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	2,690,000	4
Microsoft Copilot	40,000	171,000	4
Gemini	40,000	338,000	4
Google AI Mode	40,000	392,000	4
Google AI Overview	40,000	365,000	4
Perplexity	40,000	380,000	4
Total	**240,000**	**4,336,000**	**4**

Published: September 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: 4,336,000 cited sources · 59 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 4,336,000 cited sources and covered 59 distinct product entries. For every response the study recorded length, cited domains, source reachability, hedging, recency anchoring, truncation, refusal, and each brand mention with its ordinal position. The engines converge on a narrow vendor set: 6 of 12 brands were named by all six engines. The two leading brands took 44.4% of the mentions recorded across those brands, led by Bitwarden at 235,000 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. The engines diverge on evidence, with a mean pairwise cited-domain overlap of 0.18 and a maximum of 0.42, between Google AI Mode and Google AI Overview. Dead-link rates run from 5.3% to 23.7% against an overall rate of 10.9%, and the same three vendors lead all four markets in the same order.

Contents

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 6 of the 12 brands 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 Type II attestation, role-based access control, SAML single sign-on, SCIM provisioning, passkey support, and integration with a directory service. 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 corpus frames the decision around deployments of two hundred or more employees, where the product is chosen by security and IT teams who shortlist before contacting vendors and where 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, rather than inferring shared sourcing from shared conclusions. 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 record self-hosted deployment as a product in its own right?

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.	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 polarity score assigned to the sentence containing a brand mention, on a scale from minus one to one where higher is more favourable.	score
Cited source	One URL attached to a response by the engine as a reference.	count

Metric	Definition	Unit
Distinct domains	The number of different registrable domains cited by that engine across its responses. Reported per engine and never summed across engines.	count
Dead-link rate	Share of an engine's cited URLs that did not resolve to a reachable page when tested.	% 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 distinct product string recorded against a brand. A brand may hold several entries where engines named its products differently.	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.

Citations were extracted as the set of URLs the engine attached to a response. Each URL was reduced to its registrable domain for the distinct-domain and overlap measures, and each was tested for reachability; a URL that did not resolve to a reachable page at test time was recorded as a dead link. Reachability was tested once, during the collection window, so the dead-link results describe the state of the cited web at that moment.

Brand mentions were detected against a vendor name list and recorded with the ordinal position at which the brand appeared among the named vendors in that response, counting from one. The first-mention vendor for an engine is the brand that most often held position one. Product entries were recorded as the distinct product strings the engines used, which is why one brand may hold several entries in Appendix C. Sentiment was scored on the sentence containing each mention and reported on a scale from minus one to one.

Two implementation details are not documented by the instrumentation and are therefore not described here: the specific lexicon used to detect qualifying and date-anchoring language, and the model used to assign sentiment polarity. §6.1 states the consequence for how those two metrics should be read. Source-type classification was recorded by the instrumentation but did not separate the corpus, 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.

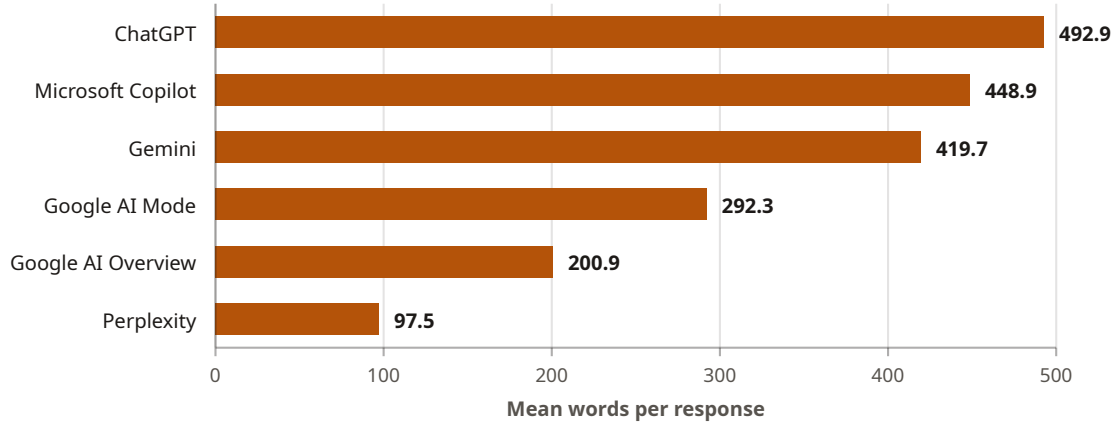

Figure 1. Mean response length by AI engine. Arithmetic mean word count per response; n = 40,000 responses per engine. Source: GrackerAI AI Search Visibility Benchmark Series, September 2026.

