FAST GROWTH Advisors

WHITE PAPER

AUGUST 2026

COUNTING BOTS, NOT LIES

Measuring verified AI traffic

> *Thirteen days of logs, two sites, every bot verified. A measurement, not a declaration.*

Authenticated

Identity confirmed by a published range, a signature or reverse resolution.

Proven impersonation

Verification was possible and failed: the address is not the vendor's.

Unverifiable

The vendor publishes nothing that allows the test. No verdict, no accusation.

Undecided

Verification should have succeeded. That is a limit of our instrument, declared here.

This document sells nothing. It shows the gap between what bots declare and what verification confirms.

SECTION 01

IN BRIEF

The AI visibility measurement market rests on a convention nobody states out loud: a bot is whatever it declares itself to be. Every automated visitor announces its identity in a line of text it writes itself, the user agent, and almost every published figure on AI crawling, scrape ratios, error rates, crawler market shares, is a sum of those declarations.

For thirteen days we applied to our own server logs the identity verification that engine publishers make possible: published address ranges, reverse resolution confirmed in both directions.

The central result fits in one line. Of 1,129 requests presenting themselves as an AI assistant fetching a page, 876 were proven impersonations, and they are what produces most of the error rate a classic tool would report. Authenticated requests, by contrast, almost never knock on the wrong door: zero missing pages across 154 requests.

This document describes the method, gives every figure with its base, states the limits of our own instrument, and ends with the one question that lets anyone assess an AI visibility tool in a single sentence.

SECTION 02

IS A DECLARATION AN IDENTITY?

The user agent dates back to the early web. It is a free text field: nothing stops any machine from writing `Googlebot` or `ChatGPT-User` in it, and many do, because the disguise opens doors. A firewall lets through what looks like a search engine. A site accepts being scraped by what it believes to be an index.

The consequence has been known to system administrators for twenty years, and is ignored by almost every marketing dashboard: counting user agents means counting declarations. The resulting figure mixes genuine bots, impersonators, and anything disguising itself to get through.

There is a way to settle it, bot by bot, request by request. Three mechanisms, from strongest to weakest. Cryptographic signature, defined by Web Bot Auth and RFC (Request For Comments) 9421, still barely deployed. The address range published by the vendor, which OpenAI makes available as JSON (JavaScript Object Notation) at stable addresses. And reverse resolution confirmed in both directions: you ask for the name attached to the address, then check that the name points back to that address. One way is not enough, an impersonator can configure the return. The round trip cannot be faked.

Two implementation traps, learned by doing. The expected domain suffix must start with a dot, otherwise `evil-googlebot.com` satisfies a `googlebot.com` test. And real suffixes are measured, not assumed: Amazon's bot resolves to `.crawl.amazonbot.amazon`, whose top level domain is `.amazon` and not `.amazon.com`.

Whoever assumes lets them through.

SECTION 03

WHY FOUR VERDICTS RATHER THAN TWO?

The binary reflex, verified or not, is not enough. It mixes two morally opposite situations: the bot that fails a possible verification, and the bot that no mechanism allows you to verify at all.

Our instrument therefore assigns four verdicts to every request. **Authenticated**: identity confirmed by one of the three mechanisms. **Proven impersonation**: verification was possible and it failed, the address does not belong to the declared vendor. **Unverifiable**: the vendor publishes neither ranges nor reverse records, no test is possible. **Undecided**: verification should have succeeded and did not, which flags a limit of the instrument rather than a fault of the bot.

The distinction is not cosmetic. ClaudeBot, Anthropic's crawler, is unverifiable: 841 requests on our main site during the window, none verified and no proven impersonation either. It is not an impersonator, it is a bot without verifiable papers. The same holds for meta-externalagent, youbot and claude-user, and for the DuckDuckGo bot, at zero verification on both sites.

The distinction shows up in the figures, and it separates names one might have thought neighbours. bytespider and google-extended are not unverifiable: their verification does resolve, and it fails. 133 requests out of 133 for the first, 292 out of 292 for the second, all proven false. Carrying a known vendor's name therefore says nothing about what sits behind it.

The question this list raises is not who cheats. It is why some vendors publish what is needed to verify their bots while others publish nothing at all, when the cost of publishing is zero and the benefit, for them as for the sites they visit, is immediate.

SECTION 04

WHAT DOES VERIFICATION SHOW OVER THIRTEEN DAYS?

Our measurements cover the window from 11 to 23 August 2026, thirteen full days, on the server logs of two B2B sites, fast-growth.fr and nomo-ia.com. The window was chosen clean: after our classification fixes of 8 August and the reclassification of history, before an infrastructure change on 24 August, with no deployment in between. Detail and aggregates agree across the whole window, with no discrepancy.

