



The Silent Infrastructure Crisis

AI Bots and the Future of Web Performance

2025-10-15

Chris Buckley
Principal Sales Engineer

The AI Bot Reality

AI Bots Are Reshaping Web Traffic

- Fastly protects **130,000+ applications**
- Inspects **6.5 trillion requests monthly**
- AI bots specifically target **Commerce** and **Media & Entertainment**

Your Origin Servers Are Under Siege

The Numbers That Matter

1 K

RPM

Peak crawler rate on
website

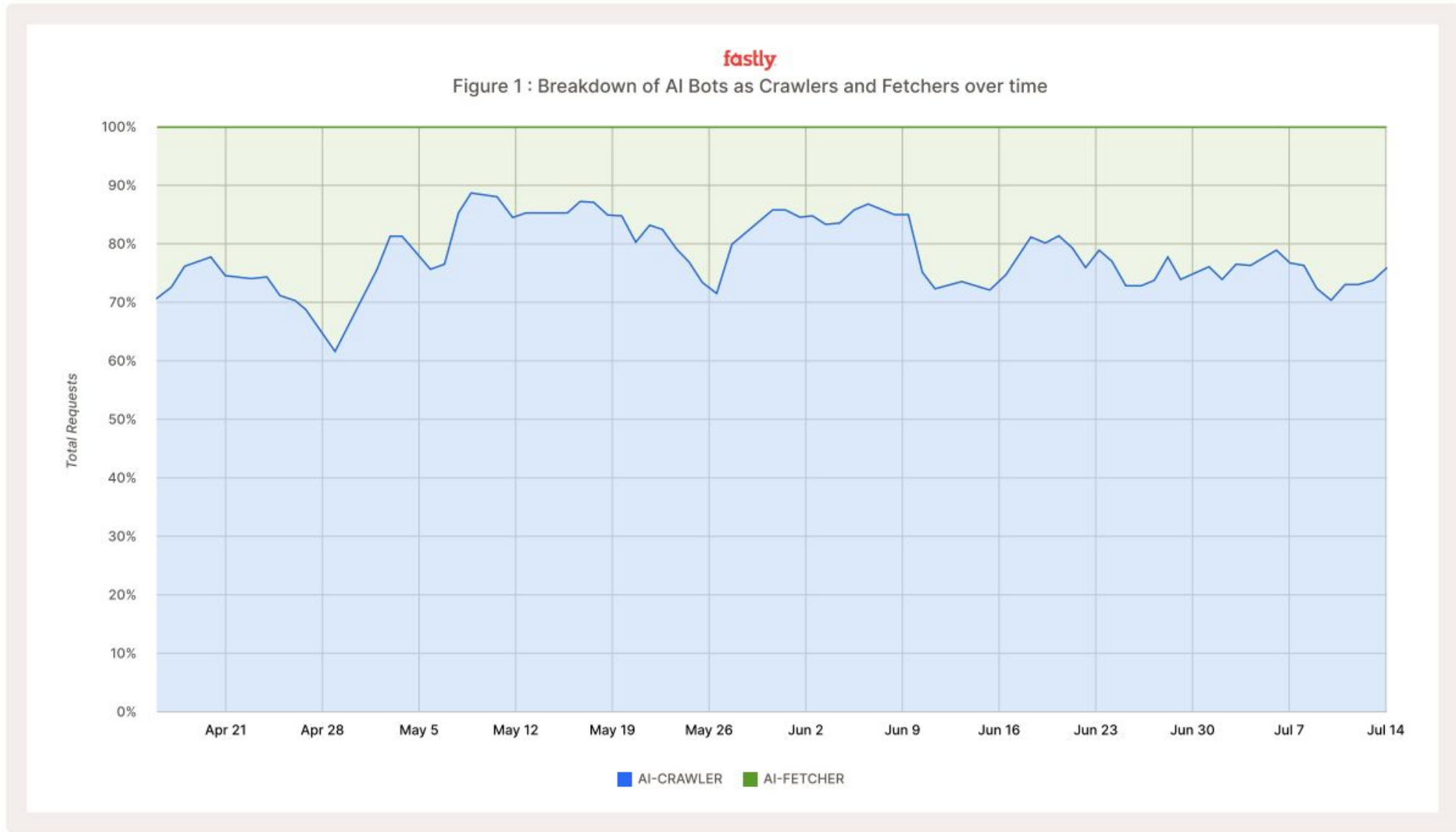
39 K

RPM

Peak fetcher rate on
website

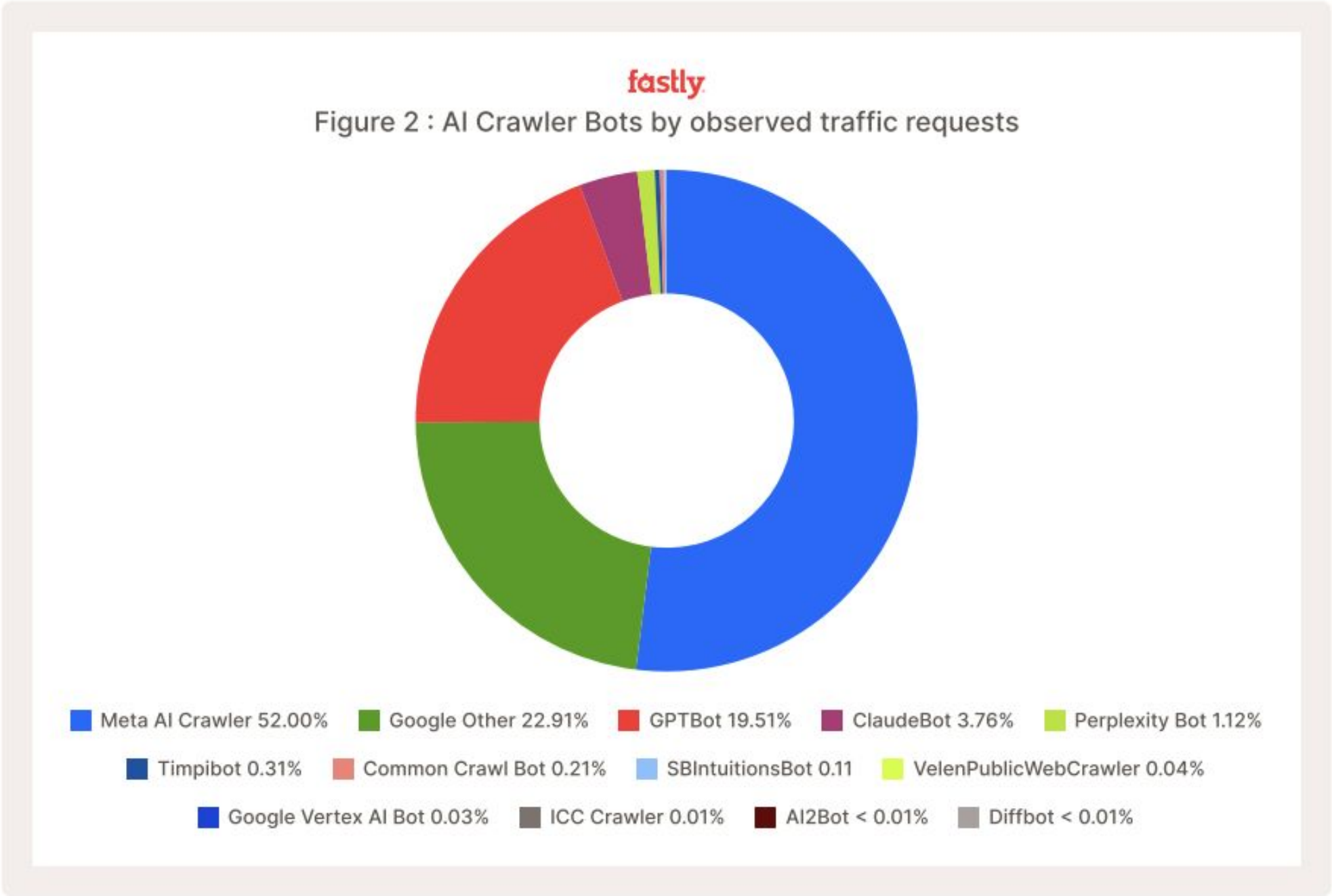
AI Bot Traffic: Crawlers vs Fetchers