4.2 Brand visibility

Brand visibility was recorded for 12 brands. Table 4 gives each brand's total mentions, the number of engines that named it, and its mean sentiment.

Table 4. Brand visibility — all 12 brands recorded; n = 240,000 responses

Brand	Total mentions	Engines recording	Mean sentiment
Bitwarden	235,000	6	0.82
1Password	214,000	6	0.73
Keeper	163,000	6	0.65
Dashlane	139,000	6	0.58
NordPass	87,000	6	0.61
Keeper Security	55,000	5	0.79
LastPass	29,000	5	0.12
Psono	28,000	6	0.62
Passbolt	24,000	5	0.59
CyberArk	16,000	2	0.61
Proton Pass	11,000	3	0.56
Vaultwarden	11,000	5	-0.23
Total	**1,012,000**	**—**	**—**

Visibility is concentrated. The two most-mentioned brands, Bitwarden and 1Password, account for 44.4% of the mentions recorded across the 12 brands, at 235,000 and 214,000 mentions, 21,000 apart. Keeper follows at 163,000 and Dashlane at 139,000; the three leading brands together take 60.5%. Figure 2 shows the distribution. Below Dashlane the counts fall to 87,000 for NordPass and to 55,000 for Keeper Security.

The tables carry Keeper and Keeper Security as two separate brand strings, as the engines wrote them, and the instrumentation did not merge them. Summed, the two strings total 218,000 mentions, which is more than the 214,000 recorded for 1Password. 6 of the 12 brands were named by all six engines: Bitwarden, 1Password, Keeper, Dashlane, NordPass and Psono. CyberArk was named by two engines, and 15,000 of its 16,000 mentions were recorded by Microsoft Copilot. Mean sentiment runs from -0.23 for Vaultwarden and 0.12 for LastPass to 0.82 for Bitwarden; the other 10 brands fall between 0.56 and 0.82.

