



Scrubnet – The Web for AI



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

Core Concept

Immediate Opportunities

Extended Services

Technical Overview

Moving Forward



Core Concept



The Problem: The Web Isn't Built for Bots

Websites are built for humans, not bots.

- Heavy UX
- Slow, bloated pages
- Hard for crawlers to extract what matters

Result: Poor discovery, wasted crawling, unreliable signals.

Scrubnet : A clean, bot-first layer that removes all the noise.



A Single Gatekeeper Isn't Sustainable

Google dominates how information is accessed. Most AI systems still depend on Google-indexed content.

Industry warning: Cloudflare's CEO said that relying on one source of truth is risky for the future of AI. The one finding a solution will hold the keys of a sustainable future for web search.

The gap: AI needs a neutral, reliable data layer not controlled by any single AI platform.

Scrubnet : A shared, independent hub where brands publish clean structured data for all AI systems.



One Integration, Not Ten

The problem: Every AI platform wants its own product feed. Brands can't maintain dozens of separate integrations.

The reality: More AIs = more feeds = more cost, more breakage, more work.

Scrubnet : One connection for brands. Distributed automatically to every AI and search crawler.

Simple, scalable, future-proof.



Summary

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.



Immediate Opportunities



Immediate Visibility Wins

Before Scrubnet becomes the official AI data layer, we can already test the impact of its current setup.

What we explore now

- Improved visibility by giving bots a cleaner version of brand content
- Better bot readability through structured feeds
- Faster discovery from simplified crawling paths
- Potential uplift in organic performance as bots index Scrubnet feeds

Goal: Prove the early SEO value while we build towards the full AI-brand data hub.



Bot Behaviour: A New Area to Explore

Bots now make up most of the web's traffic. Yet the industry rarely studies how they actually behave.

Opportunity for partner

- Deep analysis of server logs and bot patterns
- Understand how crawlers react to clean structured feeds
- Run controlled experiments on crawl frequency and indexation
- Build evidence-based case studies for clients

Scrubnet : A safe, purpose-built testing ground to study real bot behaviour at scale.



Why Brands Benefit

Faster discovery: Bots find new pages and products sooner.

More frequent crawling: Improved freshness and visibility.

Fewer technical SEO issues: No need for complex fixes on the live site.

UX freedom: Design whatever they want without harming bot access.

One integration: All AIs and search engines served from a single source.



Why LLMs Benefit

Lighter data: Small, clean feeds instead of full web pages.

Lower cost: Less bandwidth, CPU and parsing overhead.

Freshness: Reliable timestamps and structured updates.

Trusted sources: Verified brand data, not scraped noise.



Extended Services



Bot Traffic Monitoring for Clients

Opportunity: Use Scrubnet's logging and crawler-tracking tech to monitor how bots interact with client sites.

What this unlocks

- Understand which bots visit, how often and what they fetch
- Spot crawl waste, gaps and inconsistencies
- Optimise site structure, feeds and signals based on real behaviour
- Give clients clear, data-led visibility reports

Value: A new service built on real bot data, not assumptions.



AI Product Interface Tools

Opportunity: Create AI tools that understand and use a client's product data.

What this enables

- AI chat interfaces powered by the brand's own catalogue
- Product search and recommendations driven by schema and feeds
- On-site assistants that answer questions using trusted data
- Easy setup with Scrubnet as the data backbone

Value: New, practical AI features for clients without heavy development.



Automated Schema Generation

Opportunity: Use Scrubnet's tech to generate clean, accurate structured data for clients.

What this offers

- Automatic product, article and review schema
- Consistent, machine-readable data across the whole site
- Faster onboarding to the structured-data future
- Better visibility for both search and AI systems

Value: A practical service that strengthens client sites and builds partner's expertise in structured data.



Technical Overview



Scrubnet: Behind the Scenes

Daily refresh: ScrubberDuck crawls brand sites once per day using minimal, efficient fetching.

Lightweight extraction: Only essential product and content data are captured.

Three outputs (non-indexable) each designed purely for bots, not users.

- JSON
- TXT
- HTML

Result: A lean, accurate data layer kept fresh with minimal server impact.



An Ecosystem Built Only for Bots

Bot-first design: Every part of Scrubnet is engineered for crawler efficiency.

Optimised components

- HTTP headers tuned for conditional requests
- Precise last-modified and ETag logic
- Purpose-built sitemaps for fast discovery
- Server caching aligned with bot behaviour
- No UX, no tracking, no layout. Zero noise

The result: A domain where every resource exists solely to serve bots clean, structured, reliable data.



Precision Timestamping

Problem: Bots constantly re-fetch unchanged pages because timestamps are unreliable.

Scrubnet's solution, a highly accurate system of:

- Last-modified timestamps
- Consistent ETags
- Feed-level freshness signals

Impact: Bots instantly know what's new and what hasn't changed. No wasted crawling. Faster updates. Maximum efficiency.