Traditional "good vs bad" bot classification is no longer adequate



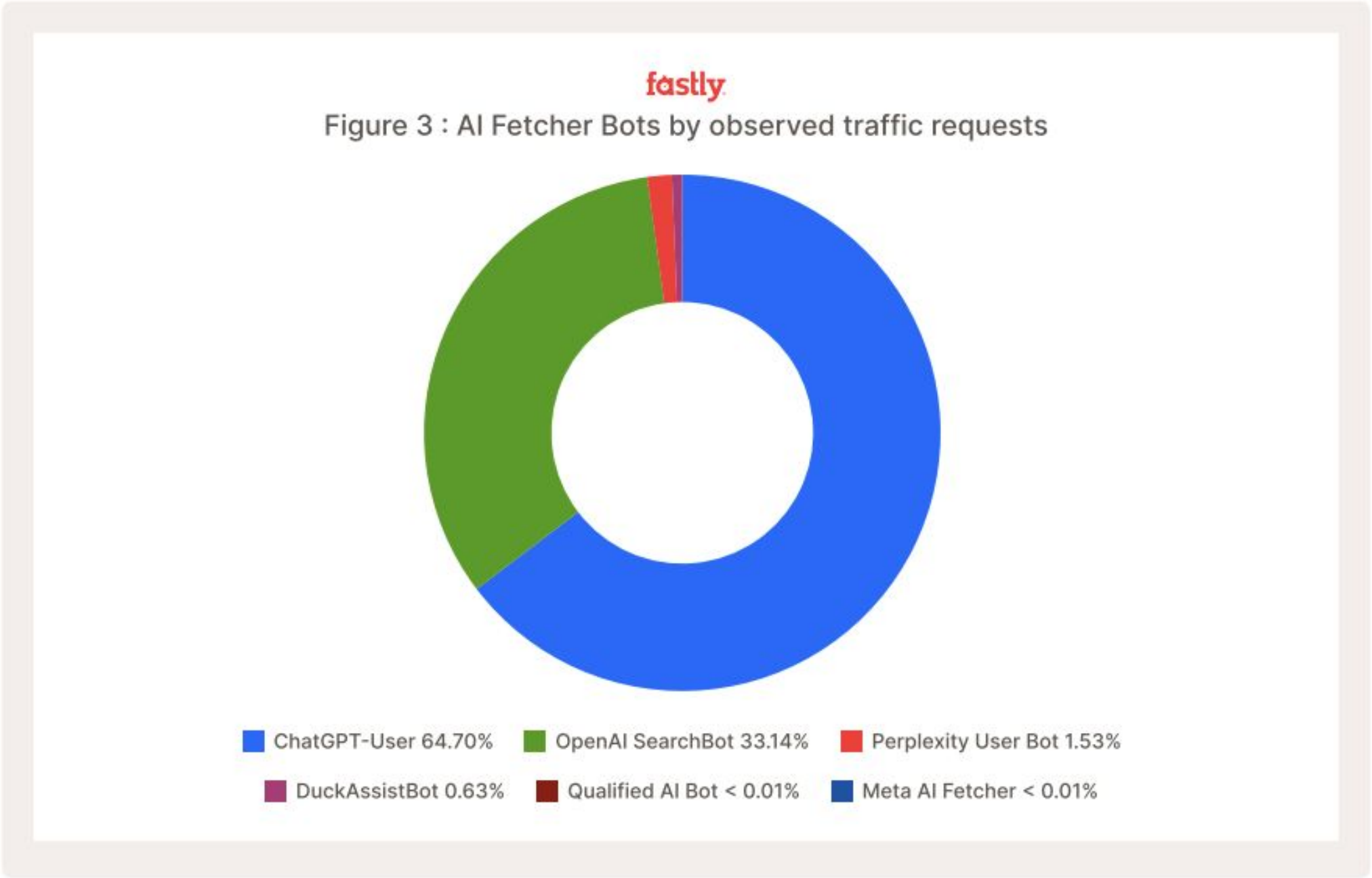
Who Controls AI Crawler Traffic?

Three companies control virtually all training data collection



Real-Time Fetching: Who's Hitting Your Site?

Unpredictable, bursty traffic patterns



What Sites Are Prime Targets?

