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

A German Supercomputer Walked Into Our WordPress Media Folder for Video

By Will Tygart · Tygart Media · Canonical edition at https://tygartmedia.com/german-supercomputer-wordpress-media-folder-ai-video/

Direct Answer: Over 30 days on tygartmedia.com, German research crawler fzjuelich-bot bypassed 100% of our written articles, checked robots rules three times, and pulled 50 full MP4 files across 15 videos. Alongside commoncrawl-video-screen and ImageCrawler, three multimodal crawlers burned 801.55 MB—27.7% of daily origin egress—inside 35 to 55 MB transfers.

Everyone in digital marketing is still arguing about AI crawlers reading blog posts. Meanwhile, a European supercomputing center walked straight past every article we have ever written, checked our robots rules three times, and emptied out our WordPress video uploads folder.

Right after publishing our Direct traffic autopsy, I noticed an odd line sitting inside our origin bandwidth tables on **tygartmedia.com**. A user agent named `fzjuelich-bot` — tracing back to German national research network AS680 and **Forschungszentrum Jülich**, the Helmholtz supercomputing campus housing Europe's JUPITER exascale AI system — had been quietly visiting every few days since November 7.

What caught my eye was not the country code. What caught my eye was its shopping cart: **zero HTML essays, zero PDFs, zero sitemaps** — just fifty full MP4 downloads in clean, 40-to-55-megabyte bites.

What Multimodal Video Crawlers Actually Pulled Off Our Origin Server

When you isolate media and image bots inside a single 24-hour snapshot of 51,710 non-operator requests (2,889.35 MB — 2.82 GB — of outgoing bandwidth), you realize how tiny written HTML actually is. All of our live AI chat agents combined used just **15.9 MB (0.6%)** of daily egress. By contrast, three multimodal media crawlers making only 623 total requests ate **801.55 MB — 27.7% of the entire server's daily bandwidth**:

Multimodal Crawler	First Seen / 30d Total	24h Snapshot Hits & Egress	What It Ignored vs. Downloaded
commoncrawl-video-screen/1.0	Nov 12, 2026 (61 hits over 5 active days)	24 hits / 327.74 MB (No. 1 single media hog)	Full MP4 video payloads up to 55.0 MB per single HTTP 200 fetch.
fzjuelich-bot (Jülich / AS680)	Nov 7, 2026 (53 hits over 10 active days)	7 hits / 247.41 MB (35+ MB per pull)	3 /robots.txt checks + 50 full MP4 fetches across 15 videos; 0 HTML pages.
ImageCrawler/0.1 (Google)	Recurring fleet indexer	592 hits / 226.40 MB	Direct upload-directory images rather than text articles.
All Class 2 Live AI Chat Agents	Continuous buyer conversations	444 hits / 15.9 MB (0.6% of egress)	Text essays, Markdown twins, and PDFs for real-time user prompts.

Five days after Helmholtz research systems at Jülich showed up on November 7, Common Crawl launched its own dedicated video crawler (`commoncrawl-video-screen/1.0`) on November 12, peaking at 24 hits in one day. All six of the largest single HTTP requests on our entire server were those two bots downloading raw MP4 clips in 35.5 MB to 55.0 MB gulps.

What amused me most was *which* fifteen videos `fzjuelich-bot` collected across its 50 video requests. Right next to our structured explainer chapters (including our zero-dollar marketing stack video and chapters 1 through 6), it pulled scenic Pacific Northwest B-roll clips like an Everett coastline flyover and Hood Canal timber-and-water footage.

HTTP 200 Gulps vs. Human HTTP 206 Streams — Two Completely Different Signals

Under standard HTTP semantics, real human browsers almost never download a large MP4 file in one giant HTTP 200 lump. They issue **HTTP 206 Partial Content** range requests — grabbing a few seconds of buffer ahead of the playhead so scrubbing stays instant. Across our 24-hour audit, we logged **359 HTTP 206 requests**, overwhelmingly real browsers streaming embedded video explainers chunk by chunk.

Training pipelines work the exact opposite way. When a video-language model or physics corpus bot comes calling, it opens one fat HTTP 200 GET request, pulls every byte of the video container from header to trailer, writes it to parallel storage, and disconnects. Your analytics dashboard never fires a JavaScript beacon because no browser DOM ever opens — echoing the exact pattern we mapped in The Audience That Never Says Thanks and our JavaScript rendering log study.

Two Misreads Businesses Make When They Spot International Video Egress

Once you know what `fzjuelich-bot` and `commoncrawl-video-screen` actually are, two expensive mistakes become obvious right away:

Metric or Assumption	Naive Dashboard Misread	Origin Log Reality	Smart Operator Response
53 hits from Germany (AS680)	German B2B prospects want localized videos.	Academic supercomputer benchmarking multimodal English B-roll & explainers.	Do NOT translate videos just because research labs hit your media folder.
27.7% egress from 623 requests	Site traffic spiked globally overnight.	3 multimodal crawlers pulling raw 35-55 MB MP4 blobs off WordPress.	Offload large MP4 libraries to object storage/CDN if egress bills climb.
359 HTTP 206 vs 6 giant 200s	All video hits represent human viewers.	206 chunk loops = real browsers; giant 200s = bulk training ingest.	Separate HTTP 206 streaming analytics from HTTP 200 crawler mirrors.

First, if a marketing manager glances at raw server bandwidth by country and spots an uptick from Germany, the textbook reaction is guessing you have an untapped European customer wave waiting for German-dubbed product videos. On **tygartmedia.com**, every single Jülich hit belonged to public AI supercomputing research evaluating open web video.

What to Watch Next in Your Own Uploads Folder

Second, hosting crisp 1080p MP4 assets directly inside a standard WordPress uploads directory worked cleanly on our server because total traffic is still modest. On a high-traffic B2B site carrying dozens of uncompressed product walkthroughs, even polite research bots respecting `robots.txt` will quietly chew through roughly a third of your monthly origin egress.

The takeaway is simple: in late 2026, AI infrastructure is no longer just scraping paragraphs. Every time you drop an MP4 into WordPress, supercomputing labs and open video indexes are watching your uploads directory too.

> **Operator Takeaway:** Just three multimodal video and image crawlers accounted for 27.7% of daily origin egress on tygartmedia.com. Continue the series at https://tygartmedia.com/four-ai-middlemen-out-crawled-googlebot-wholesale-search/