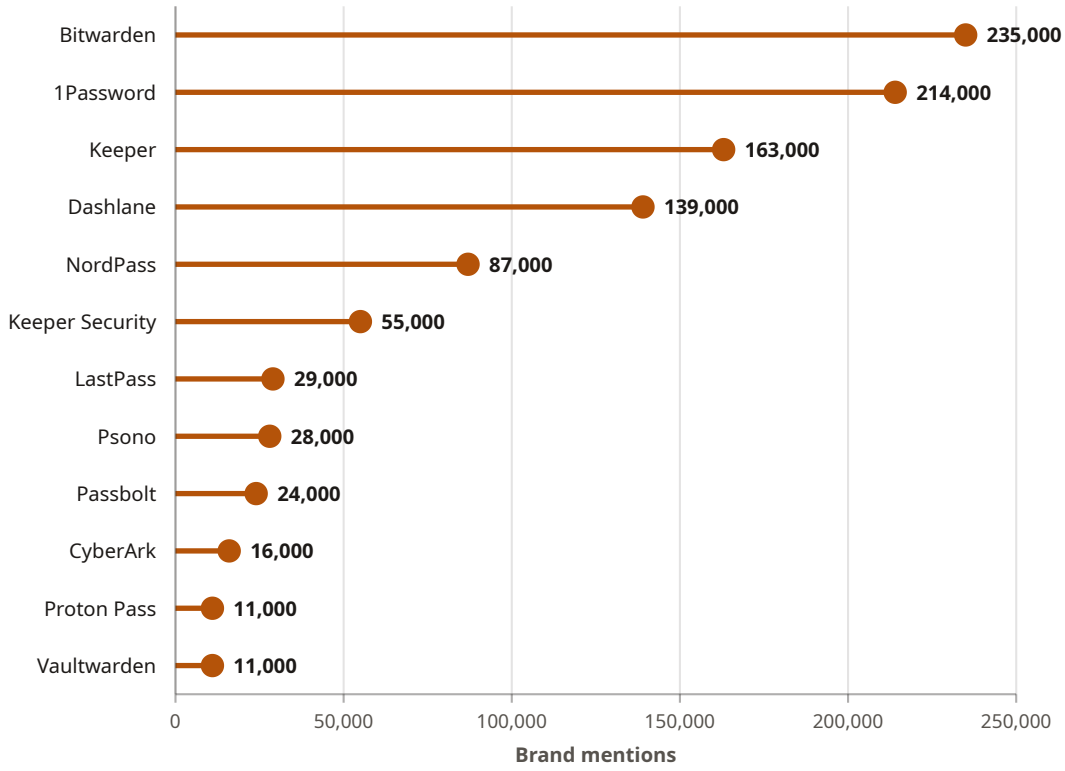

Figure 2. Brand mentions by vendor. All 12 brands recorded; n = 240,000 responses. Source: GrackerAI AI Search Visibility Benchmark Series, September 2026.

4.3 Product-level results

59 distinct product entries were recorded. An entry is one product string as an engine wrote it, so a brand holds several entries where engines named its products differently; Bitwarden holds 8, the most of any brand, and 1Password holds 6. Table 5 gives the ten most-mentioned entries.

Table 5. Product entries, ten most-mentioned — of 59 entries; n = 240,000 responses

Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
Bitwarden Enterprise	Bitwarden	206,000	1.9	0.82	6	4
1Password Business	1Password	187,000	1.6	0.80	6	4
Keeper Enterprise	Keeper	138,000	2.7	0.71	6	4
Dashlane Business	Dashlane	123,000	3.9	0.64	6	4
NordPass Business	NordPass	76,000	4.5	0.64	6	4
Keeper Security	Keeper Security	37,000	2.3	0.80	5	4
Psono	Psono	25,000	2.6	0.65	6	4
Passbolt	Passbolt	21,000	2.8	0.61	6	4
Keeper Security	Keeper	20,000	2.4	0.77	5	4
LastPass	LastPass	18,000	4.3	0.45	4	4

Product entries carry their own brand attributions, and Appendix C lists entries attributed to brands outside the 12 in Table 4. The distribution is long-tailed. 48 of the 59 entries record fewer than 10,000 mentions, 24 record exactly 1,000, and 29 were named by exactly one engine. 7 entries were named by all six engines, 19 were recorded in all four markets, and 24 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 206,000 mentions at a mean rank of 1.9, and 1Password Business records 187,000 at a mean rank of 1.6, the earlier of the two. Keeper Security appears twice in Table 5 as a product string, once attributed to the brand Keeper Security and once to Keeper.

4.4 Citation volume

The corpus contains 4,336,000 cited sources, distributed unevenly across the six engines, as Figure 3 shows.

Table 6. Cited sources and domain breadth by AI engine — n = 4,336,000 cited sources

AI engine	Cited sources	Share of corpus	Distinct domains	Most-cited domain
ChatGPT	2,690,000	62.0%	124	bitwarden.com (504,000)
Microsoft Copilot	171,000	3.9%	36	worldmetrics.org (16,000)
Gemini	338,000	7.8%	34	bitwarden.com (104,000)
Google AI Mode	392,000	9.0%	78	bitwarden.com (77,000)
Google AI Overview	365,000	8.4%	65	bitwarden.com (73,000)
Perplexity	380,000	8.8%	32	bitwarden.com (40,000)
Total	**4,336,000**	**—**	**—**	**—**

ChatGPT accounts for 62.0% of all cited sources in the corpus, 2,690,000 of 4,336,000, and cites from the widest domain population at 124 distinct domains. 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 504,000 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.

