FIRST-PARTY OBSERVATIONS WEEKLY SNAPSHOT

THE WEEK IN CRAWL DATA

The busiest day accounted for 19.72% of weekly crawl requests.

478,775

AI and search crawler requests on BrandPedia

94,417 requests on November 13. Collection continuity is unverified.

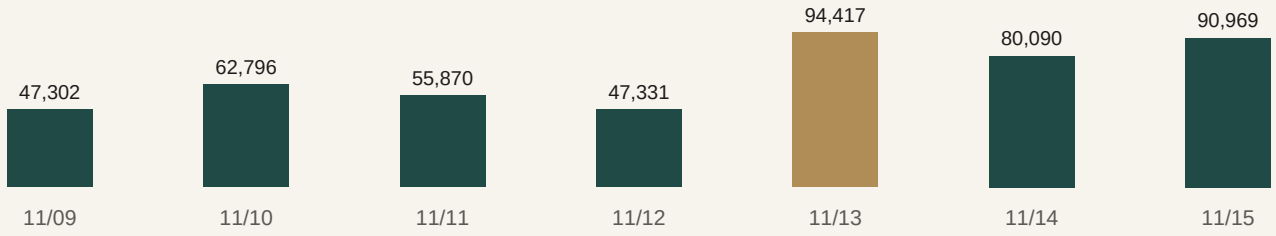
MATCHED PROFILES
84,255
Matched 2xx visits

SOURCE COMPANIES
14
22 crawler names

MULTI-SOURCE SHARE
21.05%
Of visited profiles

DAILY REQUESTS

Linear scale from zero



The gold bar marks November 13. The busiest hour was November 13, 03:00–04:00, New York time: 9,796 requests.

Window: November 9, 00:00 to November 16, 00:00, New York time; end excluded. Counts include failed responses. Collection continuity is unverified, so no week-over-week comparison is made.

SOURCES

High volume does not mean broad coverage.

Request volume measures frequency. Profile reach is the share of this week's 84,255 visited profiles with a recorded 2xx visit from each source.

Source	Requests	Share	Profiles	Reach
Amazon	156,073	32.60%	64,797	76.91%
OpenAI	147,480	30.80%	4,927	5.85%
Google	78,254	16.34%	14,421	17.12%
Perplexity	34,239	7.15%	7,061	8.38%
Meta	32,244	6.73%	2,641	3.13%
Other 9	30,485	6.37%	--	--

Reading the contrast: OpenAI generated 147,480 requests across 4,927 matched profiles; Amazon generated 156,073 across 64,797. Request counts and profile coverage tell different stories.

Crawler types behind the requests

Amazon: Amazonbot 156,068; Amzn-User 5. OpenAI: GPTBot 121,737; OAI-SearchBot 25,537; ChatGPT-User 206.

> **HOW TO READ REACH**
> Sources overlap: reach percentages do not sum to 100%. Reach uses visited profiles, not the full BrandPedia inventory. The combined distinct reach of the remaining sources is not calculated, because overlap would make a sum misleading.

Source companies are inferred from User-Agent strings. Identity-verification results were not stored. Names do not establish training use, citations, or user-triggered activity.

PROFILE COVERAGE

Profile coverage splits into exclusive source groups.

17,737 of 84,255 matched profiles (21.05%) received visits from at least two source companies.

Source count	Profiles	Share
One	66,518	78.95%
Exactly two	11,822	14.03%
Three or more	5,915	7.02%

Top 10 profiles | recorded 2xx visits

Profile	Amzn.	OpenAI	Google	Perpl.	Meta	Shape	You	MSFT	Exa	DDG	Req.
Microsoft	●	●	●	●	●	●	-	-	-	-	445
Disney	-	●	-	●	●	●	-	●	●	-	308
General Motors	●	●	●	●	●	●	-	●	-	-	283
Ford	●	●	-	●	-	●	-	●	-	-	282
Coca-Cola	●	●	-	●	●	●	-	●	-	-	281
BMW	●	●	-	●	●	●	-	-	-	-	279
Volkswagen	●	●	-	●	●	●	-	-	-	-	268
Apple	●	●	-	●	●	●	-	●	-	-	258
Unilever	●	●	-	●	●	●	-	-	-	-	232
Porsche	●	●	-	●	●	●	-	-	-	-	231

Filled circle = recorded visit; dash = no visit recorded in this window. Symbols show presence, not frequency, citations, or recommendations.

