n = 240,000 responses. Source: GrackerAI AI Search Visibility Benchmark Series, August 2026.

4.3 Product-level results

31 distinct product entries were recorded after plan, edition and deployment spellings of one product were merged. Table 5 gives the ten most-mentioned entries.

Table 5. Product entries, ten most-mentioned — of 31 entries; n = 240,000 responses

Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
Bitwarden Enterprise	Bitwarden	231,000	1.9	0.83	6	4
Keeper Enterprise	Keeper	202,000	2.6	0.73	6	4
1Password Business	1Password	197,000	1.6	0.80	6	4
Dashlane Business	Dashlane	124,000	3.9	0.64	6	4
NordPass Business	NordPass	79,000	4.5	0.64	6	4
Psono	Psono	27,000	2.7	0.66	6	4

Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
Passbolt	Passbolt	24,000	2.6	0.64	6	4
LastPass	LastPass	19,000	4.3	0.46	5	4
CyberArk	CyberArk	15,000	2.7	0.59	2	4
Proton Pass Business	Proton	13,000	4.5	0.62	3	4

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. 21 of the 31 entries record fewer than 10,000 mentions, 11 record exactly 1,000, and 14 were named by exactly one engine. 7 entries were named by all six engines, 12 were recorded in all four markets, and 11 were recorded in a single market. The full table is in Appendix C.

Mean rank among the ten most-mentioned entries runs from 1.6 to 4.5. The leading entry, Bitwarden Enterprise, records 231,000 mentions at a mean rank of 1.9. Keeper Enterprise records 202,000 at a mean rank of 2.6, and 1Password Business records 197,000 at 1.6, the earliest mean rank of the ten.

4.4 Citation volume and source type

The corpus contains 4,592,000 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 = 4,592,000 cited sources

AI engine	Cited sources	Share of corpus	Distinct domains	Most-cited domain
ChatGPT	2,730,000	59.5%	124	bitwarden.com (507,000)
Microsoft Copilot	171,000	3.7%	36	worldmetrics.org (16,000)
Gemini	147,000	3.2%	34	bitwarden.com (38,000)
Google AI Mode	691,000	15.0%	167	bitwarden.com (104,000)
Google AI Overview	471,000	10.3%	103	bitwarden.com (82,000)
Perplexity	382,000	8.3%	32	bitwarden.com (40,000)
Total	**4,592,000**	**—**	**—**	**—**

ChatGPT accounts for 59.5% of all cited sources in the corpus, 2,730,000 of 4,592,000. Google AI Mode cites from the widest domain population at 167 distinct domains, followed by ChatGPT at 124 and Google AI Overview at 103. Perplexity cites from the narrowest at 32, with Gemini at 34 and Microsoft Copilot at 36. Distinct-domain counts are reported per engine and are not summed across engines.

The most-cited domain is bitwarden.com for 5 of the six engines, at 507,000 citations in ChatGPT alone. Microsoft Copilot is the exception; its most-cited domain is worldmetrics.org at 16,000. Appendix B gives the three most-cited domains for each engine.