Highest targeted verticals / industries

- **Commerce** (product listings, reviews)
- **Media & Entertainment** (news, articles, fresh content)
- **High Tech** (documentation, technical content)

What AI Bots Cost You

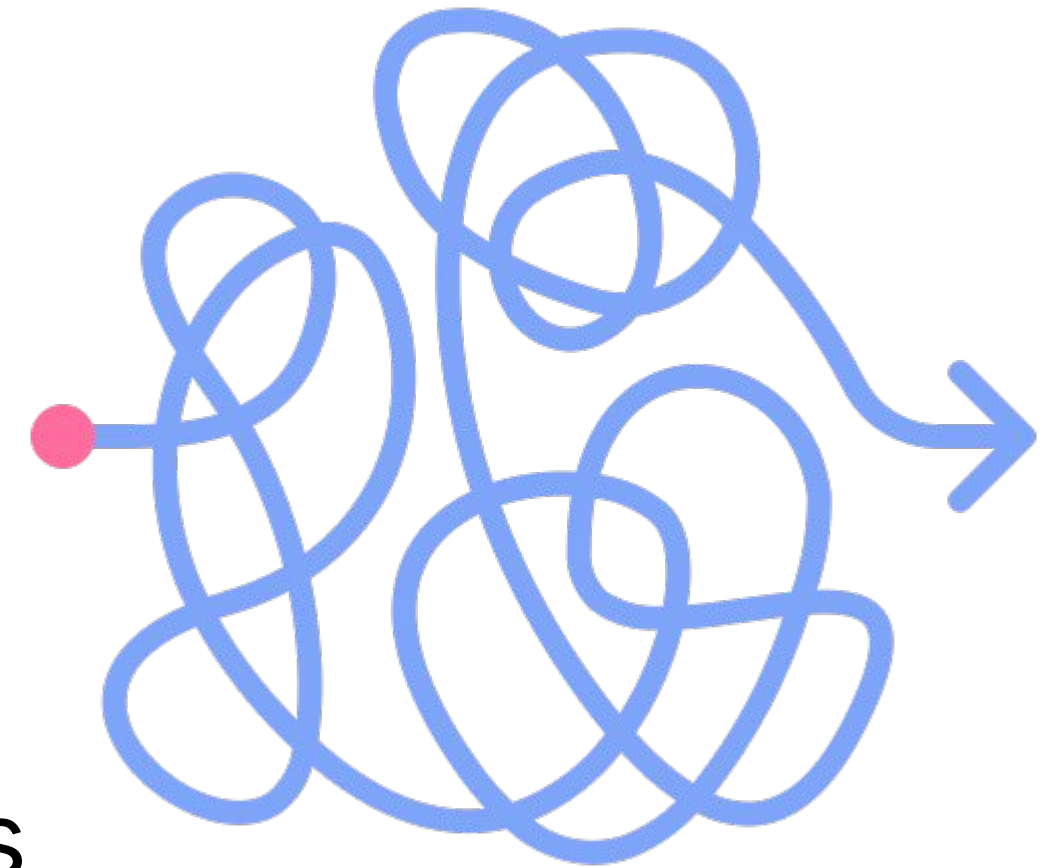
What AI Bots Cost You

- Infrastructure **scaling expenses**
- CDN/Cloud **bandwidth overages**
- Analytics **pollution**
- SEO **performance impact**



Meta vs Common Crawl

- **Common Crawl (CCBot):**
 - ✓ **Predictable** 2-week crawl cycles monthly
 - ✓ **Steadily increasing** volume
 - ✓ Can be **planned for**
- **Meta Crawler:**
 - × **Irregular** bursts: 2-3x baseline for days/weeks
 - × **No advance warning**
 - × **No coordination** with site operators



Can You Trust the Bot Identity?

Verifiable vs. Unverifiable Bots:

- **Verifiable Bots:**
 - **Publish IP ranges** (OpenAI) -
 - Support **reverse DNS** verification - Can be authenticated
- **Unverifiable Bots:**
 - **No public verification** mechanism
 - **Spoofing** risk
 - Cannot distinguish **legitimate from malicious**



The Googlebot Problem

Major search engines use the same crawlers for both search indexing AND AI training.