Abbreviations: Amzn. = Amazon; OpenAI = OpenAI; Google = Google; Perpl. = Perplexity; Meta = Meta; Shape = Shape; You = You.com; MSFT = Microsoft; Exa = Exa; DDG = DuckDuckGo. The matrix shows 10 of 14 source companies.

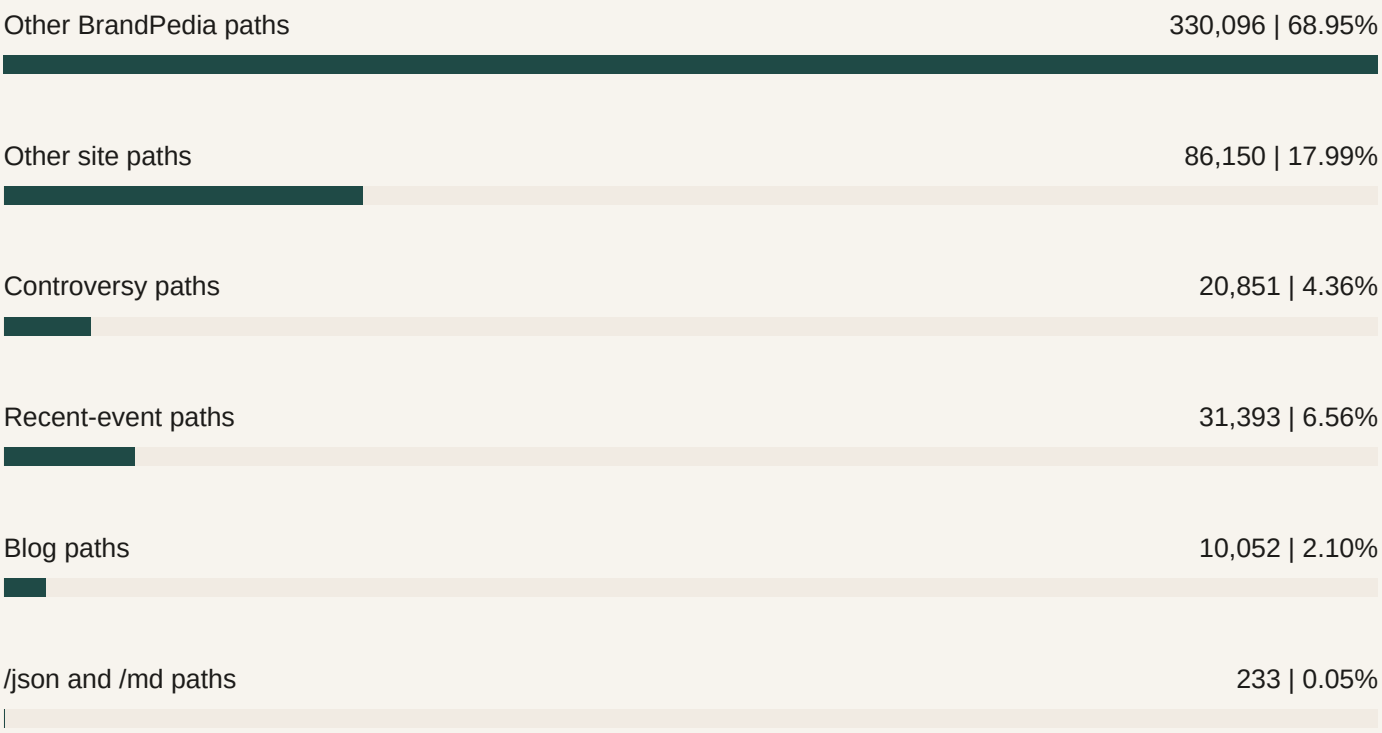
Labels use the English profile name when it is Latin text; otherwise the stable profile key is shown. Related path spellings merge by database identity. Distinct database records stay distinct.

CONTENT

Event-related paths received 52,244 requests.

Controversy paths contributed 20,851 requests; recent-event paths contributed 31,393. All path groups below include failed responses.

CONTENT PATHS | REQUESTS AND SHARE



All bars use the same linear scale from zero. The six mutually exclusive groups total 478,775 requests.

What the path categories mean

Other BrandPedia paths include main profiles and some site entry points. Blog requests span 54 distinct paths; /json and /md requests span 145. A path is not an article or a profile.

> **INTERPRETATION**
> Few explicit-format requests do not mean crawlers cannot read ordinary pages. Request counts alone do not establish successful delivery or content use.