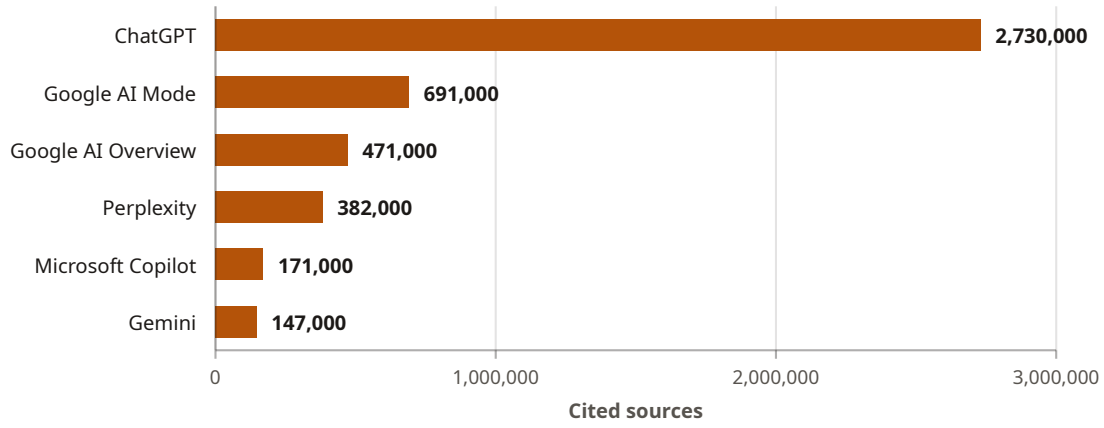

Figure 3. Cited sources by AI engine. Sources cited per engine; n = 4,592,000 cited sources. Source: GrackerAI AI Search Visibility Benchmark Series, August 2026.

Table 7. Source type by AI engine — share of each engine's cited sources by the class of the cited domain; n = 4,592,000 cited sources

AI engine	Vendor site	Other	Review aggregator	News	Forum	Social	Government
ChatGPT	78.8%	16.6%	1.8%	1.2%	1.2%	0.3%	0.1%
Microsoft Copilot	9.9%	77.8%	7.0%	5.3%	0.0%	0.0%	0.0%
Gemini	42.2%	24.5%	15.0%	17.7%	0.7%	0.0%	0.0%
Google AI Mode	34.6%	36.8%	8.5%	7.7%	5.8%	6.7%	0.0%
Google AI Overview	35.9%	36.3%	11.3%	4.5%	6.8%	5.3%	0.0%
Perplexity	18.1%	46.1%	16.2%	16.0%	3.7%	0.0%	0.0%
All engines	**58.9%**	**26.6%**	**5.6%**	**4.4%**	**2.6%**	**1.7%**	**0.1%**

Across the corpus, 58.9% of cited sources sit on a vendor site, 2,706,000 of 4,592,000, and 26.6% fall outside every named class. The mix differs by engine. ChatGPT draws 78.8% of its citations from vendor sites, the highest share, and Microsoft Copilot 9.9%, the lowest, with 77.8% of its citations in the other class. Vendor sites are the largest class for 2 of the six engines, ChatGPT and Gemini; for the other four the largest class is other. Review aggregators reach 16.2% of Perplexity's citations and 15.0% of Gemini's. Government domains account for 3,000 citations in the corpus, all recorded by ChatGPT; no academic or encyclopaedia domain was recorded by any engine. Figure 4 shows the vendor-site share by engine, and Appendix B gives the counts behind Table 7.

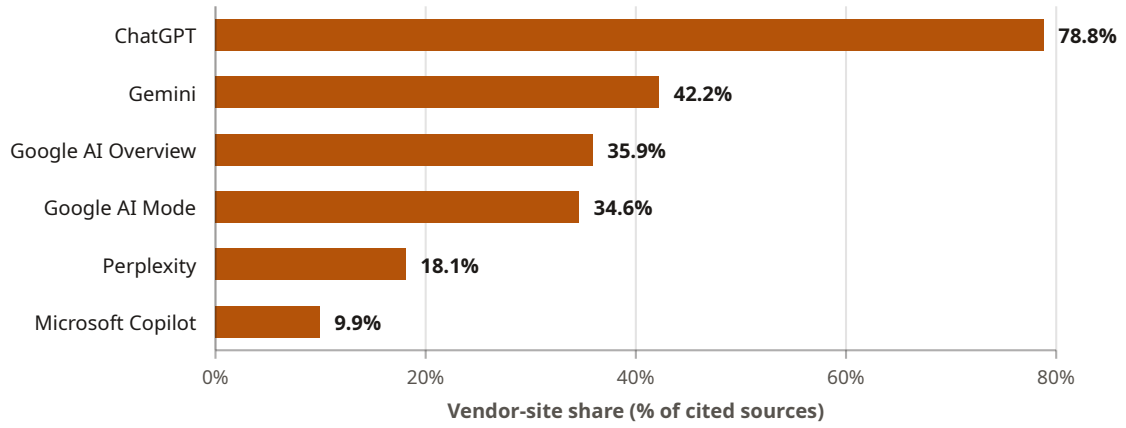

Figure 4. Share of cited sources on vendor sites by AI engine. Cited sources whose domain belongs to a vendor named in the response, as a share of that engine's cited sources; n = 4,592,000 cited sources. Source: GrackerAI AI Search Visibility Benchmark Series, August 2026.

