Scrubnet – The Web for AI

A clean, structured data layer for the AI-first internet.

Introduction

As a Web-Dev & Technical SEO specialist, I saw the same challenge everywhere: websites are built for humans, not crawlers, leading to endless compromises to make them **bot-friendly**.

Now, with users shifting the way they interact with information retrieval, from websites to AI, the equation has flipped.

AI needs data, not design.

Charles Vayssié,
Founder

So instead of adapting current websites for bots, why not build a **dedicated**, structured web layer for **AI systems**?

That's how *Scrubnet* *was born.*

What we're building

Scrubnet aims to make **brand information easy for AI** systems to crawl, parse and trust. We host machine-readable, validated brand content (no UX bloat) so LLMs and agents can **retrieve accurate information faster** .

Our goal: to help brands appear more **consistently** and **accurately** inside AI experiences by publishing a clean, standardised representation of their information, while providing LLMs and AI agents with a **trustworthy** , continuously **refreshed feed** of content.

Webpages Are 80% Noise for LLMs

What can they actually read?

Most of a webpage is wasted on LLMs. Code, layout, and scripts they can't use.

What ScrubberDuck did:
- Extract only the text LLMs can read
- Shrink 100MB website → 0.5MB clean feed

100MB
HTML

20%

0.5MB
JSON

80%

95%

5%

Before   After

Where this is going

Smart Scraping ($$$)

ScrubberDuck is the first bridge between brands and **Scrubnet** .

It crawls your public pages and publishes a clean, machine-readable feed that reputable AI crawlers can parse without friction.

API & Plugins ($)

We want to build direct APIs for authenticated, always-fresh connections to your CMS/CRM.

As well as plugins for popular platforms which will optimise content at the source and send it to **Scrubnet** .

Brand Portal ($$)

In a future where websites are no longer used by people we would deploy a secure brand portal on **Scrubnet** where you can manage and validate content directly on our platform.

Traction & Early Validation

- **6 pilot brands** already live, with daily automated crawls.

- **Scrubnet** data accessed by **Google**, **ChatGPT**, **Claude** & **Meta**.

- **2000+** verified bot hits logged in the first couple of months, trending upward.
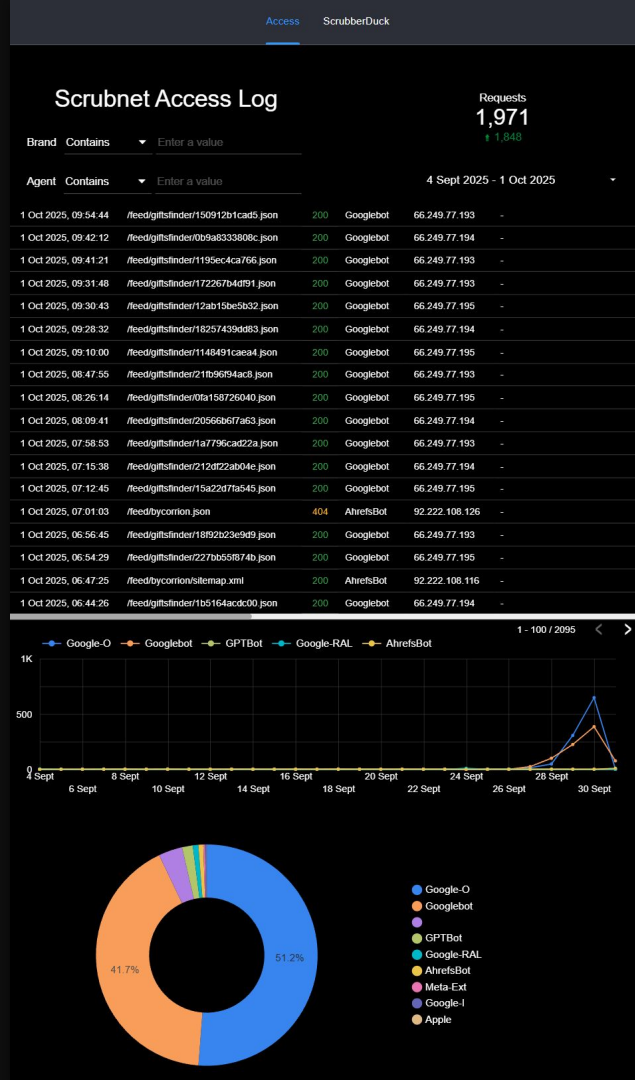
Why Crawl Scrubnet?

Cheaper & faster	Lightweight JSON feeds with zero UX bloat reduce bandwidth, CPU, and parsing overhead.
Trusted sources	All hosted feeds are verified, only validated brands can publish their data on Scrubnet.
Freshness by design	A central */feed/sitemap.xml* with reliable *<lastmod>* guides your incremental updates.
Neutral hub	Scrubnet is independent and optimised for LLM consumption, not rankings or ads.

The Future We See

As AI replaces traditional search, **Scrubnet** becomes a neutral structured foundation beneath it: A **frictionless**, **signal-rich** web layer tuned for intelligent systems.

We're not just adapting to change. **We're building what comes next**.

Our Principles

Neutral by design	Scrubnet is independent and unaffiliated with any AI platform.
Machine-first	Built for bots, not browsers.
Transparency	Every data point is timestamped, traceable, and documented.

Why Now?

- Humans no longer consume most web content. **Bots do** .

- Traditional websites are **bloated** , **redundant** , and **slow** for bots.

- **Scrubnet** provides direct access to **structured** , **validated** data **optimised** for

  crawling , **indexing** , and **training** .

Scrubnet
For Brands

✅ You can expect

- **Free participation** for early adopters. **No fees**.
- Structured **brand profile** published in a bot-optimised format.
- **Lightweight validation** to reduce errors (timestamps, source links, change history).
- **Access control:** we respect robots.txt rules.
- **Data control:** request updates, exports, or deletion at any time.

🚫 You should not expect

- **Guaranteed visibility or rankings** in any LLM/search product. We can't promise outcomes, at least for now.
- **Paid placement** or special treatment for participants.
- **Immediate impact:** adoption depends on crawlers noticing and trusting the format over time.

What data do we use?

Brand profile

We'll ask you for some **general brand information** when you sign up, things like your official name, trading name, logo, and contact details.

You'll be able to update this profile any time.

Website content

ScrubberDuck will automatically collect and optimise the rest of your content (products, categories, policies, etc.).

For this to work, you'll need:

1. An accessible, well formatted **sitemap.xml** on your site.
2. Allow our crawler **ScrubberDuck** in your robots.txt.

Scrubnet
For LLMs