Bauer Media Group

Bot Management

Learnings from Bauer Media Group

Stefan Betzold
CPMO Bauer Media Group - Digital Publishing

Automated Traffic Surpassing Human Traffic

Data from Imperva Bot Report

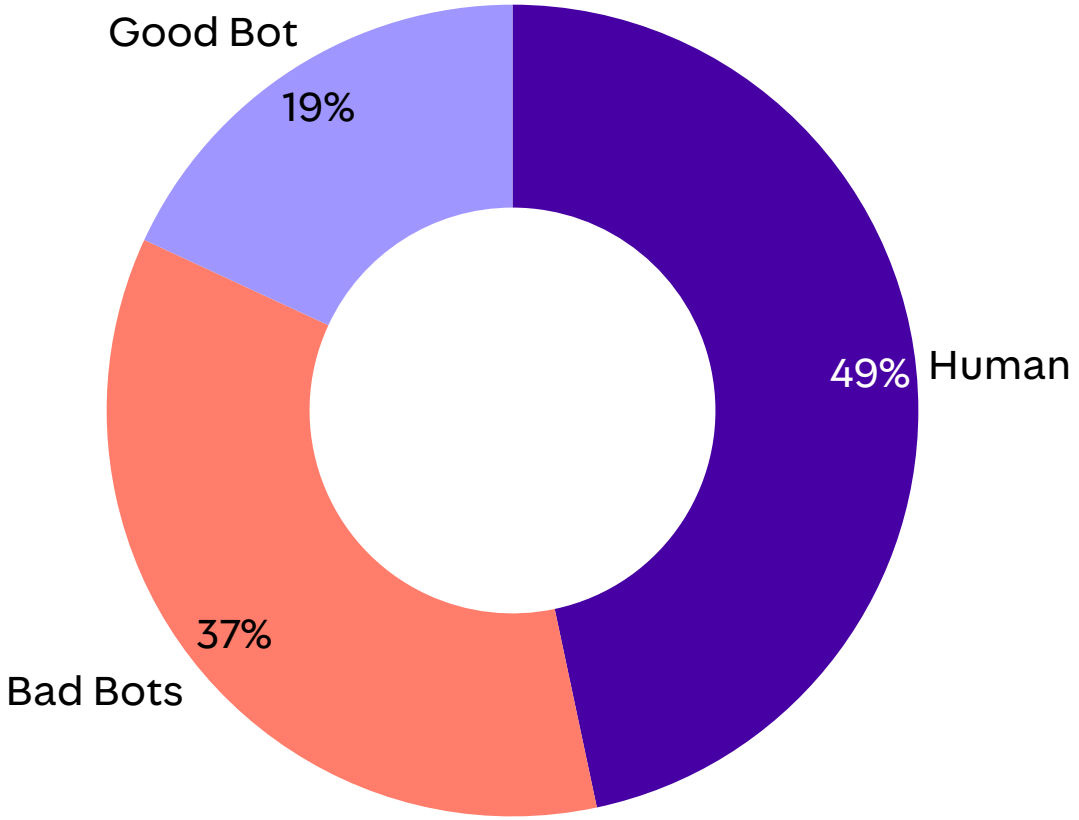
For the first time in a decade, automated traffic surpassed human activity.

Bot traffic is not new to us: Search-, SEO- and other crawlers exist for decades.

Growth is largely driven by rapid adoption of AI and LLMs.

Bad bot activity has risen for the 6th consecutive year, with malicious bots now making up 37%.

For us in **Publishing**, we have never really cared about bot protection in the past.
We have to change this now...