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 = 4,592,000 cited sources

AI engine	Cited sources	Dead-link rate	Dead links	Self-citation rate
ChatGPT	2,730,000	0.7%	19,100	0%
Microsoft Copilot	171,000	1.2%	2,050	0%
Gemini	147,000	0%	0	0%
Google AI Mode	691,000	0.7%	4,840	0%
Google AI Overview	471,000	1.1%	5,180	0%
Perplexity	382,000	2.1%	8,020	0%
Total	**4,592,000**	**0.9%**	**39,190**	**—**

The overall dead-link rate is 0.9%, 39,190 of 4,592,000 cited pages. Per-engine rates run from 0% for Gemini to 2.1% for Perplexity, and no engine exceeds 2.1%. ChatGPT records 0.7%, and because it contributes 2,730,000 of the 4,592,000 cited sources it still accounts for 19,100 of the 39,190 dead links, 48.7% of the total.

The self-citation rate is 0% for all six engines, including the three operated by the same parent as one of the search properties they could cite.

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 = 4,592,000 cited sources

	ChatGPT	Copilot	Gemini	Google AI Mode	Google AI Overview	Perplexity
ChatGPT	**1.00**	0.07	0.16	0.19	0.19	0.13
Copilot	0.07	**1.00**	0.04	0.03	0.04	0.08
Gemini	0.16	0.04	**1.00**	0.17	0.22	0.27
Google AI Mode	0.19	0.03	0.17	**1.00**	0.26	0.11
Google AI Overview	0.19	0.04	0.22	0.26	**1.00**	0.14
Perplexity	0.13	0.08	0.27	0.11	0.14	**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 0.14. The highest value is 0.27, between Gemini and Perplexity. The lowest is 0.03, between Microsoft Copilot and Google AI Mode. The three engines operated by the same parent — Gemini, Google AI Mode and Google AI Overview — record coefficients of 0.17, 0.22 and 0.26 with each other. Only one of those three pairs is among the three highest values in the matrix.

Microsoft Copilot records the lowest coefficient with every other engine: its five values run from 0.03 to 0.08, the latter with Perplexity. Only 3 of the 15 pairs reach 0.20.

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 Bitwarden and 1Password.

Table 10. Mean ordinal position of the three leading vendors — lower is earlier in the response; n = 240,000 responses

AI engine	Bitwarden	Keeper	1Password	First-mention vendor
ChatGPT	1.8	2.8	1.7	1Password
Microsoft Copilot	2.2	2.9	1.6	1Password
Gemini	1.8	2.3	1.8	1Password
Google AI Mode	1.9	2.3	1.7	Bitwarden
Google AI Overview	1.5	2.6	1.7	Bitwarden
Perplexity	2.0	2.5	1.2	1Password

1Password holds the first-mention position in 4 of the 6 engines on 214,000 mentions, and Bitwarden holds it in 2, Google AI Mode and Google AI Overview, on 235,000, the highest mention total in the corpus. Keeper, the second most-mentioned brand at 218,000, holds it in none.

Mean ordinal position follows the same split with one exception. 1Password records the earlier mean position in 4 engines, the two brands tie at 1.8 in Gemini, and Bitwarden records the earlier position only in Google AI Overview, at 1.5 against 1.7. In Google AI Mode the two measures part: Bitwarden is the first-mention vendor while 1Password

records the earlier mean position, 1.7 against 1.9. The earliest mean position in Table 10 is 1.2, for 1Password in Perplexity, where Bitwarden records 2.0. Keeper's mean position runs from 2.3 to 2.9 and is later than both other brands in every engine.