None

User-agent: GPTBot

Disallow: /

User-agent: Googlebot

Allow: /

Most Dangerous Pattern: Inference-Time Amplification

- **The Scenario:**

- 1. **Breaking news** event occurs
- 2. Thousands of users ask ChatGPT about it **simultaneously**
- 3. ChatGPT's fetcher bots hit news sites for **real-time grounding**
- 4. Traffic spike: **39,000 req/min observed**
- 5. Origin doesn't have **time to scale**
- **6. Result: DDoS-like effects without malicious intent**

Regional Variations Matter

AI Bot Traffic Varies by Region

- North America: **90% Crawlers** | **10% Fetchers**
- EMEA: **41% Crawlers** | **59% Fetchers**
- APAC/LATAM: **Mixed patterns**

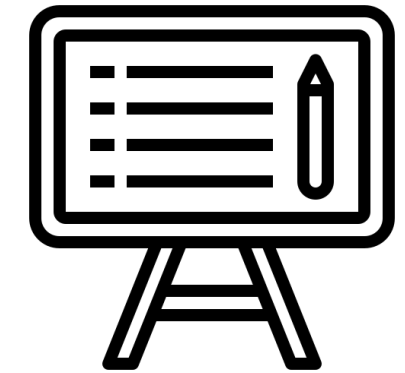


*Implication: **Different regions** require different bot management strategies*

Different Industries, Different Problems

- **High-Risk Crawler Verticals (>90%)**

- **Commerce**: 92% crawler -
- **Healthcare**: 92% crawler -
- **Public Sector**: 96% crawler -



- Risk: **Sustained high-volume** scraping of product data, medical content, government information

Different Industries, Different Problems

- **Fetcher-Heavy Verticals: -**

- **Education**: 68% fetcher (students using ChatGPT) -
- **Media & Entertainment**: 54% fetcher (news, current events) -



- Risk: **Unpredictable burst traffic** during high-interest events

Your Action Plan

Three-Layer Defense Strategy

- Layer 1: Web Standards
 - **robots.txt** directives
 - **HTTP headers** & **meta tags**
 - Works with **"polite" bots only**

```
1 User-agent: GPTBot
2 Disallow: /
3
4 User-agent: Google-Extended
5 Disallow: /
6
7 User-agent: ClaudeBot
8 Disallow: /
9
10 User-agent: googlebot
11 Allow: /
```

Your Action Plan

Three-Layer Defense Strategy

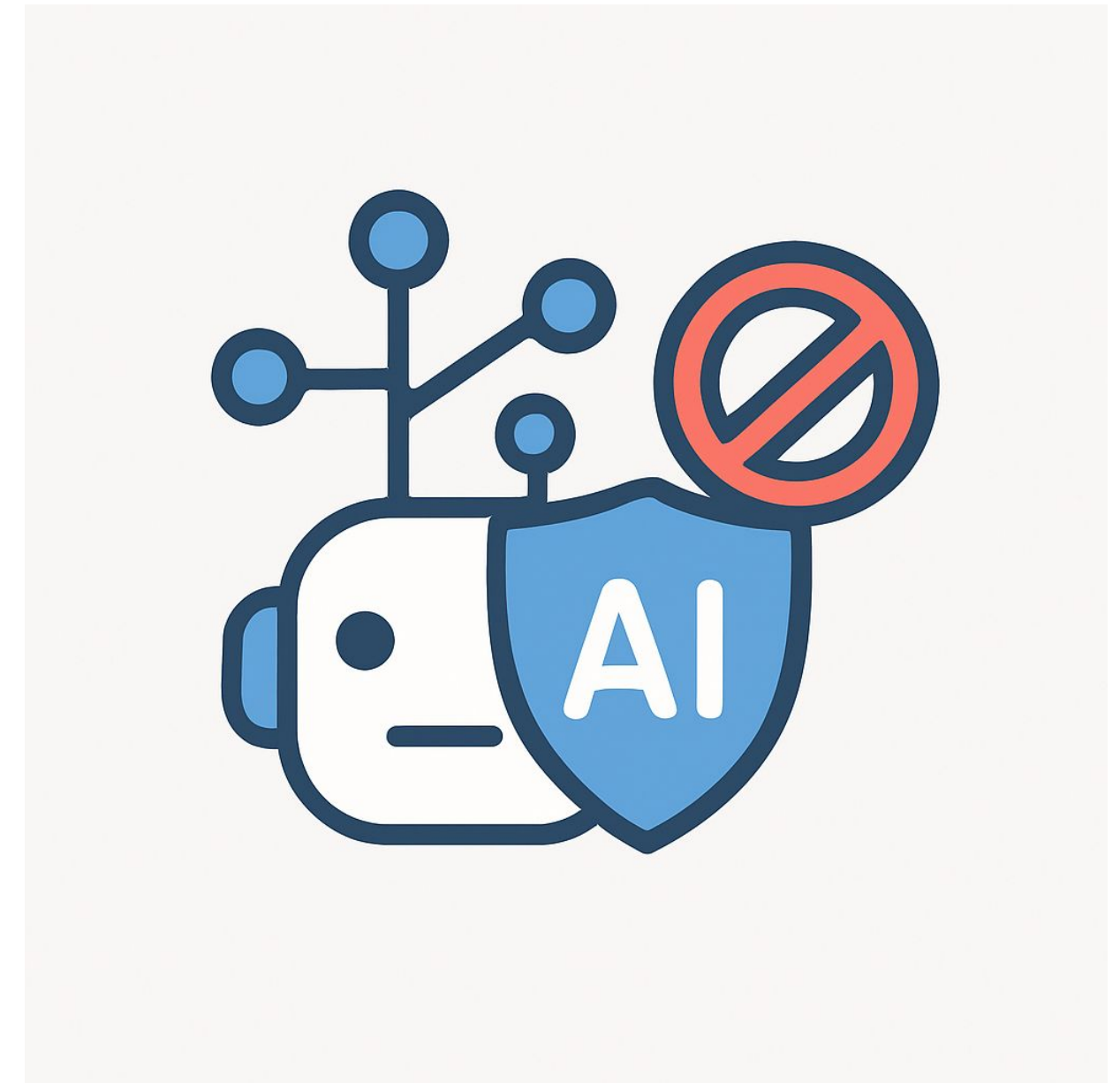
- Layer 2: Technical Controls
 - **Rate limiting** (critical for fetchers)
 - IP-based **blocking**
 - Progressive **challenges**