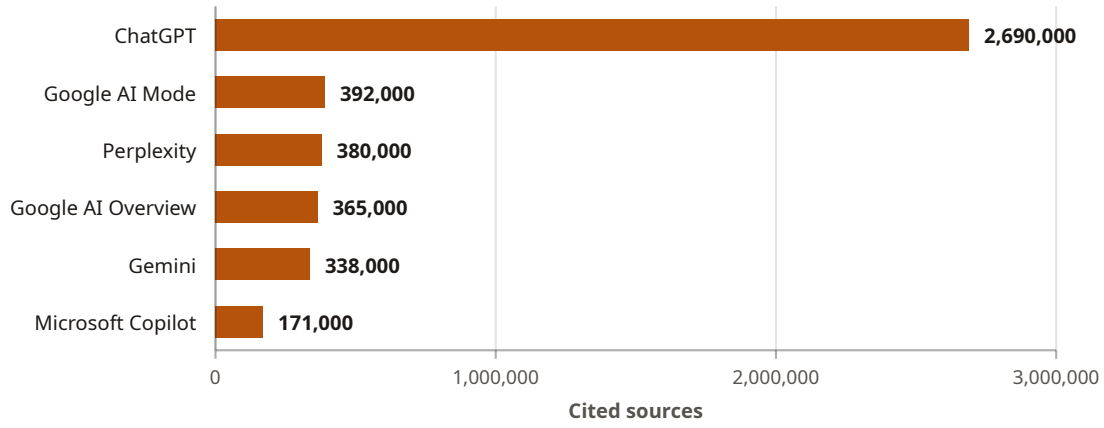

Figure 3. Cited sources by AI engine. Sources cited per engine; n = 4,336,000 cited sources. Source: GrackerAI AI Search Visibility Benchmark Series, September 2026.

4.5 Source quality and link decay

A share of the URLs the engines cited did not resolve to a reachable page when tested. Figure 4 gives the rate for each engine.

Table 7. Dead links and self-citation by AI engine — n = 4,336,000 cited sources

AI engine	Cited sources	Dead-link rate	Dead links	Self-citation rate
ChatGPT	2,690,000	6.9%	186,000	0%
Microsoft Copilot	171,000	5.3%	9,060	0%
Gemini	338,000	23.7%	80,100	0%
Google AI Mode	392,000	14%	54,900	0%
Google AI Overview	365,000	17%	62,000	0%
Perplexity	380,000	20.8%	79,000	0%
Total	**4,336,000**	**10.9%**	**471,060**	**—**

The overall dead-link rate is 10.9%, 471,060 of 4,336,000 cited URLs. Per-engine rates span a factor of 4.5, from 5.3% for Microsoft Copilot to 23.7% for Gemini, with Perplexity at 20.8% and Google AI Overview at 17%. ChatGPT records the second-lowest rate at 6.9%, and because it contributes 2,690,000 of the 4,336,000 cited sources it still accounts for 186,000 of the 471,060 dead links, 39.5% 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.

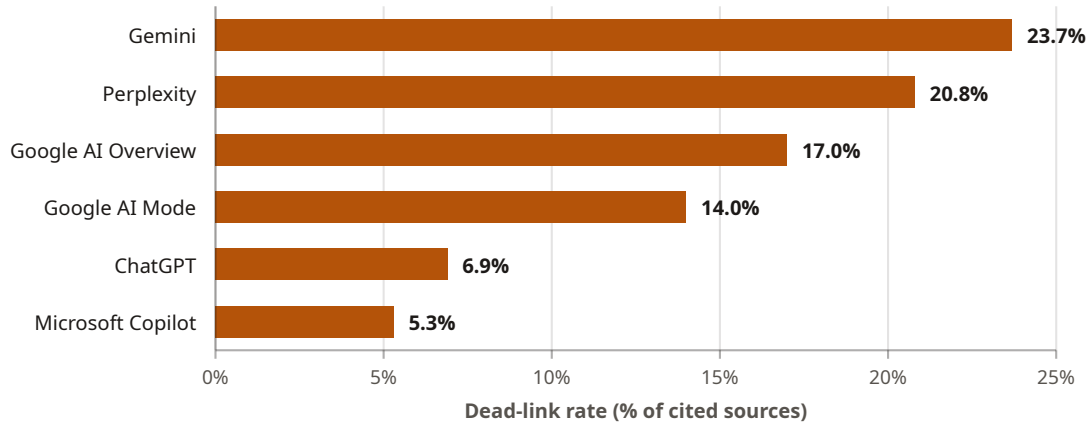

Figure 4. Dead-link rate by AI engine. Share of each engine's cited URLs that did not resolve when tested; n = 4,336,000 cited sources. Source: GrackerAI AI Search Visibility Benchmark Series, September 2026.

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 8 gives the full matrix.

Table 8. Cross-engine cited-domain overlap — Jaccard coefficient over cited-domain sets; 0 is disjoint, 1 is identical; n = 4,336,000 cited sources

	ChatGPT	Copilot	Gemini	Google AI Mode	Google AI Overview	Perplexity
ChatGPT	**1.00**	0.07	0.16	0.26	0.21	0.13
Copilot	0.07	**1.00**	0.04	0.05	0.05	0.08
Gemini	0.16	0.04	**1.00**	0.32	0.27	0.27
Google AI Mode	0.26	0.05	0.32	**1.00**	0.42	0.13
Google AI Overview	0.21	0.05	0.27	0.42	**1.00**	0.17
Perplexity	0.13	0.08	0.27	0.13	0.17	**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 0.18. The highest value is 0.42, between Google AI Mode and Google AI Overview. The lowest is 0.04, between Microsoft Copilot and Gemini. The three engines operated by the same parent — Gemini, Google AI Mode and Google AI Overview — record coefficients of 0.32, 0.27 and 0.42 with each other.