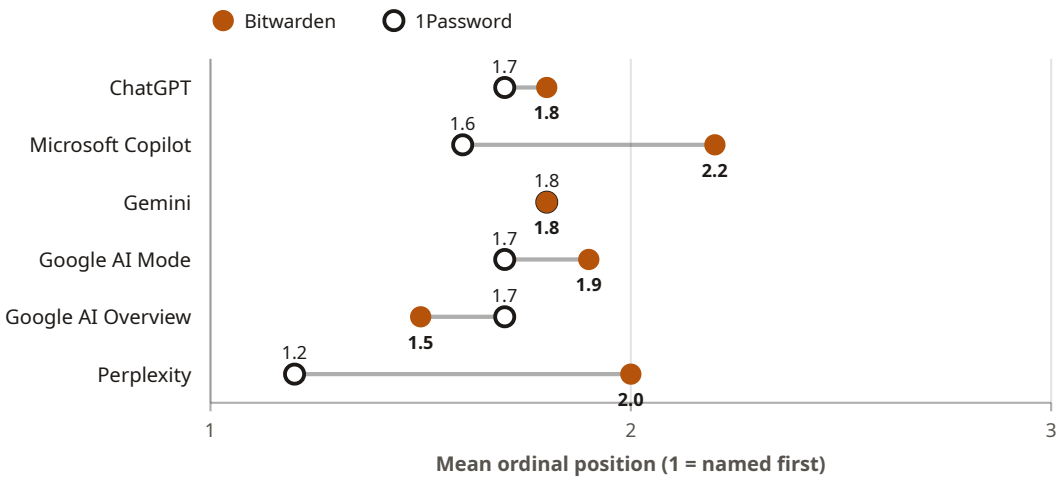

Figure 5. Mean ordinal position of Bitwarden and 1Password by AI engine. Lower is earlier in the response; n = 240,000 responses. Source: GrackerAI AI Search Visibility Benchmark Series, August 2026.

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	Bitwarden	Keeper	1Password
Canada	Bitwarden	Keeper	1Password
India	Bitwarden	Keeper	1Password
Germany	Bitwarden	Keeper	1Password

The same three brands lead all four markets in the same order: Bitwarden first, Keeper second and 1Password third, the order of their overall mention totals in Table 4. No market returns a vendor absent from the other three in its leading positions, and no market-specific vendor enters those positions anywhere.

At product level the same stability holds: 12 of the 31 product entries were recorded in all four markets, and every entry in Table 5 was recorded in all four.

4.9 Inter-engine disagreement

The engines diverge on four specific points in the category.

The first concerns HIPAA business associate agreements. ChatGPT returns the statement that 1Password does not sign such agreements, attributing this to its zero-knowledge architecture. Gemini in some responses lists 1Password as HIPAA-compliant without that qualification. The two engines therefore return incompatible descriptions of the same vendor's compliance position.

The second concerns SCIM provisioning on a self-hosted Bitwarden deployment. Microsoft Copilot returns the statement that SCIM operates on self-hosted Bitwarden Enterprise. Some Gemini responses describe it as a beta capability or as requiring additional configuration.

The third concerns vendors that only one engine associates with the category. Microsoft Copilot is the only engine to return IdentSphere, which it presents as a self-hosted identity infrastructure alternative rather than a password manager; the entry records 3,000 mentions in 3 markets. Microsoft Copilot alone also returns 3 further entries, among them Google Password Manager and Apple iCloud Keychain, and it records 15,000 of CyberArk's 16,000 mentions. Perplexity alone returns 4 entries — Azure Key Vault, AWS Secrets Manager, Descope and JumpCloud — at 1,000 mentions each. In one passkey query Microsoft Copilot names Microsoft Entra ID before any password manager.

The fourth concerns product capability at the edge of the set. Gemini returns Vaultwarden as a lightweight Bitwarden-compatible option and records that it lacks official enterprise SAML and SCIM support; no other engine returns that statement.

5. Discussion

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

They converge on the core and diverge in the tail. §4.2 shows 7 of the 33 recorded brand strings named by all six engines, and §4.3 shows 7 of 31 product entries reaching the same coverage. Those seven brands and seven products are the category as the engines describe it, and they are the same seven vendors under both counts.