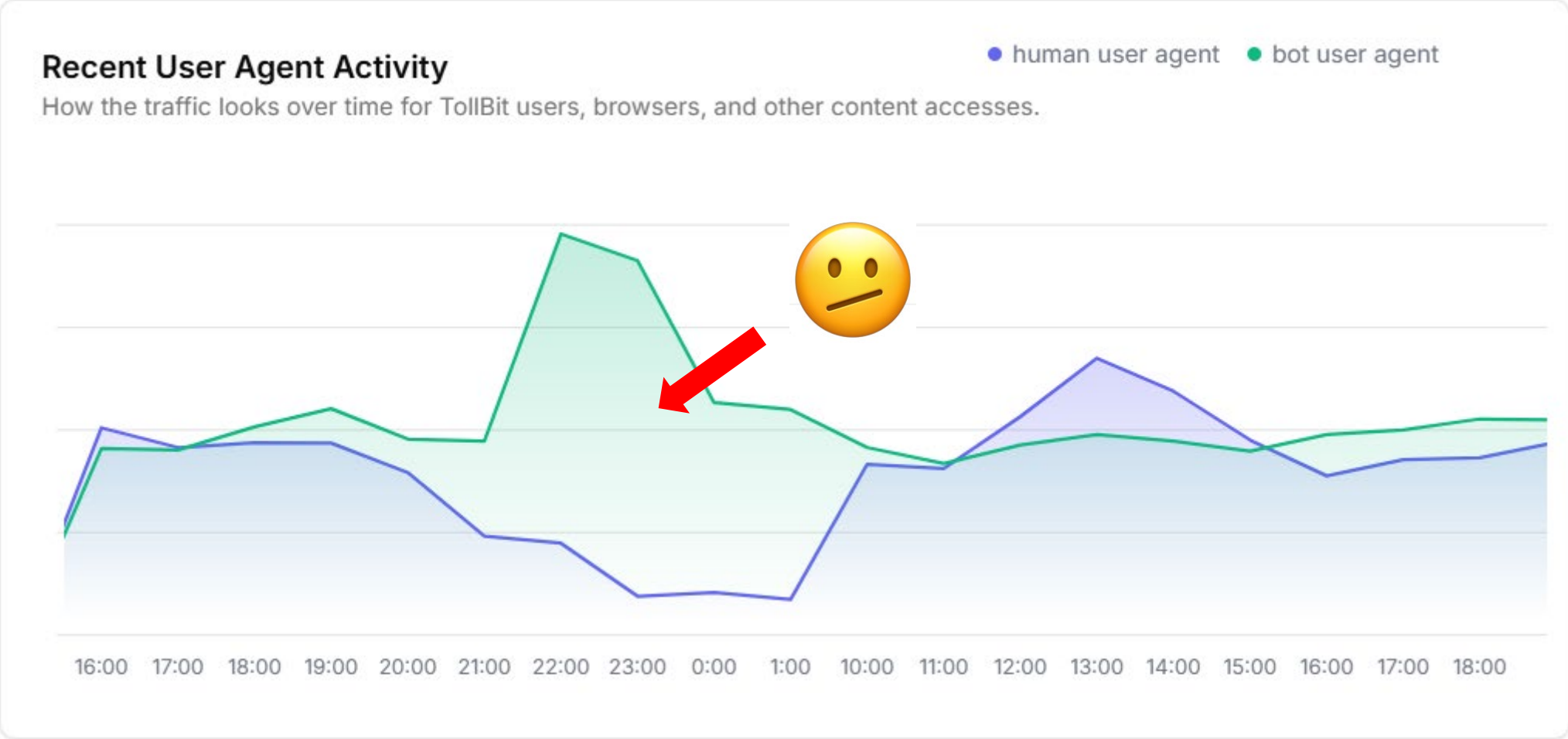


Global Internet Traffic
Imperva Bot Report 2025

imperva

Bot traffic surpassing human traffic

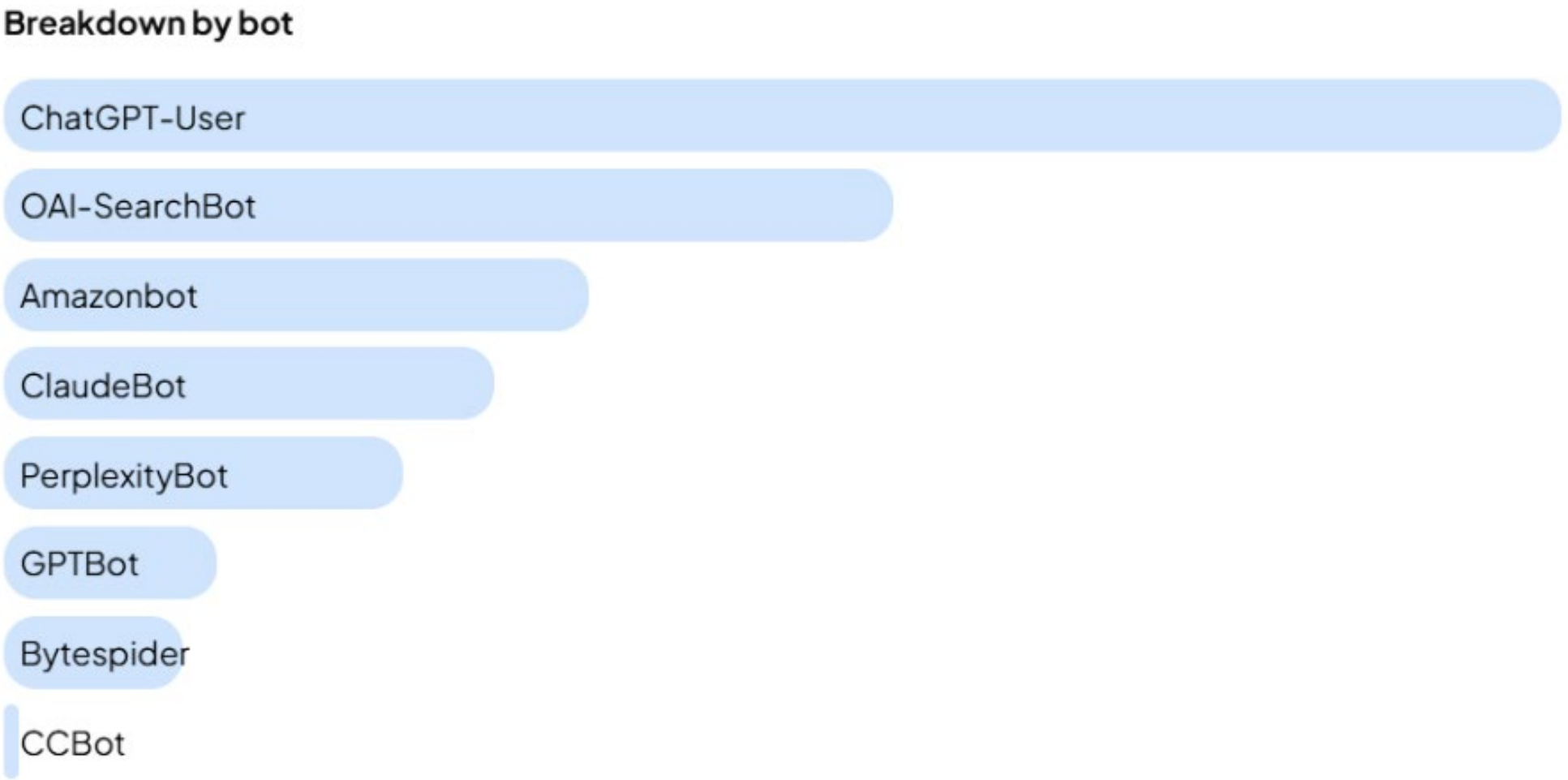
example from one Bauer Media brand in Q4



