Centinel Analytica

Beyond Robots.txt and User Agents

How Technically Enforcing Access Creates Content Scarcity and Monetization Opportunities

Simeon Räthel
Co-Founder
simeon@centinelanalytica.com

Frederick Jahn
Co-Founder
frederick@centinelanalytica.com

Publishers are losing traffic & control of their content...

The Web Is No Longer Human
more than 50% of the webs traffic is bot generated

Up to 50% in Ad Rev loss
Reduced traffic diminishes Ad Revenue

AI Crawlers steal Content
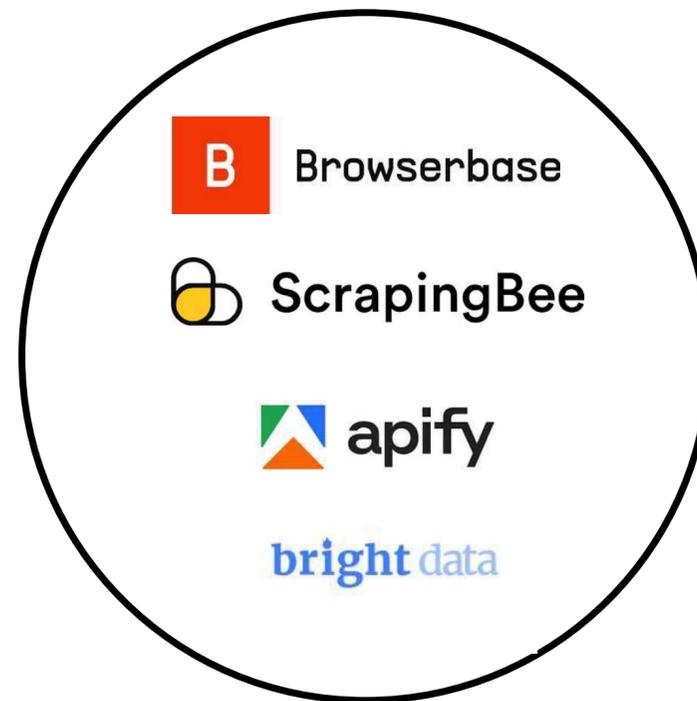Publisher Content feeds LLMs without Consent oder Compensation

AIs don´t give anything back
Average CTRs of 0.25%

AI Access Is Increasingly Intermediated

AI systems require large-scale, continuous web access

Access is increasingly outsourced to specialized data acquisition networks

Crawler ≠ Crawler

:) Declared Crawlers

Identifiable Via User Agent

GPTBot, ClaudeBot, Googlebot

May Respect Robots.Txt

Documented

:(Stealth Crawlers

Headless Browsers

Residential Proxy Networks

IP Rotation

Human-Behavior Mimicry

the seemingly obvious

Solution to stopping AI Crawlers

and how to make money from them

Rule #1

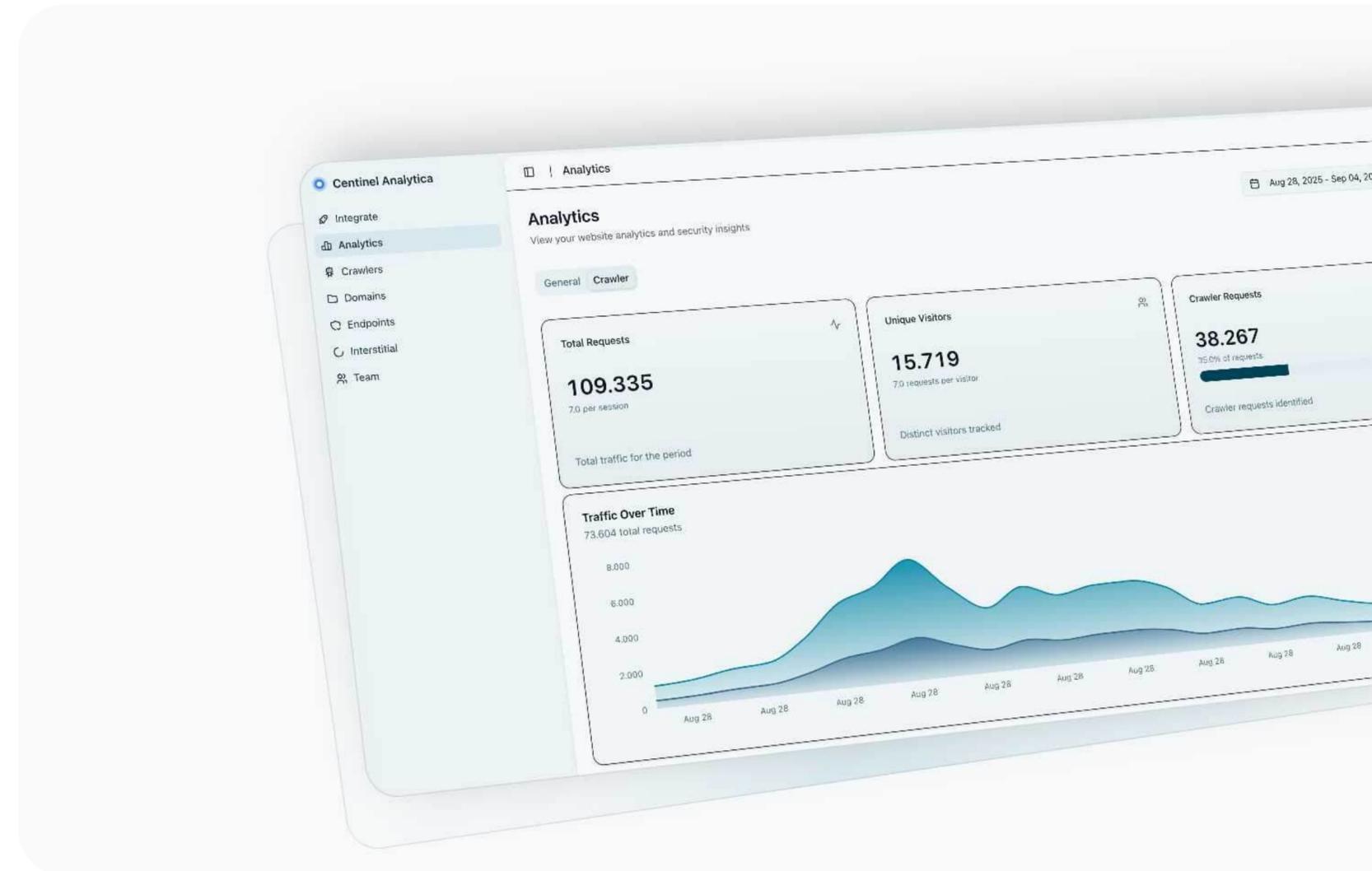# Robots.txt is <u>not</u> the answer!