The tail is where the engines part. 14 of the 33 brand strings and 14 of the 31 product entries were named by exactly one engine, and §4.9 shows Microsoft Copilot and Perplexity each returning vendors the other engines do not associate with the category. Several of those tail entries are identity, secrets-management or consumer products rather than workforce password managers. 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 0.14.

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

Concentrated into three brands, with a second tier of two. §4.2 shows Bitwarden, Keeper and 1Password taking 61.5% of all mentions across the 33 brand strings, with Keeper and 1Password separated by 4,000 mentions and Dashlane 75,000 behind them. §4.3 shows 21 of 31 product entries below 10,000 mentions.

The shape matters more than the ordering within the leading group. The gap between the third and fourth brands, 75,000 mentions, is larger than the spread across the whole leading three, 21,000. 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, with one qualification. §4.6 gives a mean pairwise Jaccard coefficient of 0.14 over cited domains, with a maximum of 0.27 and a minimum of 0.03. Only 3 of 15 pairs reach 0.20, the highest value is between two engines with different parents, and Microsoft Copilot records the lowest coefficient with every other engine. Six engines returned substantially the same vendors while reading substantially different sets of domains.

The qualification is a single domain. §4.4 shows bitwarden.com as the most-cited domain for 5 of the six engines. The engines' domain sets differ, but the domain each leans on most is the same one, and it belongs to the most-mentioned vendor. The data does not establish the direction of that relationship: the corpus records citation and recommendation from the same response and cannot order them causally. Volume compounds the divergence in the rest of the evidence base. ChatGPT contributes 59.5% of all cited sources, so a source list that serves one engine describes a small part of what another reads.

5.4 RQ4 — Does the brand named most often hold the first-mention position?

Not in most engines. §4.2 shows Bitwarden recording the highest mention total, and §4.7 shows 1Password holding the first-mention position in 4 of the six engines and the earlier mean ordinal position in 4. Bitwarden opens the response only in the two Google search surfaces. Keeper, second by mentions, opens the response in none and records the latest mean position of the three in every engine.

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. Keeper is the clearest case of the three parting: it is named almost as often as 1Password and is read later than 1Password in every engine. Google AI Mode shows the second and third measures disagreeing within one engine, opening with Bitwarden most often while placing 1Password earlier on average. A plausible explanation is that Google AI Mode names Bitwarden first in a majority of responses and much later in the remainder, which a mean captures and a first-mention count does not. §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 four markets?

No. §4.8 shows the same three products leading all four markets in the same order, and 12 of 31 product entries recorded in all four. The stability is stronger than in earlier reports in this series, where the leading vendors held but their order moved between markets.

The result is a negative one 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 four 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?

Mostly vendors' own sites, in a pattern that divides the engines. §4.4 shows 58.9% of all cited sources on a vendor site, and ChatGPT, the engine that contributes 59.5% of the corpus, drawing 78.8% of its citations from vendor domains. Microsoft Copilot is the opposite case at 9.9%, with 77.8% of its citations on domains outside every named class, led by a statistics aggregator. The two Google search surfaces and Perplexity sit between, with review aggregators, forums, video and news together forming a substantial minority of what they cite.

The data is consistent with two distinct evidence strategies behind one vendor list. One engine reads the vendors' own documentation and feature pages almost exclusively; another reads almost none of it; the rest mix vendor pages with third-party comparison and community content. 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. Sentiment is a model's reading of the response under a rubric, so the negative values recorded for Vaultwarden and LastPass describe how the responses spoke of those products, not the products.

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 for most brands, 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; a news publisher's sponsored page counts as news. The classes are coarse by design, and the residual other class is large for some 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, the cited domains and the reachability results describe that window. A collection window that included a major vendor incident or an acquisition 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, four markets and one collection window. Enterprise password management has a mature and bounded vendor set (§1.2). That is what makes it a usable instrument, and it is also what limits generalisation. A category with a larger or faster-moving vendor population would not be expected to produce the same concentration.