Your Action Plan

Three-Layer Defense Strategy

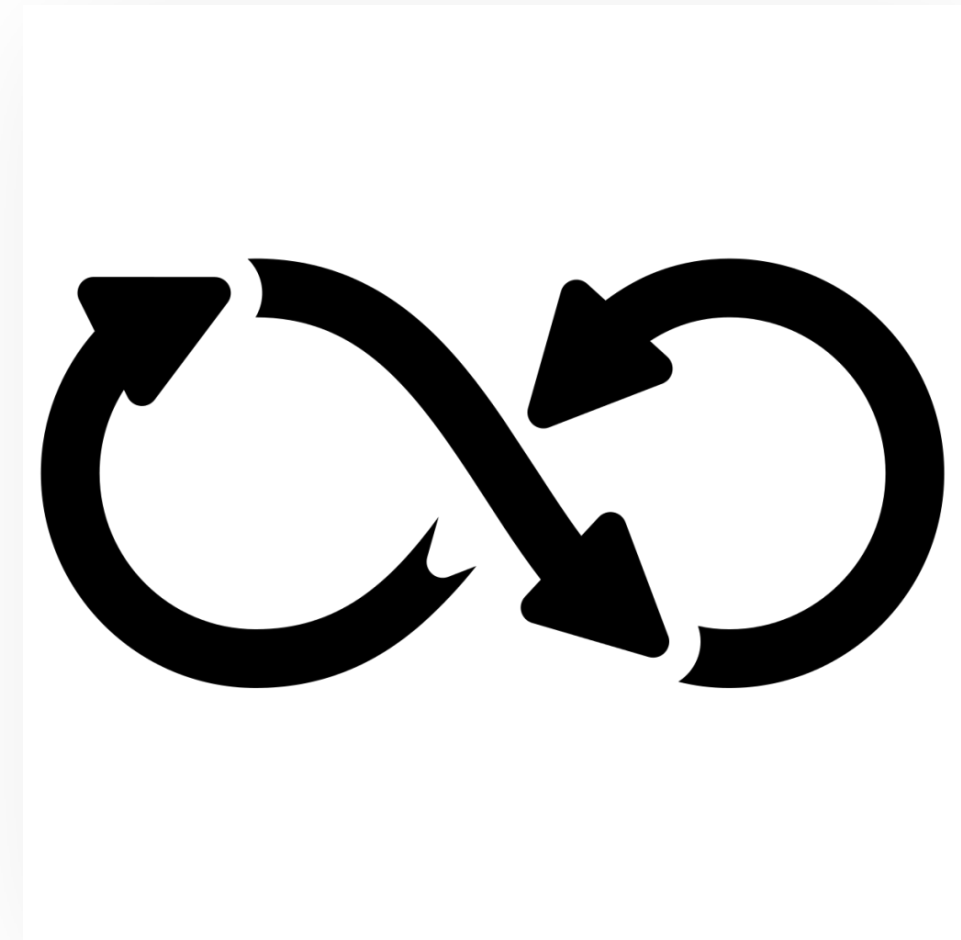
- Layer 3: Advanced Bot Management
 - Real-time **identification**
 - Dynamic **policy enforcement**
 - **Traffic shaping** vs. blocking