analytics powered by TOLLBIT

ChatGPT dominating with crawls from AI bot access

example from one Bauer Media brand in Q4



Biggest growth coming from smaller companies

example from one Bauer Media brand in Q4

1.3M
AI bot scrapes

+32.1% ↗
from previous

AI bots made **1.3M requests** to your website, **up 32.1%** from the previous

ClaudeBot Anthropic	+61.8% ↗ from previous	#4	←
PerplexityBot Perplexity	+230.5% ↗ from previous	#5	←

analytics powered by TOOLBIT

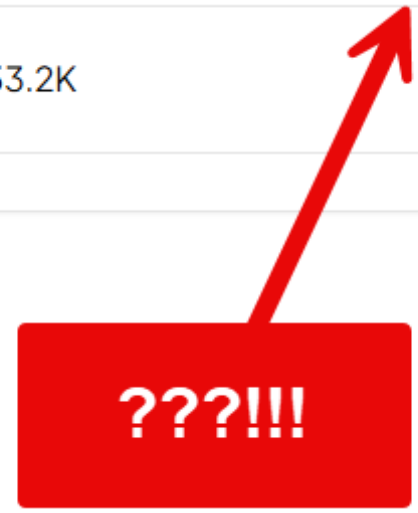
Clickthrough Rates from AI platforms on micro-level

example from one Bauer Media brand in Q4

Clickthrough Rates

How many human visitors AI platforms are sending you.

AI Platform	Referers	Scrapes	Clickthrough Rate
OpenAI (ChatGPT)	253	1.4M	0%
Perplexity	107	153.2K	0.1%



analytics powered by TOIIBIT

Not a fair deal: 500:1 to 2,800:1 crawls per referral

example from one Bauer Media brand in Q4

Referral traffic

AI companies brought **134 referrals** over the last month, **up 100.0%** from the previous period. **OpenAI** was the top referrer with **63 referrals**.

Scrape-to-referral ratio

How often AI bot scrapes result in a human visit.