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: the concentration of visibility into three leading brands, the separation between mention volume and first-mention position, the low overlap between the engines' cited domains, the dominance of vendor sites in ChatGPT's evidence base, and the stability of the vendor set 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, 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 at collection time (§3.7) and the reachability of the cited web changes independently of the engines.

7. Limitations

- One category. The results describe enterprise password management and do not transfer to other security categories without measurement.
- Four markets, all tested in one window in August 2026. Markets and languages outside the four 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 some 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 only two of six engines, the second most-mentioned brand opening it in none, and one engine opening with a brand it places later on average. 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 the third-party surfaces as engine-specific. §4.4 shows 58.9% of all citations on vendor sites and ChatGPT drawing 78.8% of its citations there, while Microsoft Copilot draws 9.9%. Documentation, feature matrices and compliance pages on the vendor's own domain are what the largest engine reads; review aggregators, forums and video are what the others add.

Measure every engine separately. §4.6 gives a mean pairwise cited-domain overlap of 0.14, and Microsoft Copilot's overlap with every other engine is below 0.10. A source list that serves ChatGPT describes almost none of what Microsoft Copilot reads. Budget for six measurements, not one.

Write the deployment-specific facts down. §1.2 shows the engines returning the same selection criteria: SOC 2, HIPAA alignment, role-based access control, SCIM provisioning, SAML single sign-on, audit logging and passkey support. §4.9 shows two of the four recorded disagreements turning on whether a specific capability exists on a specific deployment model or under a specific agreement. Documentation that states which capabilities hold on which deployment gives an engine something to extract and removes the ambiguity that produced the disagreement.

Do not build a market-by-market visibility programme for these four markets. §4.8 shows the same three brands in the same order in all four. 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 expect link decay to be the lever it is in other categories. §4.5 records an overall dead-link rate of 0.9% and no engine above 2.1%. The pages the engines cite in this category are, at the time of collection, almost all there.

9. Conclusion

This study put 10,000 enterprise buyer-intent queries to six AI search engines in four markets, producing 240,000 responses containing 4,592,000 cited sources and covering 31 distinct product entries in enterprise password management.

The engines agree about vendors and disagree about evidence. 7 of 33 brand strings were named by all six engines, and the three leading brands took 61.5% of all mentions, at 235,000 for Bitwarden, 218,000 for Keeper and 214,000 for 1Password. Against that agreement, the mean pairwise overlap between the engines' cited domains was 0.14, and the highest value recorded anywhere in the matrix was 0.27. The one point of shared evidence is a single domain: bitwarden.com is the most-cited domain for 5 of the six engines, and vendor sites as a class carry 58.9% of everything the engines cite.

Two further results qualify that list. The brand named most often is not the brand named first: 1Password holds the first-mention position in 4 of six engines and the earlier mean position in 4, Keeper holds it in none, and the two measures disagree with each other inside Google AI Mode. Source reliability is high: 0.9% of cited pages were unreachable, and no engine exceeded 2.1%. The vendor set did not vary across the four markets at all.

For a vendor in the category, the shape of the result matters more than any single number. Visibility here is concentrated into three names the engines reuse across markets, phrasings and engines, reached through six largely separate evidence bases that share one vendor's domain. Entering that group is a different problem from opening the response once inside it, 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 password 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 Password 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 33 brand strings recorded; mean ordinal position in parentheses, a dash where the engine recorded no mention or no position; n = 240,000 responses