Integration Opportunities

Build Bot Management Into Your Stack

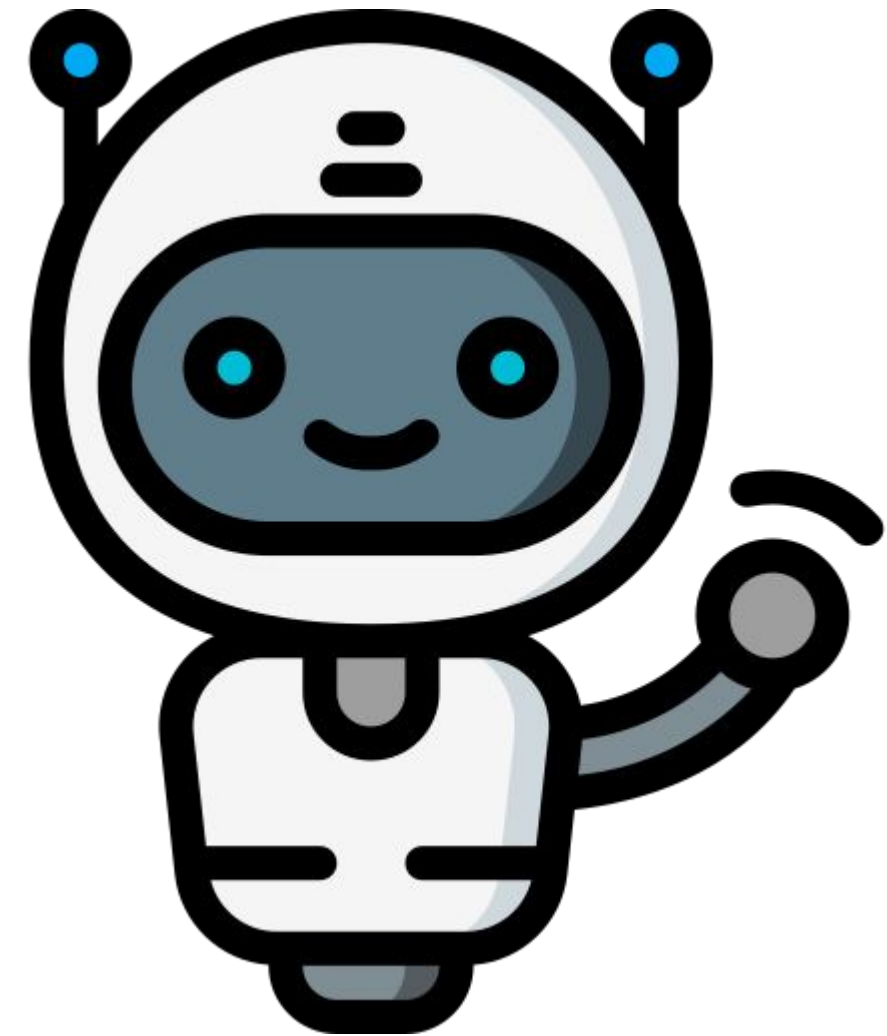
- **CI/CD Pipeline**
 - Test bot policies before deployment
 - *Benefit: Automated policy deployment*
- **Analytics Platform**
 - Separate bot metrics from user metrics
 - *Benefit: Clean data for decision-making*
- **Content Management**
 - Editorial visibility into scraping activity
 - *Benefit: Cross-team visibility and control*



If You're Building AI Services

Be a good Netizen

- **Support Verification**
- **Honor robots.txt**
- **Implement Internal Rate Limits**
- **Consider the Fetcher Problem**



Implementation Priorities

DevOps Action Items

- **Monitor First** – Get visibility before making changes
- **Rate Limiting** – Essential for 39k req/min fetcher traffic
- **Differentiated Handling** – Crawlers ≠ Fetchers
- **Regional Policies** – EMEA vs. North America patterns differ
- **Content Tiering** – Protect premium, allow basic access



Conclusion