134
Total AI referrals

361.4K
Total scrapes

2696:1
For every 2696 scrapes you get 1 AI referral

Top referrers

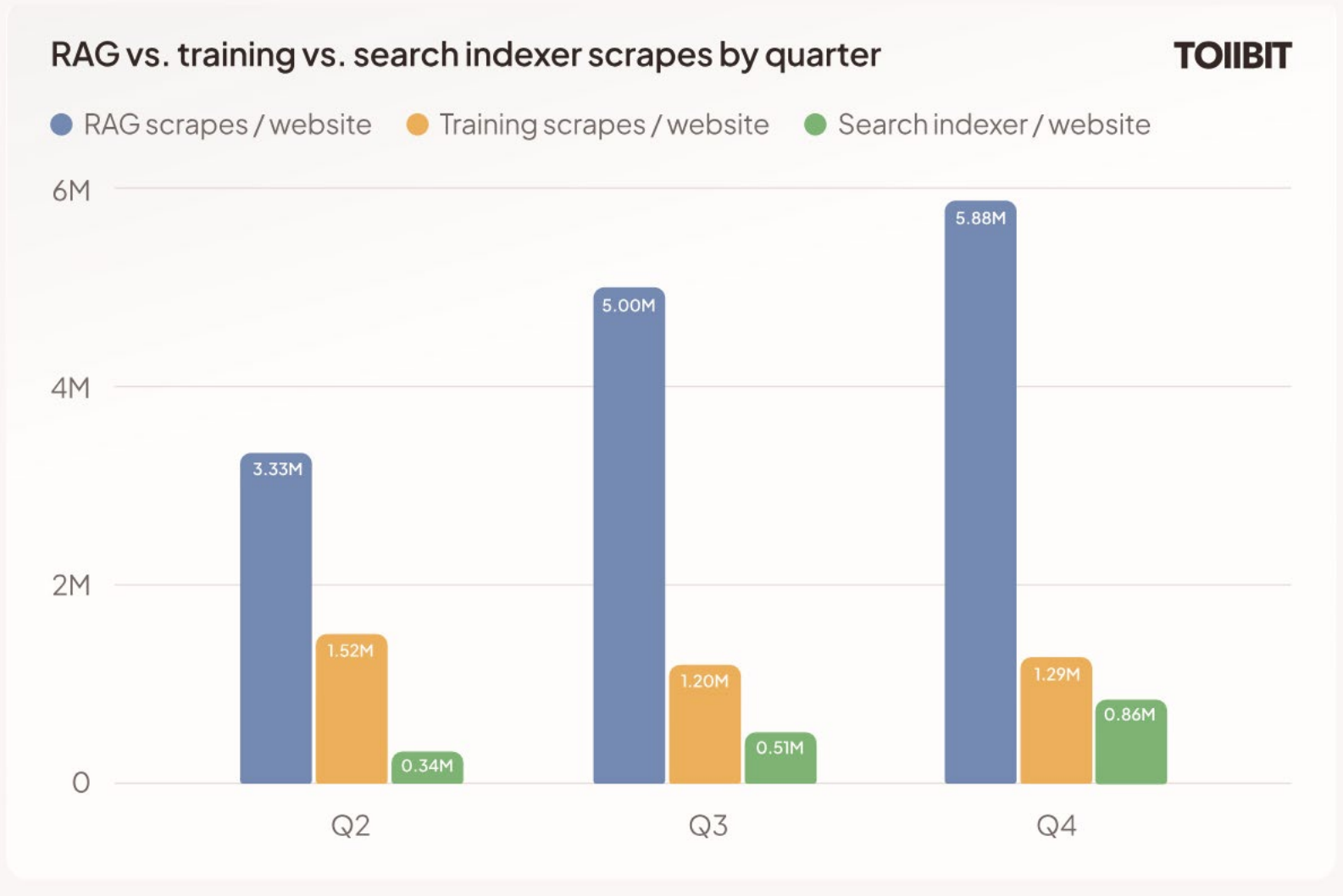
AI companies sending the most traffic to your site.

OpenAI	**63** referrals
Perplexity	**42**
Microsoft	**28**
Anthropic	**1**

analytics powered by TOIIBIT

RAG crawling massively increasing against Training

data from Tollbit state of the bot report



In Q4, the average scrapes per page for a RAG bot was ~10× more than for a Training bot

Robots.txt – is only a guidance and often ignored

example from one Bauer Media brand in Q4

Bots bypassing

AI bots ignored your robots.txt **51.2K** times over the last month. **Meta** bypassed the most with **43.0K** attempts.

See all bypassing data →

AI bots ignoring your robots.txt

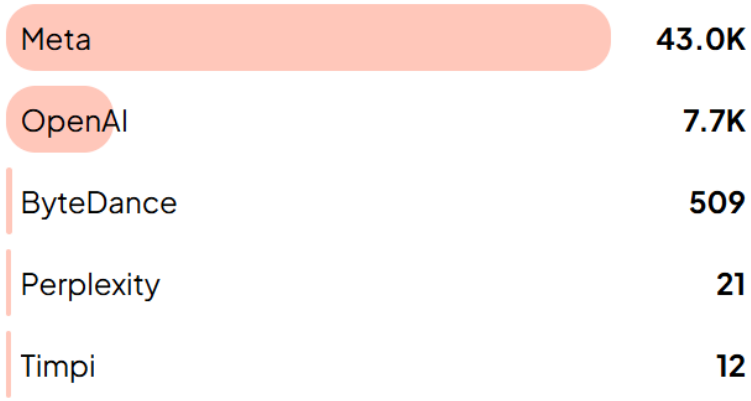
AI bots blocked in your robots.txt are still accessing your content. Set up bot paywall now

51.2K

times listed AI bots accessed your content last week

Top bypassing companies

The AI companies that ignored your robots.txt the most



analytics powered by TOILBIT

Publisher adoption of robots.txt

Miso.AI global robots analysis

22

average robots.txt publisher block list

Only 40% had bot block, 42% didn't block any bots.