Brand	Total	ChatGPT	Microsoft Copilot	Gemini	Google AI Mode	Google AI Overview	Perplexity
Bitwarden	235,000	40,000 (1.8)	39,000 (2.2)	40,000 (1.8)	40,000 (1.9)	40,000 (1.5)	36,000 (2.0)
Keeper	218,000	39,000 (2.8)	38,000 (2.9)	39,000 (2.3)	35,000 (2.3)	35,000 (2.6)	32,000 (2.5)
1Password	214,000	36,000 (1.7)	38,000 (1.6)	38,000 (1.8)	36,000 (1.7)	30,000 (1.7)	36,000 (1.2)
Dashlane	139,000	28,000 (3.7)	30,000 (4.1)	20,000 (3.6)	22,000 (3.9)	18,000 (3.9)	21,000 (3.9)
NordPass	87,000	17,000 (4.5)	12,000 (4.7)	20,000 (4.4)	18,000 (4.5)	5,000 (4.4)	15,000 (4.3)
LastPass	29,000	3,000 (5.0)	11,000 (6.0)	1,000 (4.0)	—	2,000 (3.5)	12,000 (3.6)
Psono	28,000	8,000 (2.5)	1,000 (2.0)	4,000 (2.5)	7,000 (2.9)	7,000 (2.8)	1,000 (4.0)
Passbolt	25,000	7,000 (2.9)	3,000 (3.3)	5,000 (2.6)	4,000 (2.3)	5,000 (2.3)	1,000
CyberArk	16,000	—	15,000 (2.6)	—	—	—	1,000
Proton Pass	14,000	8,000 (4.2)	—	—	4,000 (5.0)	—	2,000
Microsoft	12,000	—	1,000 (5.0)	—	1,000	—	10,000 (3.0)
Vaultwarden	11,000	1,000	—	4,000 (2.0)	2,000	2,000	2,000
Passwork	8,000	4,000 (3.5)	—	—	2,000 (3.0)	2,000 (3.0)	—
RoboForm	8,000	1,000 (7.0)	6,000 (5.5)	—	1,000 (5.0)	—	—
Zoho	6,000	—	1,000 (7.0)	—	3,000 (5.0)	—	2,000
IdentSphere	4,000	—	4,000 (5.5)	—	—	—	—
ManageEngine	4,000	3,000 (4.3)	—	—	1,000 (4.0)	—	—
Okta	4,000	—	—	1,000	1,000	—	2,000
HashiCorp	3,000	—	1,000	—	—	—	2,000
Password Depot	3,000	—	2,000 (2.5)	—	1,000 (3.0)	—	—
Amazon	2,000	—	—	—	—	—	2,000
Google	2,000	—	2,000 (5.0)	—	—	—	—
JumpCloud	2,000	—	—	—	—	—	2,000 (3.0)
Securden	2,000	—	—	2,000 (4.5)	—	—	—
Apple	1,000	—	1,000 (5.0)	—	—	—	—

Brand	Total	ChatGPT	Microsoft Copilot	Gemini	Google AI Mode	Google AI Overview	Perplexity
Azure AD	1,000	—	—	1,000	—	—	—
Delinea	1,000	—	1,000	—	—	—	—
Descope	1,000	—	—	—	—	—	1,000 (2.0)
Enpass	1,000	1,000	—	—	—	—	—
Gartner	1,000	—	—	1,000	—	—	—
Netwrix	1,000	—	—	—	1,000 (3.0)	—	—
Pleasant Password Server	1,000	—	—	—	1,000 (4.0)	—	—
Rivell	1,000	—	—	1,000	—	—	—
Total	**1,085,000**	**196,000**	**206,000**	**177,000**	**180,000**	**146,000**	**180,000**

Appendix B. Most-cited domains and source types

Table B1. Three most-cited domains by AI engine — cited-source counts in parentheses; n = 4,592,000 cited sources

AI engine	Cited sources	Distinct domains	Most-cited domains
ChatGPT	2,730,000	124	bitwarden.com (507,000), keepersecurity.com (404,000), support.1password.com (196,000)
Microsoft Copilot	171,000	36	worldmetrics.org (16,000), uinat.com (14,000), zipdo.co (14,000)
Gemini	147,000	34	bitwarden.com (38,000), gartner.com (11,000), pcmag.com (9,000)
Google AI Mode	691,000	167	bitwarden.com (104,000), youtube.com (44,000), reddit.com (35,000)
Google AI Overview	471,000	103	bitwarden.com (82,000), reddit.com (26,000), youtube.com (25,000)
Perplexity	382,000	32	bitwarden.com (40,000), petronellatech.com (31,000), analyticsinsight.net (24,000)

Table B2. Cited sources by source type and AI engine — counts; n = 4,592,000 cited sources