First table, declared against verified, by bot class.

DECLARED CLASS	FAST-GROWTH.FR	VERIFIED	NOMO-IA.COM	VERIFIED
Classic search engine	492	78.9%	567	96.5%
AI index bot	1,419	51.1%	882	46.4%
Training collection	3,907	9.5%	3,198	13.8%
User-triggered fetch	1,129	13.6%	688	3.5%

Verification rate by bot class, 11 to 23 August 2026, two B2B sites.

The overall reading is clear: classic engines are verifiable at 79% and 97%, AI agents at 3% and 14%. Same method, same window, same server.

But that global rate blends the four verdicts, and the breakdown says something other than "AI bots are opaque".

VERDICT	FAST-GROWTH.FR (6,455 REQ.)	SHARE	NOMO-IA.COM (4,768 REQ.)	SHARE
Authenticated	1,250	19.4%	875	18.4%
Proven impersonation	3,372	52.2%	3,646	76.5%
Unverifiable, silent vendor	1,041	16.1%	239	5.0%
Undecided, instrument limit	792	12.3%	8	0.2%

Breakdown into four verdicts, across all AI agent requests in the window.

The figure that matters sits in the middle. Most declared AI crawling is not unverifiable, it is proven false: 52% on one site, 76% on the other. The silent vendor nuance, the one honesty requires, covers only 16% and 5% of the volume.

DECLARED BOT	VENDOR	REQUESTS	AUTHENTICATED	IMPERSONATIONS	IMPERSONATED SHARE	DOMINANT VERDICT
amazonbot	Amazon	1,388	223	1,165	84 %	mixed
chatgpt-user	OpenAI	1,030	154	876	85 %	mixed
claudebot	Anthropic	841	0	0	0 %	unverifiable
meta-externalagent	Meta	792	0	0	0 %	unverifiable
gptbot	OpenAI	452	148	304	67 %	mixed
oai-searchbot	OpenAI	441	155	286	65 %	mixed
petalbot	Huawei	419	419	0	0 %	authenticated
perplexity-bot	Perplexity	360	52	308	86 %	mixed
google-extended	Google	292	0	292	100 %	impersonated
googlebot	Google	231	149	23	10 %	mixed
bingbot	Microsoft	213	206	7	3 %	mixed
bytespider	ByteDance	133	0	133	100 %	impersonated
applebot	Apple	107	99	8	7 %	mixed
claude-user	Anthropic	99	0	0	0 %	unverifiable
youbot	You.com	92	0	0	0 %	unverifiable

Volume and identity verdict per declared bot, 11 to 23 August 2026, fast-growth.fr. Bots with 50 requests or more. The "neither" case in the text covers requests that are neither authenticated nor proven false, for want of a verification mechanism.

The table reads in three families. Bots whose identity can be verified and holds: petalbot at 100% authentication, bingbot, applebot, googlebot. Those no mechanism allows you to verify, therefore never accused: claudebot, meta-externalagent, claude-user, youbot. And those whose name is massively used as a disguise.

Two results deserve flagging. The most impersonated name in our window is not an answer engine: it is amazonbot, with 1,165 proven impersonations across 1,388 requests. And google-extended, the agent Google associates with its training uses, shows no authenticated request at all across 292: every request carrying that name, on our server and in this window, is proven false.

The contrast with googlebot, from the same vendor, is instructive: 149 authenticated requests out of 231, and 23 impersonations. The bot Google lets you verify is verified; the one whose verification is not documented the same way serves as a mask.

SECTION 05

WHO DOES THE AI CRAWLER ERROR RATE BELONG TO?

Let us isolate the most commented category on the market: the user fetch, a request from an assistant retrieving a page on a human's behalf. On fast-growth.fr, same window, same base.

VERDICT	REQUESTS	PAGES FOUND	MISSING PAGES
Authenticated	154	99.4%	0.0%
Not verified	99	87.9%	1.0%
Proven impersonation	876	30.7%	69.1%
All verdicts combined	1,129	n.a.	53.7%

Behaviour of AI fetches by identity verdict, fast-growth.fr, 1,129 requests.

Authenticated requests find their page. Always, or nearly: not one missing page across 154 requests. Impersonations, by contrast, grope around. 69.1% of their requests target pages that do not exist, the typical behaviour of vulnerability scanning, not of content reading.

And the total, the only reading a non-verifying tool can produce, shows a 53.7% error rate. A figure manufactured entirely by impersonators.