43

Bauer Media Robots.txt blocklist

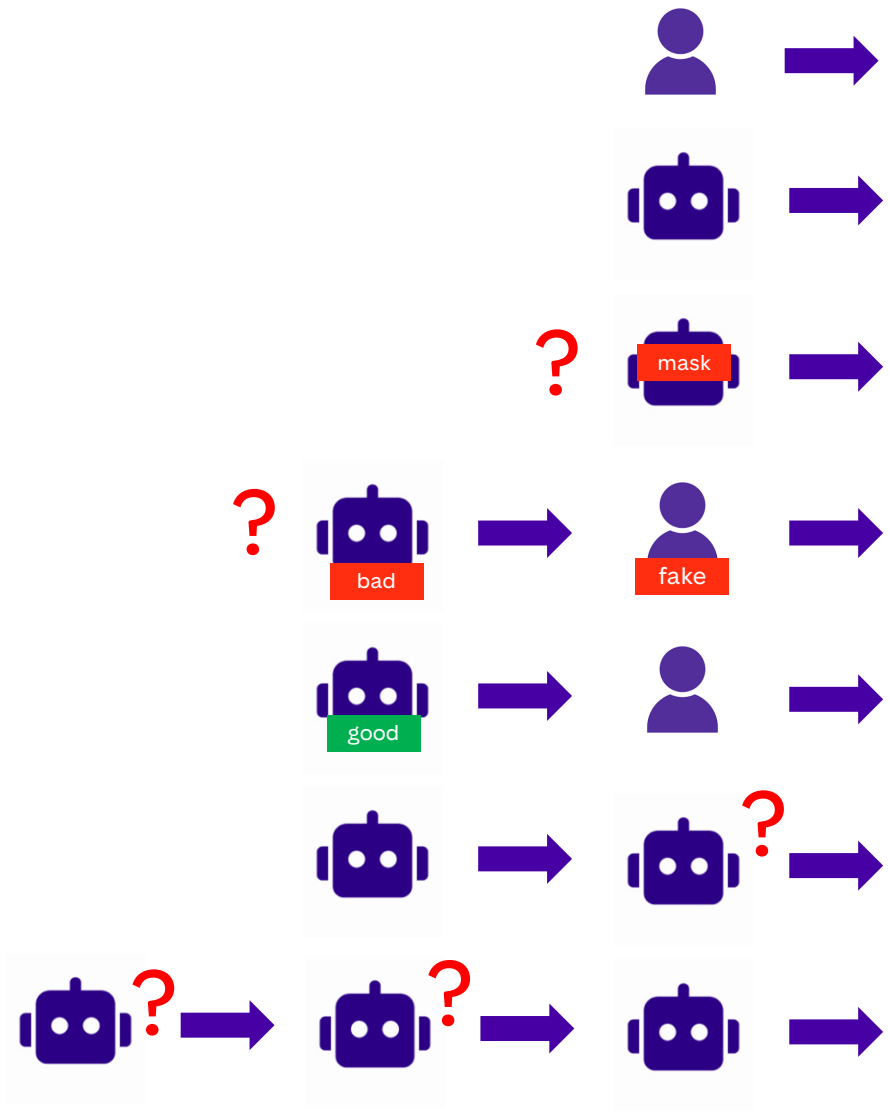
2.200

Bots accessing publisher websites

Source: Miso's Global Robots.txt Analysis (Oct 2025)
11.500 publishers analyzed.

miso

And it´s not about simple crawling bots any more!



The growing complexity of non-human access:

Direct crawling and scraping
with identification (e.g. user agent)

Masqued crawling
undeclared or using fake user agents

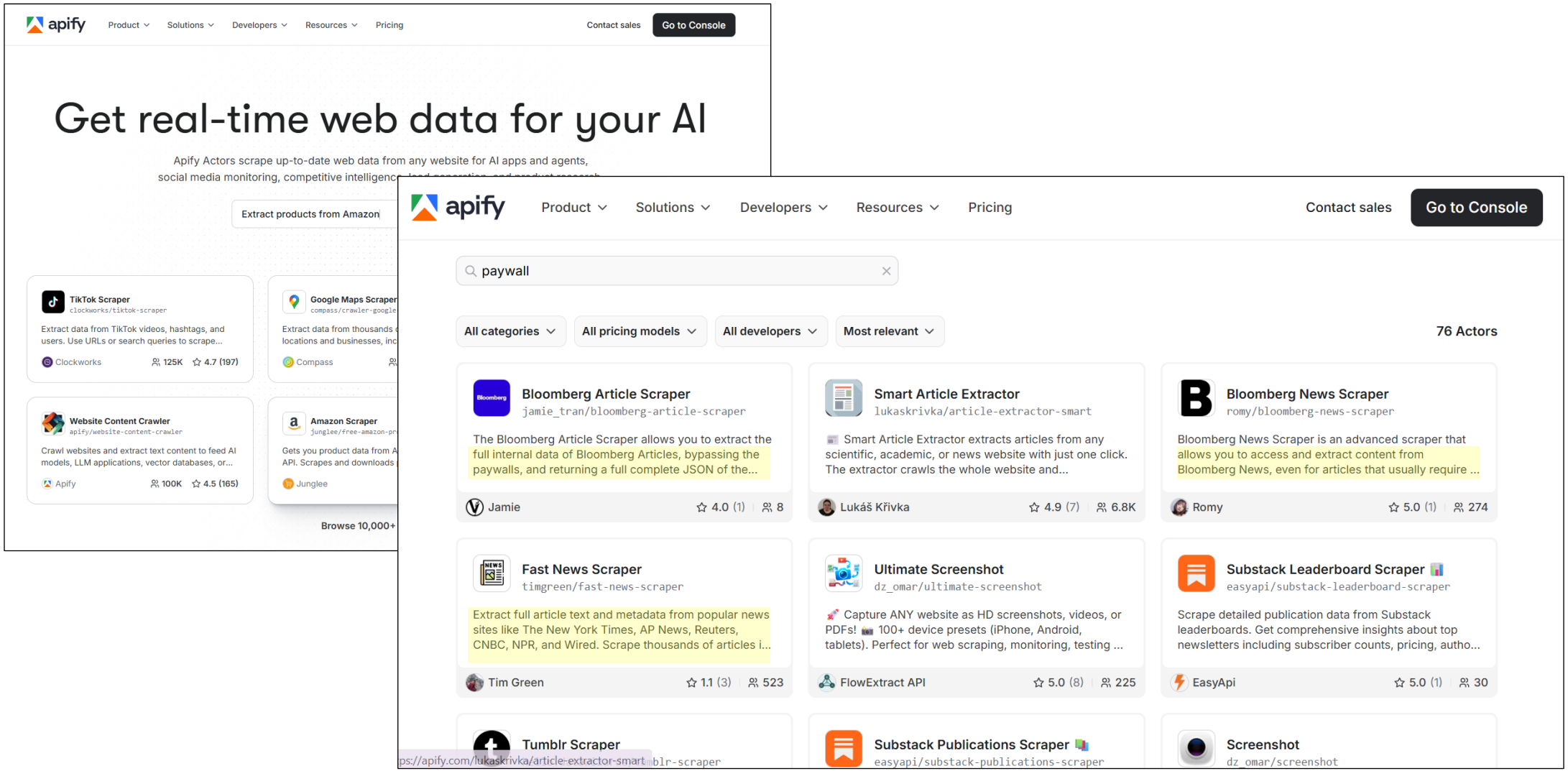
Third-party scraping services
for outsourcing the scraping process