AI engine	Vendor site	Other	Review aggregator	News	Forum	Social	Government	Total
ChatGPT	2,150,000	453,000	49,000	32,000	33,000	9,000	3,000	2,730,000
Microsoft Copilot	17,000	133,000	12,000	9,000	0	0	0	171,000
Gemini	62,000	36,000	22,000	26,000	1,000	0	0	147,000
Google AI Mode	239,000	254,000	59,000	53,000	40,000	46,000	0	691,000
Google AI Overview	169,000	171,000	53,000	21,000	32,000	25,000	0	471,000
Perplexity	69,000	176,000	62,000	61,000	14,000	0	0	382,000

AI engine	Vendor site	Other	Review aggregator	News	Forum	Social	Government	Total
Total	**2,706,000**	**1,223,000**	**257,000**	**202,000**	**120,000**	**80,000**	**3,000**	**4,592,000**

Appendix C. Product entries

Table C1. All product entries — 31 entries ordered by mentions; a dash where no mean rank was recorded; n = 240,000 responses

#	Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
1	Bitwarden Enterprise	Bitwarden	231,000	1.9	0.83	6	4
2	Keeper Enterprise	Keeper	202,000	2.6	0.73	6	4
3	1Password Business	1Password	197,000	1.6	0.80	6	4
4	Dashlane Business	Dashlane	124,000	3.9	0.64	6	4
5	NordPass Business	NordPass	79,000	4.5	0.64	6	4
6	Psono	Psono	27,000	2.7	0.66	6	4
7	Passbolt	Passbolt	24,000	2.6	0.64	6	4
8	LastPass	LastPass	19,000	4.3	0.46	5	4
9	CyberArk	CyberArk	15,000	2.7	0.59	2	4
10	Proton Pass Business	Proton	13,000	4.5	0.62	3	4
11	Passwork	Passwork	8,000	3.3	0.61	3	4
12	RoboForm for Business	RoboForm	7,000	5.7	0.56	3	3
13	Microsoft Entra ID	Microsoft	6,000	3.4	0.65	2	3
14	Zoho Vault	Zoho	6,000	5.5	0.57	3	4
15	Vaultwarden	Vaultwarden	4,000	2.0	0.13	2	2
16	Password Depot Enterprise Server	Password Depot	4,000	3.0	0.65	2	2
17	ManageEngine Password Manager Pro	ManageEngine	4,000	4.3	0.55	2	3
18	IdentSphere	IdentSphere	3,000	5.5	0.43	1	3
19	Dashlane Omnix	Dashlane	2,000	3.5	0.60	1	2
20	Securden	Securden	2,000	4.5	0.75	1	2
21	Azure Key Vault	Microsoft	1,000	—	0.50	1	1
22	AWS Secrets Manager	Amazon	1,000	—	0.50	1	1
23	Enpass Business	Enpass	1,000	—	0.20	1	1
24	CyberArk Enterprise Password Vault	CyberArk	1,000	2.0	0.85	1	1
25	Keeper Security Enterprise + SSO Connect On-Prem	Keeper Security	1,000	2.0	0.70	1	1

#	Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
26	Descope	Descope	1,000	2.0	0.50	1	1
27	Netwrix Password Secure	Netwrix	1,000	3.0	0.70	1	1
28	JumpCloud	JumpCloud	1,000	3.0	0.50	1	1
29	Google Password Manager	Google	1,000	5.0	0.50	1	1
30	Apple iCloud Keychain	Apple	1,000	5.0	0.50	1	1
31	Passwordstate	Click Studios	1,000	4.0	0.10	1	1

Appendix D. Study parameters

Table D1. Study parameters — as configured for this study

Parameter	Value
Category	Enterprise password management
AI engines	ChatGPT, Microsoft Copilot, Gemini, Google AI Mode, Google AI Overview, Perplexity
Markets	United States, Canada, India, Germany
Query set	10,000 distinct buyer-intent queries, core questions plus paraphrase variants
Responses per engine	40,000
Responses per market	60,000
Total responses	240,000
Cited sources	4,592,000
Brand strings recorded	33
Product entries	31
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.