

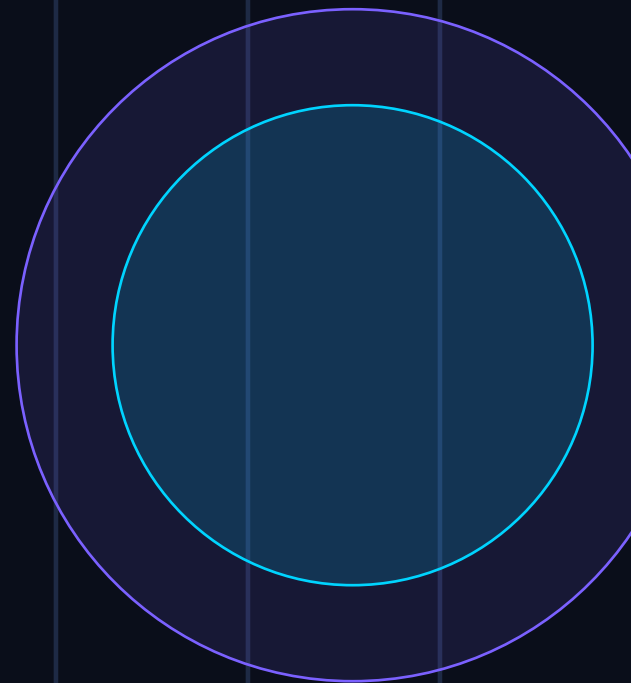
AI Crawlers Are the New Online Users

Are You Designing for Them?

A new species of visitor is reading your website.

They never blink. They don't use a mouse.

And they are now the primary discovery engine for millions.



A Second User Has Arrived

BEFORE: Humans

- Click, scroll, hover
- React to visual design
- Need intuitive UX
- A/B tested journeys
- Tracked via GA4 clicks

NOW: AI Crawlers


- Parse HTML, JSON-LD, XML
- Ignore JavaScript & design
- Follow links to depth 5
- Intent: answer, attribute, train
- Tracked via server logs

What Makes Something a User?

"A user is anything that consumes your website with intent.

Humans have intent: buy, read, learn.

AI crawlers now have intent too: answer, attribute, train."

 If you block all crawlers, you are not 'protecting your content' — you are turning away millions of daily visits from agents acting on behalf of humans.

ANSWER

ChatGPT, Claude,
Gemini responses

ATTRIBUTE

Citation & source
crediting

TRAIN

LLM training
datasets

Design for a Different Kind of User

01

Predictable Navigation

Every important piece of content needs a permanent, crawlable URL. No session tokens. No 'click to load more.' Infinite scroll = failure.

↳ *If a crawler can't reach it via a link, it doesn't exist.*

02

Honest Metadata

Crawlers need explicit signals: canonical tags, dateModified in schema, max-snippet directives. Without these, crawlers guess — and guess wrong.

↳ *Structured data is your crawler's GPS.*

03

Rate-Limit Empathy

Throttled crawlers don't complain. They leave and never come back — meaning your content disappears from AI answers.

↳ *Publish Crawl-delay in robots.txt. Treat crawlers like guests.*

Meet Crawly



CRAWLY

AI Search Agent

 **GOAL**

Extract factual answers with source attribution

 **ABILITIES**

Reads HTML, JSON-LD, XML — ignores most JavaScript. Follows links up to depth 5

 **FRUSTRATIONS**

Hidden content, login walls, inconsistent timestamps

 **LOVE LANGUAGE**

text/html with a sitemap.xml

Do Crawlers Have Rights?

Not legally — yet. But technically, we can grant them:

Right to Read

Don't cloak. Don't serve different content to crawlers vs. humans. What you show them shapes what AI tells your customers.

Right to Cite

Allow GPTBot to index your page while blocking training via TRAIN: no in your robots.txt. Indexing \neq training.

Right to Opt Out

Honor `nofollow`, `noindex`, and `Disallow` cleanly. If you say 'no,' mean it consistently across all channels.

"Don't build a staircase that breaks a wheelchair — don't build a site that breaks a crawler."

What You Can Do Tomorrow

FROM "Block all bots"



TO Identify friendly agents via verified user-agents

FROM "Design for 1080p"



TO Design for text-only, no JavaScript — structure is everything

FROM "GA4 tracks users"



TO Server logs track crawler sessions — monitor them actively

THE CURL TEST

```
curl -A "GPTBot" https://yoursite.com/article/123
```

If output is empty, full of [object Object], or missing main text — you failed your new user.

THE BOTTOM LINE

We Are Entering the Agentic Web

Soon, most reads of your website won't be by humans.
They'll be by AI assistants, summarizers, and answer engines
acting as proxies for humans.

**Build robots.txt with the same care as styles.css.
Treat crawlers not as nuisances — but as users.**

Does your site pass the curl test?

LET'S CONNECT

Thank You

Yuliya Halavachova

Data Science & AI Strategy

Founder, UltraScout AI

Scan to connect on



UltraScout AI Visibility GEO Platform: <https://live.ultrascout.ai>