

The 2026 AI Crawler Report

We analysed 14,227 real AI-crawler visits, checking each one against the operator's true network identity: its own live-published IP ranges and forward-confirmed reverse DNS. The result is a clear picture of which AI crawlers can prove who they are, and why an IP outside a published range is not automatically fake.

14,227 crawler visits · 18 AI crawlers · ~2,015 unique IPs · 10 June–6 August 2026 · every figure measured, none simulated

56%

verified: IP inside the operator's live-published range, or confirmed by reverse DNS

34%

outside a published range: from an IP not on the operator's list. A flag, not proof of fraud

10%

unverifiable: the operator publishes nothing to check against

Who publishes the means to verify them

The useful divide is not real versus fake. It is which operators publish IP ranges you can check at all.

OpenAI, Anthropic, Apple, Perplexity and Google do. **ByteDance (Bytespider), Common Crawl, Cohere, Huawei (PetalBot) and You.com** publish nothing, so their crawlers cannot be verified by IP by anyone. The news this year: Anthropic reversed its long-standing no-publish stance in 2026, and ClaudeBot moved to 92% verified overnight.

The busiest crawlers, and how much of each verified

AI CRAWLER	VISITS	VERIFIED	OUTSIDE RANGE	UNVERIFIABLE	PUBLISHES RANGES
ChatGPT-User (OpenAI)	7,888	53%	47%	0%	Yes
ClaudeBot (Anthropic)	1,962	92%	8%	0%	Yes (new in 2026)
Bytespider (ByteDance)	1,062	0%	13%	87%	No
GPTBot (OpenAI)	971	82%	18%	0%	Yes
OAI-SearchBot (OpenAI)	946	69%	31%	0%	Yes
PerplexityBot (Perplexity)	477	66%	34%	0%	Yes
Amazonbot (Amazon)	357	62%	2%	36%	Yes (via rDNS)
CCBot (Common Crawl)	317	0%	4%	96%	No

Full 18-crawler dataset (CSV and JSON) at unsourced.app/ai-crawler-report-2026.

Why "outside the range" is a flag, not a verdict

When a crawler's IP sits outside its operator's published list, that is worth flagging. On its own it is not proof of fraud. Published lists lag real infrastructure, so an out-of-range IP can be undisclosed operator infrastructure as easily as a spoofer. And the pattern matters: a spoof campaign clusters on a few addresses, while traffic spread thinly across hundreds of cloud IPs reads as infrastructure. So we report

"outside the published range" as exactly that, and reserve "fake" for the traffic that earns it. The figure worth acting on is the verified one.

Method. Every AI-crawler visit Unsourced recorded between 10 June and 6 August 2026 was resolved using two independent network proofs: whether the IP falls inside the operator's live-published ranges, and forward-confirmed reverse DNS (the IP's reverse record must resolve to the operator's domain and that host must resolve back to the same IP). Published-range coverage: OpenAI, Anthropic, Apple, Perplexity, Google. Aggregated by operator, not an internet-wide census.

© Unsourced · the evidence layer for AI search · unsourced.app/ai-crawler-report-2026 · Free to cite with attribution.