

COMPTER DES ROBOTS, PAS DES MENSONGES



Mesurer le trafic IA vérifié

Treize jours de journaux, deux sites, chaque robot vérifié. Une mesure, pas une déclaration.

Authentifié

L'identité est confirmée par une plage publiée, une signature ou une résolution inverse.

Usurpation prouvée

La vérification était possible et elle a échoué : l'adresse n'est pas celle de l'éditeur.

Invérifiable

L'éditeur ne publie rien qui permette le test. Aucun verdict, aucune accusation.

Non tranché

La vérification aurait dû aboutir. C'est une limite de notre instrument, déclarée ici.

Ce document ne vend rien. Il montre l'écart entre ce que les robots déclarent et ce que la vérification confirme.

SECTION 01

EN RÉSUMÉ

Le marché de la mesure de visibilité IA repose sur une convention que personne n'énonce : un robot est ce qu'il déclare être. Chaque visiteur automatique annonce son identité dans une ligne de texte qu'il écrit lui-même, le user-agent, et la quasi-totalité des chiffres publiés sur le crawl IA, ratios d'aspiration, taux d'erreur, parts de marché des robots, sont des additions de ces déclarations.

Nous avons appliqué à nos propres journaux serveur, pendant treize jours, la vérification d'identité que les éditeurs de moteurs rendent possible : plages d'adresses publiées, résolution inverse confirmée dans les deux sens.

Le résultat central tient en une ligne. Sur 1 129 requêtes se présentant comme un assistant IA venu chercher une page, 876 étaient des usurpations prouvées, et ce sont elles qui produisent l'essentiel du taux d'erreur que mesurerait un outil classique. Les requêtes authentifiées, elles, ne se trompent presque jamais de porte : zéro page inexistante sur 154 requêtes.

Ce document décrit la méthode, donne tous les chiffres avec leurs bases, déclare les limites de notre propre instrument, et se termine par la seule question qui permette d'évaluer un outil de visibilité IA en une phrase.

SECTION 02

UNE DÉCLARATION VAUT-ELLE UNE IDENTITÉ ?

Le user-agent date des débuts du web. C'est un champ libre : rien n'empêche une machine quelconque d'y écrire `Googlebot` ou `ChatGPT-User`, et beaucoup le font, parce que le déguisement ouvre des portes. Un pare-feu laisse passer ce qui ressemble à un moteur de recherche. Un site accepte d'être aspiré par ce qu'il croit être un index.

La conséquence est connue des administrateurs système depuis vingt ans, et ignorée de presque tous les tableaux de bord marketing : compter des user-agents, c'est compter des déclarations. Le chiffre obtenu mélange les robots authentiques, les imposteurs, et tout ce qui se déguise pour passer.

Il existe pourtant un moyen de trancher, robot par robot, requête par requête. Trois mécanismes, du plus fort au plus faible. La signature cryptographique, définie par Web Bot Auth et la RFC (Request For Comments) 9421, encore peu déployée. La plage d'adresses publiée par l'éditeur, qu'OpenAI met à disposition en JSON (JavaScript Object Notation) à des adresses stables. Et la résolution inverse confirmée dans les deux sens : on demande le nom associé à l'adresse, puis on vérifie que ce nom pointe bien vers cette adresse. L'aller simple ne suffit pas, un imposteur peut configurer le retour. L'aller-retour, non.

Deux pièges de mise en œuvre, appris en le faisant. Le suffixe de domaine attendu doit commencer par un point, sans quoi `evil-googlebot.com` satisfait le test `googlebot.com`. Et les suffixes réels se mesurent, ils ne se supposent pas : le robot d'Amazon résout en `.crawl.amazonbot.amazon`, dont le domaine de premier niveau est `.amazon` et non `.amazon.com`.

Qui suppose laisse passer.

SECTION 03

POURQUOI QUATRE VERDICTS, ET PAS DEUX ?

Le réflexe binaire, vérifié ou pas, ne suffit pas. Il mélange deux situations moralement opposées : le robot qui échoue à une vérification possible, et le robot qu'aucun mécanisme ne permet de vérifier.

Notre instrument attribue donc quatre verdicts à chaque requête. **Authentifié** : l'identité est confirmée par l'un des trois mécanismes. **Usurpation prouvée** : la vérification était possible et elle a échoué, l'adresse n'appartient pas à l'éditeur déclaré. **Invérifiable** : l'éditeur ne publie ni plage ni enregistrement inverse, aucun test n'est possible. **Non tranché** : la vérification aurait dû aboutir et n'a pas abouti, ce qui signale une limite de l'instrument, pas une faute du robot.

La distinction n'est pas cosmétique. ClaudeBot, le robot d'Anthropic, est invérifiable : 841 requêtes sur notre site principal pendant la fenêtre, zéro vérifiée et zéro usurpation prouvée. Ce n'est pas un imposteur, c'est un robot sans papiers vérifiables. Même chose pour meta-externalagent, youbot et claude-user, et pour le robot de DuckDuckGo, à zéro vérification des deux côtés.

La distinction se lit dans les chiffres, et elle sépare des noms qu'on aurait pu croire voisins. bytespider et google-extended, eux, ne sont pas invérifiables : leur vérification aboutit, et elle échoue. 133 requêtes sur 133 pour le premier, 292 sur 292 pour le second, toutes prouvées fausses. Porter un nom d'éditeur connu ne dit donc rien de ce qu'il y a derrière.

La question que cette liste pose n'est pas de savoir qui triche. Elle est de savoir pourquoi certains éditeurs publient de quoi vérifier leurs robots quand d'autres n'en publient rien, alors que le coût de publication est nul et que le bénéfice, pour eux comme pour les sites visités, est immédiat.