The external comparison follows. Vercel, in the December 2024 reference study, publishes 34.82% missing pages for ChatGPT and 34.16% for Claude, against 8.22% for Googlebot. Those rates, quoted everywhere as proof that AI crawlers waste resources, sit in the range of our unfiltered total, never in the range of our authenticated traffic.

So we offer a hypothesis, given for what it is: the error rate attributed to AI crawlers may measure, to a substantial extent, the behaviour of those impersonating them. We do not know the detail of Vercel's method and claim nothing about their figures. We show that on our servers, with a public and replayable method, the gap between the declarative reading and the verified reading reaches 53.7 points.

SECTION 06

WHY DO TWO SITES GIVE DIFFERENT RATES?

An attentive reader of the first table will have noticed that classic engines are verified at 78.9% on one site and 96.5% on the other. Eighteen points apart, same method, same resolver. An anomaly?

We resolved the addresses one by one. 67 distinct addresses declared themselves Googlebot on fast-growth.fr: 7 with a genuine Google resolution, 3 pointing to third parties, 57 with no resolution at all. On nomo-ia.com, 20 declared addresses, including the same 7 genuine ones.

The real Googlebots are identical on both sides. It is the fake ones that make the difference. Among the impersonators, machines hosted with providers where anyone can rent a machine by the hour.

The gap is therefore not a flaw in the method. It measures each site’s exposure to impersonators, a quantity no declarative tool can produce, since it does not know they are impersonators. A site that is more visible, older or more attacked carries more fake Googlebots. The control figure does control something.

SECTION 07

DO ACQUISITION CHANNELS LIE TOO?

Identity verification corrects the counting of bots. It does not correct everything, and two blind spots deserve documenting, because they affect every tool on the market, ours included.

The referral channel first. Over the same window, 12,087 requests in our logs carried a referring site, the classic definition of traffic coming from elsewhere. 10,740 of them, 88.9%, were internal navigation: a visitor already on the site clicking through to another page of the same site, which the classification chain filed under acquisition for want of a rule. The fourth channel of our acquisition screen was wrong nearly 90% of the time. We fixed the rule. How many acquisition screens elsewhere still count their own navigation as referred visits?

AI referrals next. Over the window, four requests referred by an AI interface. That figure is a floor, not a measurement: the native ChatGPT and Claude applications send no provenance header, so an unknowable share of that traffic arrives as direct. Any "zero AI referrals", ours included, reads as "zero referrals carrying a header", never as "zero visitors sent".

On that note, the scrape ratios in circulation.

SOURCE	RATIO	WINDOW	SCOPE
Cloudflare	70,900 : 1	June 2025	Anthropic bots
Cloudflare	≈ 4,580 : 1	28 days, late June 2026	Cloudflare network
Microsoft Clarity	≈ 6,000 : 1	August 2026	published example

Three ratios between pages scraped and visitors referred. All real, none comparable to the others.

Windows, scopes and methods differ. Quoting a ratio without its window is quoting a headline, not a measurement. And since these ratios are computed on undercounted referrals, they mechanically overstate, as their own authors point out.

SECTION 08

WHO ACTUALLY GETS THROUGH, ONCE PAPERS ARE CHECKED?

Once traffic is filtered on confirmed identities, the real landscape is surprising. These orders of magnitude were measured over seventeen days, two sites, before the main window, and have been stable since.

On fetches triggered by a real prompt, OpenAI accounts for all verified traffic. On AI indexing, the leading bot is neither OpenAI nor Anthropic: it is PetalBot, Huawei's crawler, at 55% and 68% depending on the site. On training collection, amazonbot runs about forty times the volume of GPTBot.

And the most important one is invisible by construction. Gemini never visits your server: it reads Google's index, built by Googlebot for classic search, which means that a fetch dashboard, however well verified and however complete its identity checks, stays structurally blind to one of the market's main answer engines.

Your tools show you who comes by. They cannot show you who reads you without coming by.

SECTION 09

WHAT ARE THE LIMITS OF OUR OWN MEASUREMENT?

A document that faults the market for its declarative habits owes it to itself to be beyond reproach about its own gaps. Here they are.

The 792 undecided requests on fast-growth.fr, 12.3% of AI agent volume, concern a single bot, meta-externalagent, whose verification succeeds on one site and not the other. It is neither an impersonation nor an impossibility: it is a gap in our instrument, unresolved to date, counted separately and declared here rather than quietly fixed.

The volumes are those of two small company sites: 154 authenticated fetches in thirteen days on the main site, 24 on the second. Too few to carry a figure on their own. We are not describing the web, we are describing a method and its effect on the numbers.