Microsoft Copilot records the lowest coefficient with every other engine: its five values run from 0.04 to 0.08, the latter with Perplexity. Outside the same-parent group the highest value is 0.27, between Gemini and Perplexity, followed by 0.26 between ChatGPT and Google AI Mode.

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 9 gives all three for the three most-mentioned brands, and Figure 5 shows the two leading brands.

Table 9. Mean ordinal position of the leading vendors — lower is earlier in the response; n = 240,000 responses

AI engine	Bitwarden	1Password	Keeper	First-mention vendor
ChatGPT	1.8	1.7	2.9	1Password
Microsoft Copilot	2.2	1.6	2.8	1Password
Gemini	1.8	1.8	2.3	1Password
Google AI Mode	2.0	1.7	2.3	Bitwarden
Google AI Overview	1.5	1.7	2.7	Bitwarden
Perplexity	2.0	1.2	2.5	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 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 2.0. The earliest mean position in Table 9 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 leading brands in every engine.

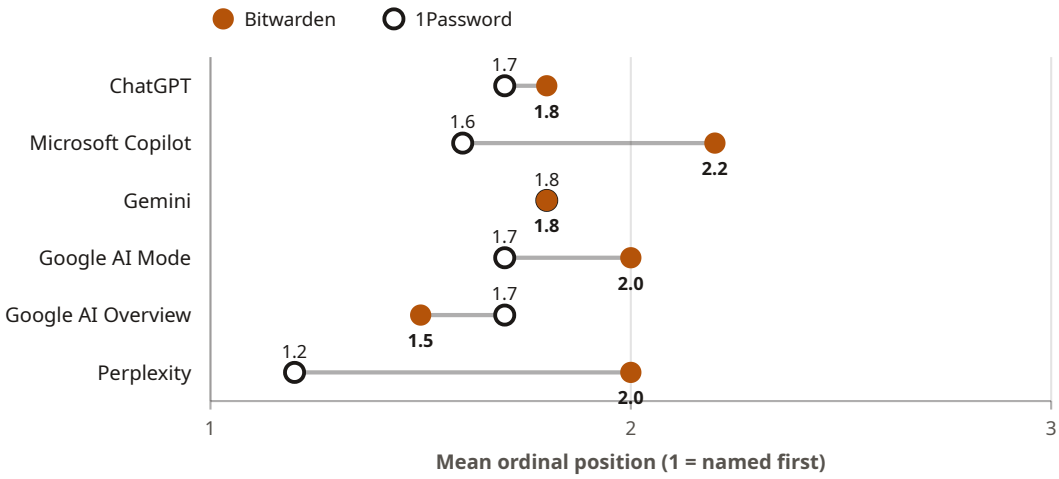

Figure 5. Mean ordinal position of Bitwarden and 1Password by AI engine. Lower is earlier in the response; n = 240,000 responses. Source: GrackerAI AI Search Visibility Benchmark Series, September 2026.

4.8 Geographic variance

Table 10 gives the three leading products in each market.

Table 10. Leading products by market — n = 60,000 responses per market

Market	First	Second	Third
United States	Bitwarden Enterprise	1Password Business	Keeper Enterprise
Canada	Bitwarden Enterprise	1Password Business	Keeper Enterprise

Market	First	Second	Third
India	Bitwarden Enterprise	1Password Business	Keeper Enterprise
Germany	Bitwarden Enterprise	1Password Business	Keeper Enterprise

The same three products lead all four markets in the same order: Bitwarden Enterprise first, 1Password Business second and Keeper Enterprise third. No market returns a vendor absent from the other three in its leading positions, and no market-specific vendor enters those positions anywhere. Mean sentiment by market falls between 0.64 and 0.68.

At product level the same stability holds: 19 of the 59 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. Gemini treats 1Password as supportive of HIPAA requirements without that distinction. 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. ChatGPT returns the statement that SCIM operates on self-hosted Bitwarden Enterprise. Microsoft Copilot in some responses omits the claim and in others qualifies it.

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 alternative rather than a conventional password manager; the entry records 3,000 mentions in 3 markets. Microsoft Copilot also 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.

