

Cloudflare Report: Googlebot Tops AI Crawler Traffic via @sejournal, @MattGSouthern

Cloudflare published its sixth annual Year in Review, offering a comprehensive look at Internet traffic, security, and AI crawler activity across 2025.

The report draws on data from Cloudflare's network, which spans more than 330 cities across 125 countries and handles over 81 million HTTP requests per second on average.

The AI crawler findings stand out. Googlebot crawled far more web pages than any other AI bot, reflecting Google's dual-purpose approach to crawling for both search indexing and AI training.

Googlebot Top AI Crawler Traffic

Cloudflare analyzed successful requests for HTML content from leading AI crawlers during October and November 2025. The results showed Googlebot reached 11.6% of unique web pages in the sample.

That's more than 3 times the pages seen by OpenAI's GPTBot at 3.6%. It's nearly 200 times more than PerplexityBot, which crawled just 0.06% of pages.

Bingbot came in third at 2.6%, followed by Meta-ExternalAgent and ClaudeBot at 2.4% each.

The report noted that because Googlebot crawls for both search indexing and AI model training, web publishers face a

difficult choice. Blocking Googlebot's AI training means risking search discoverability.

Cloudflare wrote:

"Because Googlebot is used to crawl content for both search indexing and AI model training, and because of Google's long-established dominance in search, Web site operators are essentially unable to block Googlebot's AI training without risking search discoverability."

Related: Complete Crawler List For AI User-Agents

AI Bots Now Account For 4.2% of HTML Requests

Throughout 2025, AI bots (excluding Googlebot) averaged 4.2% of HTML requests across Cloudflare's customer base. The share fluctuated between 2.4% in early April and 6.4% in late June.

Googlebot alone accounted for 4.5% of HTML requests, slightly more than all other AI bots combined.

The share of human-generated HTML traffic started 2025 at seven percentage points below non-AI bot traffic. By September, human traffic began exceeding non-AI bot traffic on some days. As of December 2, humans generated 47% of HTML requests while non-AI bots generated 44%.

Crawl-to-Refer Ratios Show Wide Variation

Cloudflare tracks how often AI and search platforms send traffic to sites relative to how often they crawl. A high ratio means heavy crawling without sending users back to source sites.

Anthropic had the highest ratios among AI platforms, ranging from approximately 25,000:1 to 100,000:1 during the second half of the year after stabilizing from earlier volatility.

OpenAI's ratios reached as high as 3,700:1 in March. Perplexity maintained the lowest ratios among leading AI platforms, generally below 400:1 and under 200:1 from September onward.

For comparison, Google's search crawl-to-refer ratio stayed much lower, generally between 3:1 and 30:1 throughout the year.

User-Action Crawling Grew Over 20X

Not all AI crawling is for model training. "User action" crawling occurs when bots visit sites in response to user questions posed to chatbots.

This category saw the fastest growth in 2025. User-action crawling volume increased more than 15 times from January through early December. The trend closely matched the traffic pattern for OpenAI's ChatGPT-User bot, which visits pages when users ask ChatGPT questions.

The growth showed a weekly usage pattern starting in mid-February, suggesting increased use in schools and workplaces. Activity dropped during June through August when students were on break and professionals took vacations.

AI Crawlers Most Blocked In Robots.txt

Cloudflare analyzed robots.txt files across nearly 3,900 of the top 10,000 domains. AI crawlers were the most frequently blocked user agents.

GPTBot, ClaudeBot, and CCBot had the highest number of full

disallow directives. These directives tell crawlers to stay away from entire sites.

Googlebot and Bingbot showed a different pattern. Their disallow directives leaned heavily toward partial blocks, likely focused on login endpoints and non-content areas rather than full site blocking.

Civil Society Became Most-Attacked Sector

For the first time, organizations in the “People and Society” vertical were the most targeted by attacks. This category includes religious institutions, nonprofits, civic organizations, and libraries.

The sector received 4.4% of global mitigated traffic, up from under 2% at the start of the year. Attack share jumped to over 17% in late March and peaked at 23.2% in early July.

Many of these organizations are protected by Cloudflare’s Project Galileo.

Gambling and games, the most-attacked vertical in 2024, saw its share drop by more than half to 2.6%.

Other Key Findings

Cloudflare’s report included several additional findings across traffic, security, and connectivity.

Global Internet traffic grew 19% year-over-year. Growth stayed relatively flat through mid-April, then accelerated after mid-August.

Post-quantum encryption now secures 52% of human traffic to Cloudflare, nearly double the 29% share at the start of the year.

ChatGPT remained the top generative AI service globally. Google Gemini, Windsurf AI, Grok/xAI, and DeepSeek were new entrants to the top 10.

Starlink traffic doubled in 2025, with service launching in more than 20 new countries.

Nearly half of the 174 major Internet outages observed globally were caused by government-directed shutdowns. Cable cut outages dropped nearly 50%, while power failure outages doubled.

European countries dominated Internet quality metrics. Spain topped the list for overall Internet quality, with average download speeds above 300 Mbps.

Why This Matters

The AI crawler data should affects how you think about bot access and traffic.

Google's dual-purpose crawler creates a competitive advantage. You can block other AI crawlers while keeping Googlebot access for search visibility, but you can't separate Google's search crawling from its AI training crawling.

The crawl-to-refer ratios help quantify what publishers already suspected. AI platforms crawl heavily but send little traffic back. The gap between crawling and referring varies widely by platform.

The civil society attack data matters if you work with nonprofits or advocacy organizations. These groups now face the highest rate of attacks.

Looking Ahead

Cloudflare expects AI metrics to change as the space continues

to evolve. The company added several new AI-related datasets to this year's report that weren't available in previous editions.

The crawl-to-refer ratios may change as AI platforms adjust their search features and referral behavior. OpenAI's ratios already showed some decline through the year as ChatGPT search usage grew.

For robots.txt management, the data shows most publishers are choosing partial blocks for major search crawlers while fully blocking AI-only crawlers. The year-end state of these directives provides a baseline for tracking how publisher policies evolve in 2026.

Featured Image: Mamun_Sheikh/Shutterstock

<https://www.searchenginejournal.com/cloudflare-report-googlebot-tops-ai-crawler-traffic/563303/>