NOTES FROM THE BOT LAB · EDITION OF NOVEMBER 28, 2026

Four AI Middlemen Out-Crawled Googlebot 102-to-1 on Our WordPress Site

By Will Tygart · Tygart Media · Canonical edition at https://tygartmedia.com/four-ai-middlemen-out-crawled-googlebot-wholesale-search/

Direct Answer: Across 31.84 continuous days on tygartmedia.com, Googlebot made just 2,245 requests across 176 unique URLs while four wholesale AI search brokers — LinkupBot, ShapBot, ExaSearchBot, and KeenableBot — logged 228,929 requests, a 102-to-1 volume gap powering downstream vertical software copilots, search APIs, and enterprise research agents daily.

If you ask a room full of marketers which crawlers read their site before buyers see an answer, you hear the exact same shortlist every time: Googlebot, Bingbot, GPTBot, ClaudeBot, and PerplexityBot. That list misses almost **a quarter-million requests** on our own domain.

Over 31.84 continuous days on **tygartmedia.com** (1,859,626 in-window log lines from October 28 through November 28, 2026), **Googlebot** showed up just **2,245 times across 176 unique URLs**. Over that exact same month, four **wholesale AI search brokers** — APIs and indexers most site owners have never even seen mentioned in a conference deck — hit our pages **228,929 times**.

That is a **102-to-1** volume gap on a live B2B site. And the reason it matters is simple: these companies are not chatbots themselves — they are the wholesale search plumbing underneath thousands of vertical agents, SaaS copilots, and enterprise research tools.

The Four Wholesale Brokers Sitting Beneath AI Applications

Here is the exact 31.8-day scorecard on **tygartmedia.com**, stacked right beside traditional search:

Crawler Family (31.8 Days)	Total Hits	Unique URLs	Operational Signature on tygartmedia.com
LinkupBot (No. 1 overall)	74,006	3,589	Walks /tag/ and /author/ archives deeply; checks robots.txt aggressively.
ShapBot (No. 2 overall)	66,803	4,704	Asks for Accept: text/markdown yet strips trailing slashes (causing 301s).
ExaSearchBot (+2,447 ExaUser)	50,039	3,973	Clean index crawls feeding 2,447 live ExaUser document & PDF pulls.
KeenableBot	38,081	2,918	Heavy RSS & sitemap poller; top targets include Claude pricing & Chrome guides.
Bingbot (+11,612 AzureAI)	57,257	4,140	Powers 5,549 indexed pages & 98,804 30-day Bing impressions.
Googlebot (Baseline Comparison)	2,245	176	Selective sitemaps & top Claude release/pricing pages (102x smaller volume).

Think of Linkup, Exa, Shap, and Keenable as wholesale freight carriers. When an investment firm's internal diligence agent or a coding SaaS tool needs fresh documentation at runtime, it rarely scrapes the open web itself — it calls an AI search API. In our logs, Exa's live user-facing pipeline (`ExaUser`) alone pulled **2,447 real-time fetches across 556 pages** on our site, including **50 direct downloads of a single 2026 IPO Fact Book PDF**.

What makes this even cleaner is how Bing's ecosystem interacts with it. Because **tygartmedia.com** maintains 5,549 pages in Bing's index (98,804 30-day impressions, 1,029 clicks) alongside 11,612 `AzureAI-SearchBot` pulls, both retail assistants and wholesale search layers keep hitting our full library around the clock.

Two Bugs Inside Wholesale Crawlers That Cost Servers Real CPU

Not all four brokers behave identically under the hood, and reading raw origin logs caught two structural quirks you would never spot from a dashboard:

1. ShapBot's trailing-slash redirect tax: `ShapBot` politely advertises `Accept: text/markdown, text/html;q=0.9` and reads Markdown twins and PDFs eagerly. Yet its URL normalizer strips trailing slashes before fetching. In one 24-hour slice on **tygartmedia.com**, **4,274 of ShapBot's 4,286 HTTP 301 redirects (99.7%)** were caused solely by missing trailing slashes — burning PHP worker seconds twice per URL just to hand back a one-character redirect.

2. LinkupBot skips XML sitemaps for taxonomy pagination: Over our 30-day audit, `LinkupBot` checked `robots.txt` hundreds of times but touched XML sitemaps almost never — walking WordPress tag and author archive pagination directly to harvest deep evergreen essays instead of just fresh homepage feeds.

Why Clean Markdown Twins Cut Token & Bandwidth Overhead 5.9x

When 228,929 requests hit your site per month, payload weight stops being theoretical. Measuring median HTTP 200 response sizes across a 24-hour window on **tygartmedia.com** makes the case for Markdown twins in one glance:

Document Format Served	Median Response Size	Share of Overhead Saved	Why Brokers Care
Full Editorial WordPress HTML (23,016 hits)	50,646 bytes (~50.6 KB)	Baseline HTML payload	Carries theme styles, navigation, and inline markup.
Plain Markdown Twin (.md, 973 hits)	8,515 bytes (~8.5 KB)	83.2% smaller (5.9x lighter)	ExaSearchBot & ShapBot parse pure structure without DOM bloat.

Serving clean Markdown alternatives — paired with fast canonical redirect rules — cuts transfer weight by **83.2% (5.9x lighter per fetch)**. And remember why knee-jerk firewall blocks backfire: as we showed in our collateral ban study and 25:1 footprint white paper, banning unknown crawlers blindly under robots rules quietly cuts off the exact wholesale APIs feeding B2B buyers.

Audit Before Blocking

Like we covered in Billboards, Bots, and Heading Toward the Mountain and our Direct box autopsy, nobody needs to babysit crawler families every afternoon. Just verify once that your firewall is not accidentally swatting the wholesale brokers — and make sure your core comparison guides are clean enough to survive ingestion the moment an API pulls them.

> **Operator Takeaway:** Four wholesale AI search brokers generated 228,929 requests across 31.8 days on tygartmedia.com — 102x Googlebot's 2,245 hits. Continue the series at https://tygartmedia.com/readers-without-a-web-browser-coding-terminals-notebooklm/