The fourth concerns product capability. Gemini alone returns the claim that Psono Enterprise Edition supports SAML and audit logging, and in the same responses records that native SCIM support could not be confirmed. No other engine returns either 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 6 of the 12 recorded brands named by all six engines, and §4.3 shows 7 of 59 product entries reaching the same coverage. The difference between those two figures is the answer: the engines agree about companies and disagree about product names. A brand is recognised across all six; the string an engine uses for that brand's product is not, and Bitwarden alone holds 8 such strings.

The tail is where the engines part. 29 of the 59 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 0.18.

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

Concentrated into two brands, with a second tier of two. §4.2 shows Bitwarden and 1Password taking 44.4% of the mentions recorded across the 12 brands, separated from each other by 21,000 mentions and from Keeper by 51,000. §4.3 shows 48 of 59 product entries below 10,000 mentions.

The Keeper result qualifies the ordering. §4.2 records Keeper and Keeper Security as separate strings; summed, they exceed 1Password's total. The data is consistent with a category whose second position depends on how the engines spell a vendor's name. The study cannot determine from the corpus whether the two strings were used for the same product line in every response. What the measurement does establish is a boundary: a brand is either inside the group the engines reuse across queries and markets, or in a tail where mention counts fall away quickly.

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 0.18 over cited domains, with a maximum of 0.42 and a minimum of 0.04. The highest value is between the two Google search surfaces, 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 62.0% of all cited sources and cites from 124 distinct domains against 32 for Perplexity, 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, which are also the pair with the highest cited-domain overlap in §4.6.

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. Google AI Mode shows the second and third 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 19 of 59 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 record self-hosted deployment as a product in its own right?

Yes, in a small number of entries concentrated in one engine family. §4.3 and Appendix C record 7 product entries whose string carries a self-hosted or on-premise qualifier, totalling 16,000 mentions. Google AI Mode named all 7, and 6 of the 7 were named only by Google AI Mode, Google AI Overview or Gemini. The largest, Bitwarden Enterprise Self-Hosted, records 6,000 mentions at a mean rank of 1.0, tied with 4 other entries for the earliest mean rank in Appendix C, and a mean sentiment of 0.90.

The data is consistent with the two Google search surfaces treating deployment model as part of the product's identity, which matches their first-mention preference for Bitwarden in §4.7 and the corpus's association of Bitwarden with self-hosting in §1.2. The other engines record self-hosting as a property of a product rather than as a product. The measurement does not establish whether that difference changes which vendor a buyer sees first; it establishes only that the same capability is written into the product string by some engines and not by others.

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 records two brand strings for Keeper and cannot say from the corpus whether they refer to one product line.

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 was scored by an undocumented model (§3.7), so the negative values recorded for Vaultwarden and LastPass should be read as a property of the scorer applied to those sentences, not as a finding about 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 scored on the sentence containing each mention by a model the instrumentation does not document (§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 scorer 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.

Truncation is an inference from the terminal characters of a response, not an observation of a process. It cannot separate a response that stopped mid-sentence from one whose last sentence lacked terminal punctuation, and it is reported here as a characteristic of the recorded text rather than as a property of the engine. Hedging and recency anchoring rest on undocumented lexicons (§3.7) and are subject to the same limit. The same detector was 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 two leading brands, the separation between mention volume and first-mention position, the elevated overlap between the two Google search surfaces, and the stability of the vendor set across markets. All four 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 in particular, 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.
- Two processing steps are undocumented by the instrumentation: the lexicon behind the hedging and recency measures, and the sentiment model. Both are stated where they are used.
- Brand strings were recorded as the engines wrote them and were not merged, so Keeper and Keeper Security are reported separately throughout. Source-type classification did not separate the corpus and is not reported.

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, 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.

Publish one brand string and one product string. §4.2 records the second-placed vendor split across two brand names, and §4.3 records 8 product strings for the leading vendor. Any measurement that counts strings will divide a vendor's share of voice across the variants the engines happen to use. Consistent naming in documentation, pricing pages and third-party listings is the only lever a vendor holds over that.

Measure every engine separately. §4.6 gives a mean pairwise cited-domain overlap of 0.18, 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 specifics the engines reuse, and write the self-hosted case explicitly. §1.2 shows the engines returning the same selection criteria: SOC 2 Type II, role-based access control, SAML single sign-on, SCIM provisioning, passkey support and directory integration. §4.9 shows two of the four recorded disagreements turning on whether a specific

capability exists on a specific deployment model. 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 products 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.