The window is short, thirteen days, and will be extended. Finally, the classification of the categories themselves, index, training, fetch, relies on vendors' declarations about the role of their bots, declarations we cannot verify.

SECTION 10

WHAT QUESTION SHOULD YOU ASK YOUR MEASUREMENT TOOL?

For an executive, this whole document folds into one question, to be put to a tool or a provider: what does bot identification rest on in this report?

If the answer is the user agent, the volumes shown mix bots and impersonators in proportions this document lets you imagine.

The standard, meanwhile, is moving. The cryptographic signatures of Web Bot Auth, described by RFC 9421, would give strong verification without reverse resolution, and OpenAI already publishes its address ranges at stable locations. The day signing becomes the norm, the question will change sides: it will be put to the bots that do not sign, and to the vendors that publish nothing.

Until then, verification in both directions remains the best tool available, and it is within reach of anyone who reads their server logs. Counting correctly is possible today, and it is the precondition for any AI visibility strategy: before optimising for bots, make sure they are bots.

SECTION 11

FREQUENTLY ASKED QUESTIONS

How do I verify the identity of a bot visiting my site?

Three mechanisms exist, from strongest to weakest: the cryptographic signature defined by RFC 9421, the address range published by the vendor, and reverse resolution confirmed in both directions. The last one means asking for the name attached to the address, then checking that the name points back to that address. One way is not enough: an impersonator can configure the return.

Why is the AI crawler error rate so high in published studies?

Because it is computed on declarations. In our logs, authenticated requests show zero missing pages across 154 requests, while proven impersonations show 69.1%. The unfiltered total, the only reading available without identity verification, comes out at 53.7%. A tool that does not verify therefore attributes to engines the behaviour of those imitating them.

Is ClaudeBot suspicious since it is never verified?

No. It is unverifiable, which differs from an impersonation. Anthropic publishes neither address ranges nor reverse records allowing the test, so no verdict can be reached. Confusing the two would mean accusing a bot for want of being able to identify it.

Can the scrape ratios published by different players be compared?

Not unless you compare their windows and scopes. The three ratios in circulation, 70,900 to 1, about 4,580 to 1 and about 6,000 to 1, are all real but cover different periods, networks and methods. A ratio quoted without its window is a headline, not a measurement.

Is my AI visibility tool reliable?

The question to put to it fits in one sentence: what does bot identification rest on in this report? If the answer is the user agent, the volumes mix genuine bots and impersonators. If the answer mentions published address ranges or reverse resolution in both directions, the tool knows what it is talking about.

SECTION 12

GLOSSARY

User agent

A line of text an automated visitor writes itself to announce its identity. A free field, therefore forgeable: counting it means counting declarations.

Reverse resolution in both directions

A check that asks for the name attached to an address, then confirms that the name points back to that same address. Only the round trip proves anything, since one way can be configured by an impersonator.

Proven impersonation

The verdict when verification was possible and failed: the address does not belong to the declared vendor. To be distinguished from unverifiable.

Unverifiable

The verdict when the vendor publishes neither address ranges nor reverse records. No test is possible, no fault is established.

User fetch

A request from an assistant retrieving a page on a human's behalf, as opposed to indexing and training collection.

Web Bot Auth

A mechanism for cryptographically signing bot requests, built on RFC 9421. It would provide strong verification without relying on reverse resolution.

Scrape ratio

The ratio between pages scraped by a vendor's bots and visitors it refers back. Sensitive to window and scope, therefore not comparable across sources.

SECTION 13

SOURCES

1. Vercel and MERJ (web analytics firm). *The rise of the AI crawler*, 17 December 2024. Missing page rates of 34.82% for ChatGPT, 34.16% for Claude and 8.22% for Googlebot. vercel.com
2. OpenAI, published address ranges for its bots, in JSON, at stable addresses. gptbot.json, searchbot.json, chatgpt-user.json
3. RFC 9421, *HTTP Message Signatures*, IETF (Internet Engineering Task Force), February 2024. Technical basis for Web Bot Auth. rfc-editor.org
4. Cloudflare. *From Googlebot to GPTBot: who's crawling your site in 2025*. Ratios between pages scraped and visitors referred. blog.cloudflare.com
5. Microsoft Clarity. *Scrape-to-Referral insights*, 13 August 2026. A ratio of about 6,000 to 1 in the published example. clarity.microsoft.com
6. Fast Growth Advisors measurements, server logs of fast-growth.fr and nomo-ia.com, window from 11 to 23 August 2026. Bases stated with every figure in this document.