Residential IP proxies, presenting as ordinary user traffic, that match human audience.

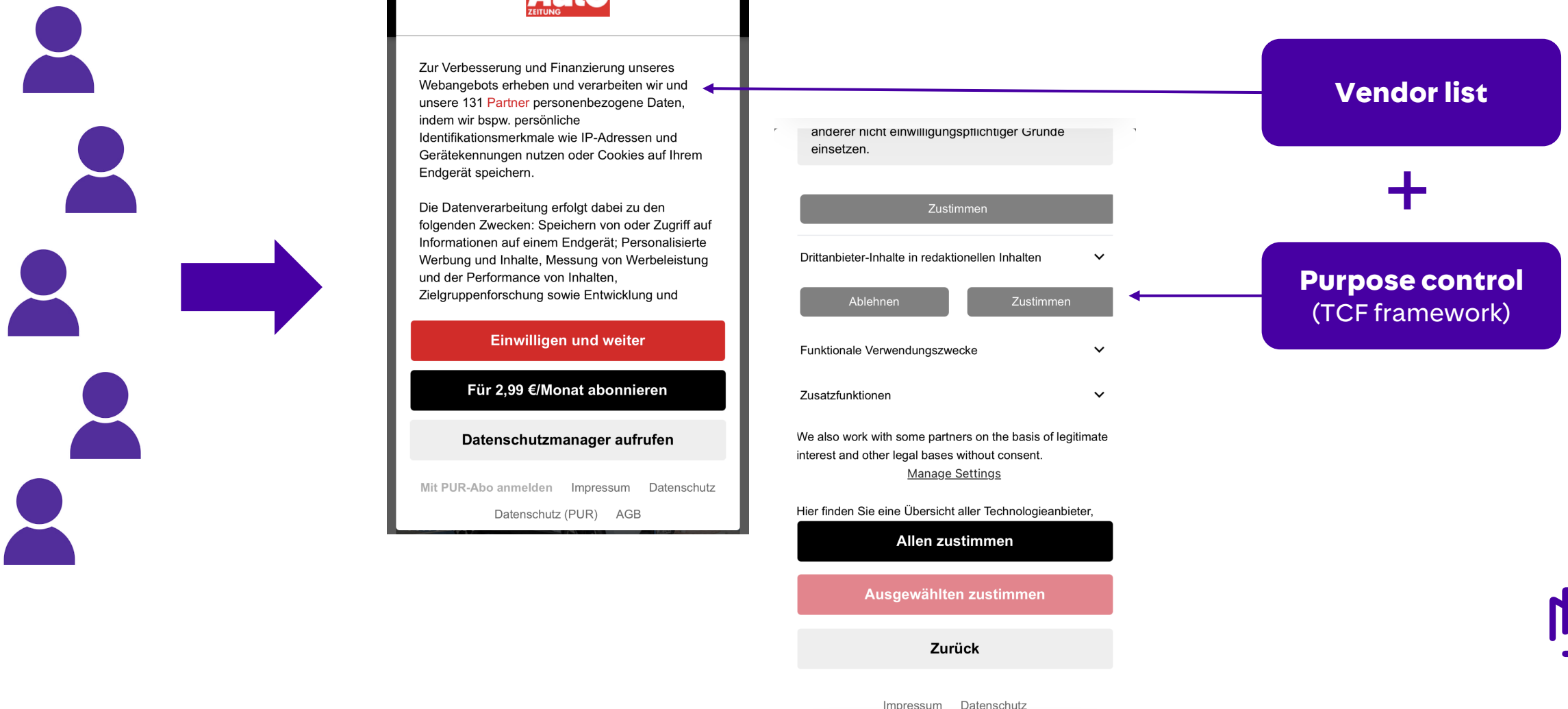
Circumvention services that evade bot detection / IP protection measures.