Maintain the pages the engines cite. §4.5 records an overall dead-link rate of 10.9%, reaching 23.7% for Gemini. A vendor's own retired or moved URLs are part of that population, and a citation pointing at an unreachable page carries the brand name without carrying the reader.

9. Conclusion

This study put 10,000 enterprise buyer-intent queries to six AI search engines in four markets, producing 240,000 responses containing 4,336,000 cited sources and covering 59 distinct product entries in enterprise password management.

The engines agree about vendors and disagree about evidence. 6 of 12 brands were named by all six engines, and the two leading brands took 44.4% of the mentions recorded across those brands, at 235,000 for Bitwarden and 214,000 for 1Password. Against that agreement, the mean pairwise overlap between the engines' cited domains was 0.18, and the highest value recorded anywhere in the matrix was 0.42, between the two Google search surfaces. The one point of shared evidence is a single domain: bitwarden.com is the most-cited domain for 5 of the six engines.

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, and the two measures disagree with each other inside Google AI Mode. Source reliability varies by a factor of 4.5 across the engines, from 5.3% to 23.7% of cited URLs unreachable, against an overall rate of 10.9%. 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 two names the engines reuse across markets, phrasings and engines, reached through six largely separate evidence bases, and split by how the engines spell a vendor's name. 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, the brand and product mention records with their ordinal positions, and the URL reachability results. The dataset for this report was generated in August 2026.

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. (September 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 — the 10 brands with a per-engine breakdown; 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 (2.0)	40,000 (1.5)	36,000 (2.0)
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)
Keeper	163,000	35,000 (2.9)	23,000 (2.8)	24,000 (2.3)	28,000 (2.3)	21,000 (2.7)	32,000 (2.5)
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 (4.6)	12,000 (4.7)	20,000 (4.4)	18,000 (4.5)	5,000 (4.4)	15,000 (4.3)
Keeper Security	55,000	4,000 (2.0)	15,000 (3.3)	15,000 (2.1)	7,000 (1.8)	14,000 (2.3)	—
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	24,000	7,000 (2.9)	3,000 (3.3)	5,000 (2.6)	4,000 (2.3)	5,000 (2.3)	—
CyberArk	16,000	—	15,000 (2.6)	—	—	—	1,000
Total	**990,000**	**178,000**	**187,000**	**167,000**	**162,000**	**142,000**	**154,000**

Proton Pass and Vaultwarden were recorded in the share-of-voice table only, at 11,000 mentions each, and carry no per-engine breakdown.

Appendix B. Most-cited domains

Table B1. Three most-cited domains by AI engine — cited-source counts in parentheses; n = 4,336,000 cited sources

AI engine	Cited sources	Distinct domains	Most-cited domains
ChatGPT	2,690,000	124	bitwarden.com (504,000), keepersecurity.com (394,000), support.1password.com (196,000)
Microsoft Copilot	171,000	36	worldmetrics.org (16,000), zipdo.co (14,000), uinat.com (14,000)
Gemini	338,000	34	bitwarden.com (104,000), techrepublic.com (32,000), deel.com (25,000)
Google AI Mode	392,000	78	bitwarden.com (77,000), youtube.com (29,000), reddit.com (22,000)
Google AI Overview	365,000	65	bitwarden.com (73,000), youtube.com (22,000), gartner.com (21,000)
Perplexity	380,000	32	bitwarden.com (40,000), petronellatech.com (31,000), analyticsinsight.net (24,000)

Appendix C. Product entries

Table C1. All product entries — 59 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	206,000	1.9	0.82	6	4
2	1Password Business	1Password	187,000	1.6	0.80	6	4
3	Keeper Enterprise	Keeper	138,000	2.7	0.71	6	4
4	Dashlane Business	Dashlane	123,000	3.9	0.64	6	4
5	NordPass Business	NordPass	76,000	4.5	0.64	6	4
6	Keeper Security	Keeper Security	37,000	2.3	0.80	5	4
7	Psono	Psono	25,000	2.6	0.65	6	4
8	Passbolt	Passbolt	21,000	2.8	0.61	6	4
9	Keeper Security	Keeper	20,000	2.4	0.77	5	4
10	LastPass	LastPass	18,000	4.3	0.45	4	4
11	CyberArk	CyberArk	15,000	2.7	0.59	2	4
12	Proton Pass for Business	Proton	8,000	4.1	0.62	3	4
13	Bitwarden Enterprise / Teams	Bitwarden	7,000	2.1	0.86	3	3
14	RoboForm for Business	RoboForm	7,000	5.7	0.56	3	3
15	Bitwarden Enterprise Self-Hosted	Bitwarden	6,000	1.0	0.90	4	4