Moving Forward



Traction & Early Validation

- 11 pilot brands already live, with daily automated crawls.
- **Scrubnet** data accessed by **Google**, **Meta**, **Bing**, **ChatGPT** ...
- **150k+** verified bot hits logged in just 28 days, trending upward.
- First signs of impact on performance are not negative...

Path contains: /brand/ Bot name: Googlebot Date from: 28/10/2025 Date to: 25/11/2025

Apply filters

Filtered Requests

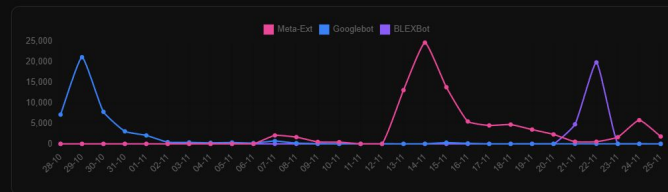
Total 161,736

Current filters apply to the table and chart below.

Time	Status	Bot	Path	IP	Bytes	User-Agent
2025-11-25 13:36:36	200	Meta-Ext	/feed/smartuk/html/86d151b2359c.html	2a03:2880:f804:1f::	3,863	meta
2025-11-25 13:35:34	200	Meta-Ext	/feed/worldwidemenus/html/824003c89d81.html	2a03:2880:f800:3d::	2,402	meta
2025-11-25 13:35:05	404	Meta-Ext	/feed/giftsfinder/txt/a3739fe35776.txt	2a03:2880:f800:1a::	196	meta
2025-11-25 13:32:28	200	Meta-Ext	/feed/smartuk/html/686058761a75.html	2a03:2880:f800:3c::	4,232	meta
2025-11-25 13:32:23	200	Meta-Ext	/feed/smartuk/html/02534c287296.html	2a03:2880:f800:fd::	6,044	meta
2025-11-25 13:31:58	200	Meta-Ext	/feed/giftsfinder/txt/def73759bd34.txt	2a03:2880:f800:1c::	1,909	meta
2025-11-25 13:31:46	200	Meta-Ext	/feed/smartuk/html/fbbf303a874d.html	2a03:2880:f800:29::	3,760	meta
2025-11-25 13:30:58	200	Meta-Ext	/feed/giftsfinder/txt/aa1b8f8f461d.txt	2a03:2880:f800:37::	2,638	meta
2025-11-25 13:30:56	200	Meta-Ext	/feed/smartuk/html/e09c63881961.html	2a03:2880:f800:40::	4,278	meta

Showing 1 to 50 of 161736 Page 1 of 3235 Next

By Crawlers Overtime



By Crawlers, IPs and Status

Crawler	Requests	IP	Requests	Status	Requests
Meta-Ext	86,918	66.249.65.38	35,476	200	138,716
Googlebot	44,240	59.120.220	24,612	301	20,537
BLEXBot	24,612	66.249.65.39	8,503	404	2,477
Google-O	4,549	66.249.65.40	4,345	403	4
GPTBot	585	2a03:2880:f800:20::	1,251	304	2

1 - 5 / 16 1 - 5 / 765 1 - 5 / 5



Partnership: First Steps

What I propose: Add a small group of volunteer clients to **Scrubnet**.

What partner brings: Expert analysts to monitor crawl activity and organic impact.

What we learn together

- How bots interact with web servers and resources
- Early visibility gains if any
- Which services could evolve into full client offerings

Goal: A low-risk pilot that builds evidence and shapes the next phase of Scrubnet while improving partner's internal Tech SEO skill set and improving client offering.



THANK YOU



Annex



Introduction



Charles Vayssié,
Founder

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.

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.*



Webpages Are 80% Noise for LLMs

Readable by LLMs 

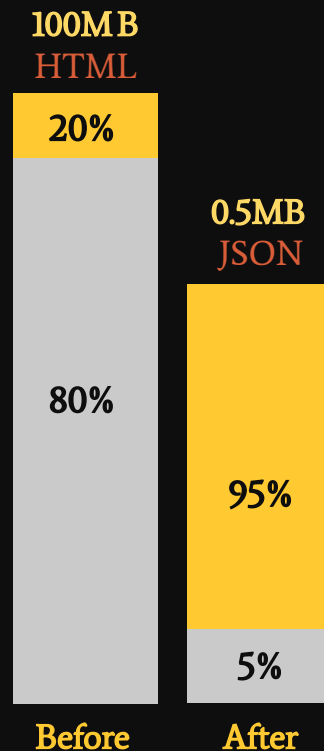
Ignored by LLMs 

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



A Neutral Publisher Layer for All AI Platforms

Scrubnet acts as a universal data hub between brands and AI platforms.

Brands publish their product feeds once, and Scrubnet distributes clean, structured data to every AI crawler.

LLMs get consistent, machine-first metadata.

Brands get one simple integration.

Scrubnet becomes the shared layer that powers product discovery across the entire AI ecosystem.



ScrubberDuck Subscription

1 Website

£5 /month

Up to 50,000 URLs

- ScrubberDuck activation
- Machine-readable JSON feed hosted on Scrubnet
- Change tracking with timestamps
- Access logs and basic monitoring
- Cancel anytime; no lock-in



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.



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.