Robots.txt is a Signal - not a Control Mechanism

Robots.txt is a voluntary guideline that allows publishers to inform crawlers which content may be indexed or crawled.

Why robots.txt is not sufficient

1 **Voluntary, not enforceable** (Crawlers can ignore robots.txt — every third crawler does)

2 **No technical protection** (Content remains freely accessible and scrapeable)

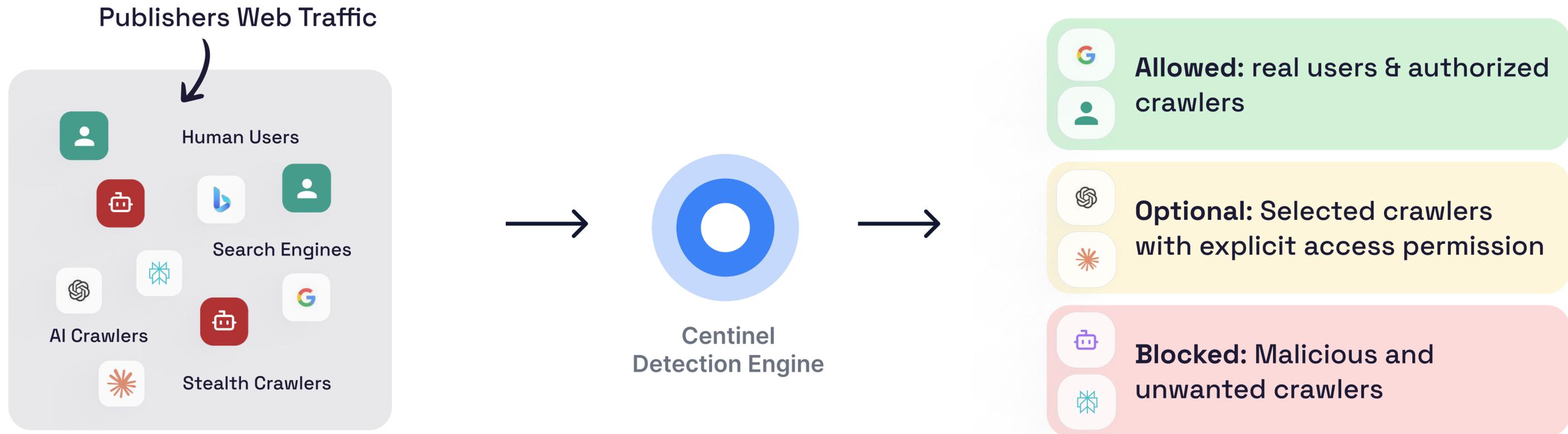**3** **No transparency or control** (Publishers cannot see who complies — and who does not)

Logfile Analysis is not sufficient because it misses stealth crawlers

Rule #2

Get a <u>bot protection solution!</u>

How does a bot protection work?

Publishers Web Traffic

Human Users

Search Engines

AI Crawlers

Stealth Crawlers

Centinel
Detection Engine

Allowed: real users & authorized crawlers

Optional: Selected crawlers with explicit access permission

Blocked: Malicious and unwanted crawlers

1 **Traffic Analysis**

Advanced Fingerprinting and Behavioral Analysis identify unwanted Crawler

2 **Real-time Blocking**

Malicious Bots are stopped immediately at the CDN, CMS, or server level

3 **Tangible Insights**

Dashboards display bot traffic, crawler intent, and crawler origin

Rule #3

Make money from it!

unfortunately, this step will
take some further time for
marketplaces to evolve

Monetization is still in development...

AI content marketplaces are intended to enable publishers to license their content for AI usage in a controlled and compensated way — for example, per crawl, per access, or per usage.

Current opportunities and challenges

1 **Structure and transparency for AI access:** Content usage becomes traceable, measurable, and rule-based.

2 **Shared market logic instead of individual negotiations:** Publishers do not need to negotiate separately with each AI provider.

3 **Lack of enforcement:** Without technical protection, crawlers can continue to access content for free — participation remains voluntary.

Microsoft

TOllBIT

CLOUDFLARE®

Dappier

P paywalls.net

Why Scarcity Enables Marketplaces

Key Fundamentals to consider:

1 the "times of training AI models" are mainly over

2 RAG-bots scrape pages 10× more than training bots

3 RAG-bot scraping rates grew over 50% the past year

RAG systems require continuous retrieval.

When access is **enforceable**,
it can be **redirected**.

From **uncontrolled scraping**
to **structured licensing**.

Through **infrastructure providers**.
Through **marketplaces**.
Through **enterprise integrations**.

Scarcity creates the rails for monetization.

Centinel

Last words

the formula to success...

great content
+ protection layer
+ monetization layer
= AI Money $$$