Cloud-based headless browsers
from new AI browsers, often used with a residential IP proxy, making the request appear to come from a normal human user.

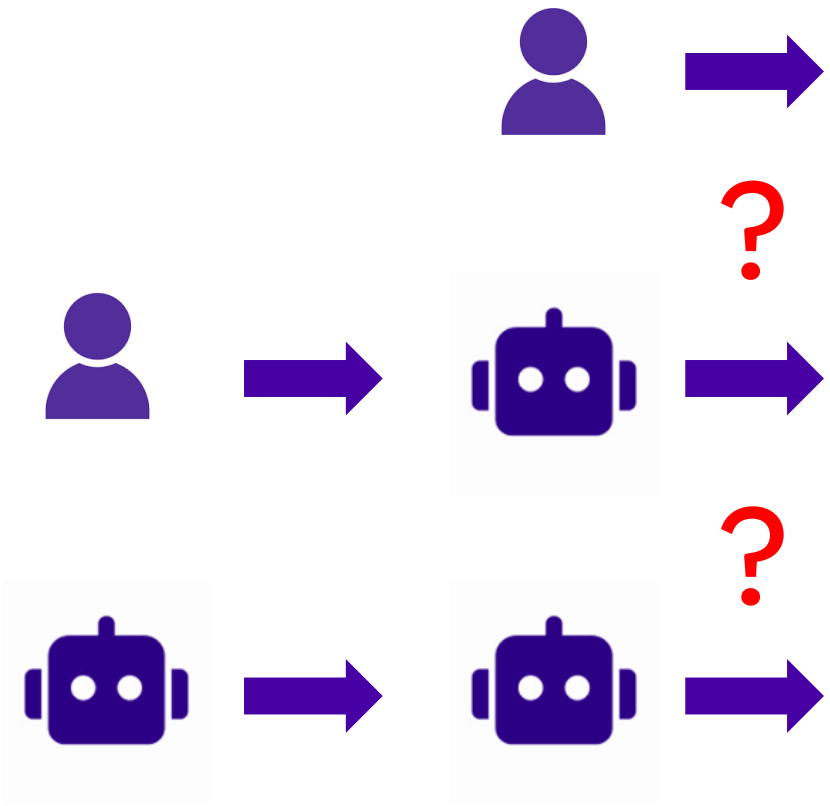
Easier than ever to order an AI crawling agent to extract contents and data from any website



Human access: We implemented granular access and data control for Human traffic with TCF framework