#	Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
16	Bitwarden Teams & Enterprise	Bitwarden	6,000	1.7	0.86	3	4
17	Microsoft Entra ID	Microsoft	6,000	3.4	0.65	2	3
18	Passwork	Passwork	6,000	3.3	0.55	2	4
19	Zoho Vault	Zoho	6,000	5.5	0.57	3	4
20	1Password Teams & Business	1Password	4,000	1.3	0.89	2	4
21	Keeper Enterprise Password Manager	Keeper	4,000	1.5	0.88	1	4
22	Vaultwarden	Vaultwarden	4,000	2.0	0.13	2	2
23	ManageEngine Password Manager Pro	ManageEngine	4,000	4.3	0.55	2	3
24	Proton Pass Business	Proton Pass	4,000	5.5	0.60	2	4
25	Passbolt Pro Self-Hosted	Passbolt	3,000	2.0	0.83	2	2
26	Password Depot Enterprise Server	Password Depot	3,000	2.7	0.63	2	2
27	Keeper Business & Enterprise	Keeper	3,000	3.0	0.77	2	3
28	IdentSphere	IdentSphere	3,000	5.5	0.43	1	3
29	Bitwarden Enterprise Server	Bitwarden	2,000	1.0	0.85	1	2
30	1Password Teams/Business	1Password	2,000	1.0	0.85	2	2
31	1Password Business & Enterprise	1Password	2,000	1.5	0.85	2	2
32	Bitwarden Teams and Enterprise	Bitwarden	2,000	2.0	0.85	2	2
33	Passwork On-Premise Edition	Passwork	2,000	3.0	0.78	1	2
34	Psono Enterprise Self-Hosted	Psono	2,000	3.5	0.70	2	2
35	Dashlane Omnix	Dashlane	2,000	3.5	0.60	1	2
36	Bitwarden Self-Hosted Official Enterprise	Bitwarden	1,000	1.0	0.90	1	1
37	Bitwarden Enterprise + Self-Hosted	Bitwarden	1,000	1.0	0.90	1	1
38	Azure Key Vault	Microsoft	1,000	—	0.50	1	1
39	AWS Secrets Manager	Amazon	1,000	—	0.50	1	1
40	Enpass Business	Enpass	1,000	—	0.20	1	1
41	CyberArk Enterprise Password Vault	CyberArk	1,000	2.0	0.85	1	1
42	1Password Business and Enterprise	1Password	1,000	2.0	0.85	1	1
43	Keeper Security Enterprise + SSO Connect On-Prem	Keeper Security	1,000	2.0	0.70	1	1
44	Descope	Descope	1,000	2.0	0.50	1	1
45	1Password Business / Enterprise Plans	1Password	1,000	3.0	0.80	1	1
46	Netwrix Password Secure	Netwrix	1,000	3.0	0.70	1	1
47	JumpCloud	JumpCloud	1,000	3.0	0.50	1	1

#	Product	Brand	Mentions	Mean rank	Mean sentiment	Engines	Markets
48	NordPass Teams & Business	NordPass	1,000	4.0	0.75	1	1
49	NordPass Teams	NordPass	1,000	4.0	0.75	1	1
50	NordPass Business & Enterprise	NordPass	1,000	4.0	0.75	1	1
51	LastPass Teams / Enterprise Plans	LastPass	1,000	4.0	0.75	1	1
52	Securden Business Password Manager	Securden	1,000	4.0	0.70	1	1
53	Pleasant Password Server	Pleasant Password Server	1,000	4.0	0.70	1	1
54	Dashlane (Team & Business)	Dashlane	1,000	4.0	0.70	1	1
55	Securden	Securden	1,000	5.0	0.80	1	1
56	Proton Pass Team Plans	Proton Pass	1,000	5.0	0.75	1	1
57	Google Password Manager	Google	1,000	5.0	0.50	1	1
58	Apple iCloud Keychain	Apple	1,000	5.0	0.50	1	1
59	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	4,336,000
Brands recorded	12
Product entries	59
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.
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.
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 sentence-level polarity scores for a brand's mentions, 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.