SECTORS

41.97% are Other or unclassified.

35,361 of 84,255 covered profiles fall into these buckets. This classification gap limits interpretation of the selected-sector counts.

CLASSIFICATION OF VISITED PROFILES



Named sectors: 58.03% Other / unclassified: 41.97%

Six selected sectors	Visited profiles
Consumer discretionary	11,280
Communication services	10,738
Industrials	7,100
Consumer staples	7,075
Financials	4,548
Information technology	3,332

These six sectors are a selected subset, including sectors with zero visited profiles when that is what the extract shows. Counts use classifications at extraction time for brands, companies, and organizations matched to 2xx visits.

> **COUNTS ARE NOT PREFERENCES**
> Sector populations differ. Without comparable inventory denominators, larger counts do not establish crawler preference or higher whole-sector coverage.

Next: reduce classification gaps and retain end-of-week inventory snapshots before calculating sector-level coverage rates.

RESPONSE QUALITY

0.461% of requests returned HTTP errors.

2,209 of 478,775 requests returned status 400 or above. Stored rows are GET and HEAD document requests, including failures. The table does not store the method itself.

HTTP status	Requests	Share
200	470,309	98.23%
308	6,257	1.31%
403	4	<0.001%
404	2,191	0.458%
429	14	0.003%

Investigate the failures

Review the paths behind 2,191 not-found responses (404) and 4 forbidden responses (403). This snapshot provides aggregate counts; it does not identify the affected URLs or explain the access failures.

Successful response is not proof of use

470,309 returned 200; 6,257 returned 308. HEAD requests do not return a page body. A non-error status does not establish that content was read.

> **KEEP COLLECTION HEALTH SEPARATE**
> A low observed error rate does not prove every request was logged. Check outages, delays, retention, and duplicates before making period comparisons.

NEXT ACTIONS

What to check before the next comparison.

Turn the observed patterns into specific checks. These are investigation priorities, not evidence of AI citations, recommendations, or business impact.

Validate the daily concentration

Compare collection-health records with the November 13 pattern before treating it as demand. No day in this window fell below 2,000 recorded requests.

Measure repeat visits and source diversity

Use equally complete windows to compare returning profiles and multi-source share. Preserve the source definitions and profile identity rules.

Review failures and classification gaps

Inspect the URLs behind the 2,191 404 and 4 403 responses, and prioritize the 35,361 profiles in Other or unclassified. Break response success down by content group before describing content as delivered.

> **NEXT REPORT: COMPARABILITY GATE**
> Publish week-over-week changes only after both windows are sufficiently complete and use the same definitions. Otherwise show the current snapshot and explain what prevents comparison.

Explore a brand profile

Open BrandPedia profiles

Profile pages show cumulative statistics, which use a different window from this weekly snapshot.

METHODS AND SOURCES

A snapshot of observed site traffic.

Reporting window: November 9, 00:00 to November 16, 00:00, 2026, New York time. The end timestamp is excluded. The article is published on Monday.

Collection: sources are classified by User-Agent. SEO tools and other_bot rows are excluded. Aggregates keep one row per hit_key. 0 extra stored rows shared a hit_key and were not counted again. No log rows were deleted. A full calendar window does not prove uninterrupted capture.

Profile matching: only matched 2xx visits count toward coverage. /brand-wiki/ and /brandpedia/ paths merge by database identity at extraction, after removing a profile-key suffix of .json or .md. Matched profile requests total 346,082; 1,407 unmatched path keys representing 4,100 successful requests are excluded. Profile paths under /brand-wiki/ contributed 382,565 requests and /brandpedia/ contributed 0 requests in the sitewide deduplicated window. Both prefixes normalize to the same profile key before matching.

History: logging began October 28, 2026 at 11:17 EDT. 1,409 profiles first appear in available logs this week; first observed does not mean first discovered by AI.

Limits: identity verification was not stored. Requests are not unique users, citations, recommendations, training adoption, or conversions. No end-of-week inventory snapshot is part of this extract, so whole-database coverage is not calculated.

Extraction: November 19, 2026 at 02:43 EST. Methodology 2026-11-19.1. Current classifications and later corrections can change a rebuild; a sent issue keeps this frozen file.

Read and verify

BrandPedia methodology

Frozen data is attached to this PDF as **BrandPedia-Crawl-Data-2026-11-09.json**. Keep the PDF and the attachment together. The website article for this issue is not published, so this PDF includes no report URL or QR code.