Non-Human access: We cannot identify and control bot and agent traffic



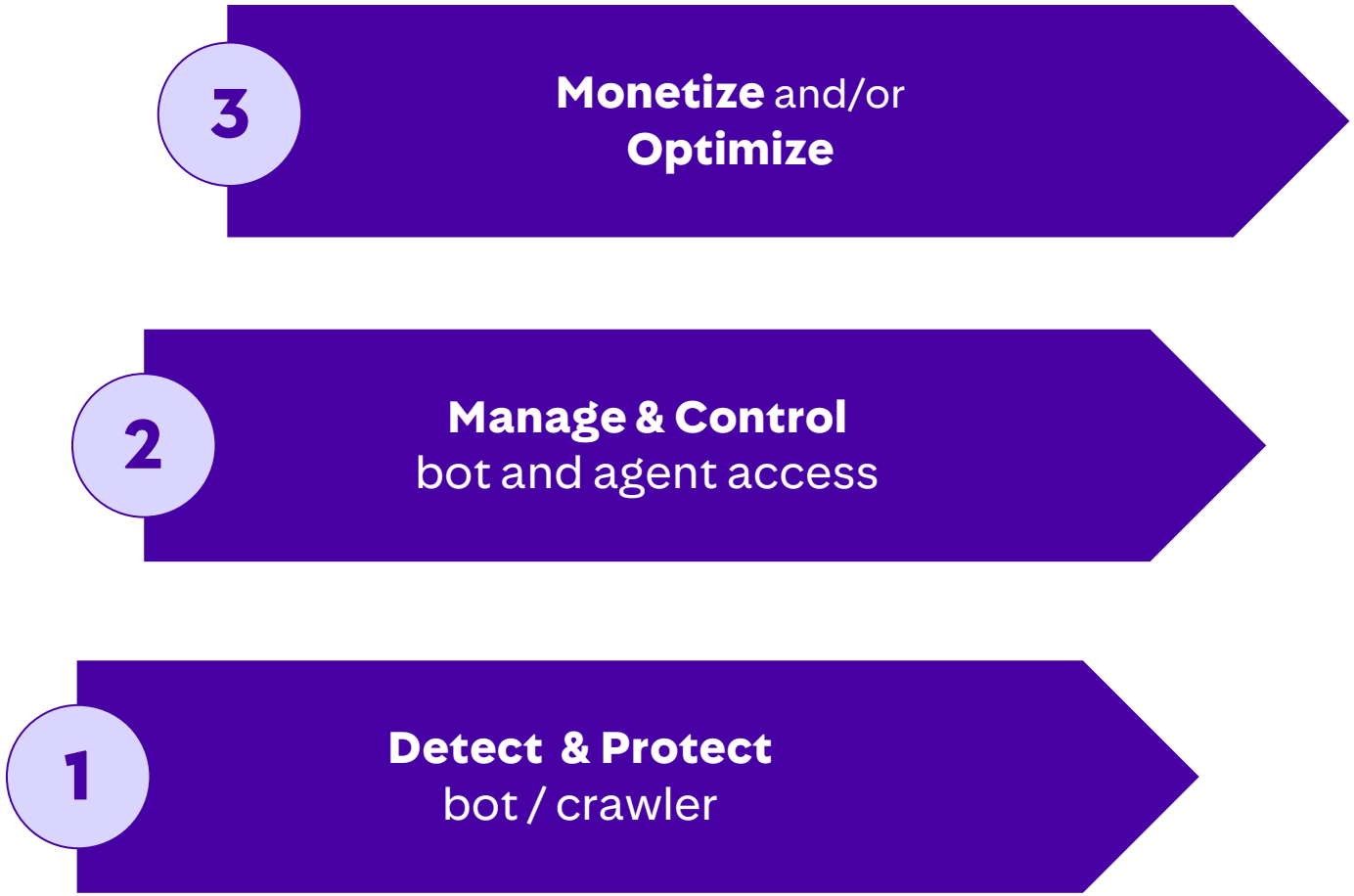
No Vendor identification

- No standards for bot / agent identification
- Bots not respecting robots.txt
- Undisclosed scraping / headless browsers
- Data brokers and third-party scraping outfits

No purpose control

- No differentiation of e.g. Search Index, RAG, training, transaction
- Googlebot runs search and RAG for AI/O and AI mode
- Good actor agents cannot be differentiated from bad actors

What we should do



What we should do

Policy & Standardization & Governance

3 **Monetize** and/or **Optimize**

2 **Manage & Control** bot and agent access

1 **Detect & Protect** Bot / Crawler

Content Transformation & Delivery

Detect non-human access and protect from unwanted crawling

1 Detect & Protect

Detect and classify automated traffic (AI crawlers, scrapers, agents) vs. human users.
Invest in bot protection (before you start thinking about monetization)

Action plan:

- Update and actively manage your robots.txt file
- Measure and analyze bot access and monitor closely (add bot-specific reporting to business dashboards)
- Evaluate bot protection tools: via CDN partner or 3rd party solution
- Start protecting! Turn on edge-level bot classification

Partnering options:

- CDN / Edge security: Cloudflare, Akamai, Fastly,…
- Specialist bot mitigation: Centinel Analytica, Datadome, Human, Imperva,…

Policy & Standardization & Governance

Support the standardization efforts:

- Lobby work as part of association work (e.g. MVFP), Brussel needs to understand!
- New protocol standards: W3C TDMRep, IETF AI PREF group

Start setting rules: move from “anonymous crawling” to “permissioned access”

2 Manage & Control

Create granular controls to manage bot and agent access for different purposes.

Action plan:

- Evaluate your “highest value contents” to be protected
- Define content access tiers: **granular controls** and rule sets by
  - content type (news, evergreen, paywalled, archives…),
  - bot purpose (training, summary, grounding, display…)
  - content use categories (free / snippet / license / premium / neverAllowed)
- Provide controlled endpoints (instead of letting bots hit the human website).
- Policy, rights, and governance (the “rules” layer)

Partnering options:

- Connect to bot management / cybersecurity providers
- Emerging landscape of new players: e.g. tollbit. But a lot of room for innovation!

Standards:

- Protocol standards: auth/token access patterns, llms.txt “AI-ready guidance”, RSL

Optimize and monetize: Convert legitimate demand into revenue

3 Optimize & monetize

Prepare content transformation into a cleaned-up, for bots optimized format.
Transform bot/agent traffic from cost into revenue; negotiate, meter, and license access.

Action plan:

- Reduce cost-to-serve with content transformation („reduce crawling costs“)
  - Structured feeds/endpoints for licensed access (e.g. MCP)
  - Clean content cache” for agents (faster, cheaper, controlled)
- Decide monetization routes: Direct deals, vs. Marketplace vs. Pay-per-use models
- Package your offer in commercial products (easy to sell + easy to buy), e.g.: “Licensed RAG / grounding access” (usage-based) vs. “Summaries + attribution” bundle“ vs. Full-text display” premium tier (strict terms)
- Explore and test new user propositions in conversational web.

Content Transformation & Delivery

Partnering options:

- Marketplaces: Tollbit, CloudFlare pay-per-crawl, Microsoft PCM, Prorata
- Make your own deal with LLMs, but this will be tough…

Standards:

- IAB Tech Lab ComP, RSL (really simple licensing), Nlweb

What´s next?

- **For AI companies**
  “Treat content as strategic infrastructure - not free raw material.”
- **For publishers & creators**
  “Block and manage AI bots, demand transparency, and build licensing standards.”
- **For regulators**
  “Create fair, simple rules that reward high-quality, diverse content - before it disappears.”

Bauer Media Group

Thank you!