SECTION 04

QUE DONNE LA VÉRIFICATION SUR TREIZE JOURS ?

Nos mesures portent sur la fenêtre du 11 au 23 août 2026, treize jours pleins, sur les journaux serveur de deux sites B2B, fast-growth.fr et nomo-ia.com. Fenêtre choisie propre : après nos correctifs de classification du 8 août et le reclassement de l'historique, avant un changement d'infrastructure du 24, sans aucun déploiement entre les deux. Détail et agrégats concordent sur toute la fenêtre, sans écart.

Premier tableau, le déclaré contre le vérifié, par classe de robot.

CLASSE DÉCLARÉE	FAST-GROWTH.FR	VÉRIFIÉES	NOMO-IA.COM	VÉRIFIÉES
Moteur de recherche classique	492	78,9 %	567	96,5 %
Robot d'index IA	1 419	51,1 %	882	46,4 %
Collecte d'entraînement	3 907	9,5 %	3 198	13,8 %
Fetch déclenché par un utilisateur	1 129	13,6 %	688	3,5 %

Taux de vérification par classe de robot, du 11 au 23 août 2026, deux sites B2B.

La lecture d'ensemble est nette : les moteurs classiques sont vérifiables à 79 et 97 %, les agents IA à 3 et 14 %. Même méthode, même fenêtre, même serveur.

Mais ce taux global mélange les quatre verdicts, et la décomposition dit autre chose que « les IA sont opaques ».

VERDICT	FAST-GROWTH.FR (6 455 REQ.)	PART	NOMO-IA.COM (4 768 REQ.)	PART
Authentifié	1 250	19,4 %	875	18,4 %
Usurpation prouvée	3 372	52,2 %	3 646	76,5 %
Invérifiable, éditeur muet	1 041	16,1 %	239	5,0 %
Non tranché, limite de l'instrument	792	12,3 %	8	0,2 %

Décomposition en quatre verdicts, sur l'ensemble des requêtes d'agents IA de la fenêtre.

Le chiffre qui compte est au milieu. La majorité du crawl IA déclaré n'est pas invérifiable, elle est prouvée fausse : 52 % sur un site, 76 % sur l'autre. La nuance des éditeurs muets, celle qu'il faut faire par honnêteté, ne couvre que 16 % et 5 % du volume.

ROBOT DÉCLARÉ	ÉDITEUR	RE- QUÊTES	AUTHENTI- FIÉES	USURPA- TIONS	PART USURPÉE	VERDICT DOMINANT
amazonbot	Amazon	1388	223	1165	84 %	mixte
chatgpt-user	OpenAI	1030	154	876	85 %	mixte
claudebot	Anthropic	841	0	0	0 %	invérifiable
meta- externalagent	Meta	792	0	0	0 %	invérifiable
gptbot	OpenAI	452	148	304	67 %	mixte
oai-searchbot	OpenAI	441	155	286	65 %	mixte
petalbot	Huawei	419	419	0	0 %	authentifié
perplexitybot	Perplexity	360	52	308	86 %	mixte
google-extended	Google	292	0	292	100 %	usurpé
googlebot	Google	231	149	23	10 %	mixte
bingbot	Microsoft	213	206	7	3 %	mixte
bytespider	Byte- Dance	133	0	133	100 %	usurpé
applebot	Apple	107	99	8	7 %	mixte
claude-user	Anthropic	99	0	0	0 %	invérifiable
youbot	You.com	92	0	0	0 %	invérifiable

Volume et verdict d'identité par robot déclaré, du 11 au 23 août 2026, fast-growth.fr. Robots de 50 requêtes ou plus. La colonne « ni l'un ni l'autre » du texte correspond aux requêtes ni authentifiées ni prouvées fausses, faute de mécanisme de vérification.

Ce tableau se lit en trois familles. Les robots dont l'identité se vérifie et tient : petalbot à 100 % d'authentification, bingbot, applebot, googlebot. Ceux qu'aucun mécanisme ne permet de vérifier, donc jamais accusés : claudebot, meta-externalagent, claude-user, youbot. Et ceux dont le nom sert massivement de déguisement.

Deux résultats méritent d'être signalés. Le nom le plus usurpé de notre fenêtre n'est pas celui d'un moteur de réponse : c'est amazonbot, avec 1 165 usurpations prouvées sur 1 388 requêtes. Et google-extended, l'agent que Google associe à ses usages d'entraînement, ne présente aucune requête authentifiée sur 292 : la totalité du trafic portant ce nom, sur notre serveur et sur cette fenêtre, est prouvée fausse.

Le contraste avec googlebot, du même éditeur, est instructif : 149 requêtes authentifiées sur 231, et 23 usurpations. Le robot que Google permet de vérifier l'est ; celui dont la vérification n'est pas documentée de la même façon sert de masque.

SECTION 05

À QUI APPARTIENT LE TAUX D'ERREUR DES CRAWLERS IA ?

Isolons la catégorie la plus commentée du marché : le fetch utilisateur, la requête d'un assistant qui va chercher une page à la demande d'un humain. Sur fast-growth.fr, même fenêtre, même base.

VERDICT	REQUÊTES	PAGES TROUVÉES	PAGES INEXISTANTES
Authentifié	154	99,4 %	0,0 %
Non vérifié	99	87,9 %	1,0 %
Usurpation prouvée	876	30,7 %	69,1 %
Tous verdicts confondus	1 129	n.s.	53,7 %

Comportement des fetchs IA selon le verdict d'identité, fast-growth.fr, 1 129 requêtes.

Les requêtes authentifiées trouvent leur page. Toujours, ou presque : aucune page inexistante sur 154 requêtes. Les usurpations, elles, tâtonnent. 69,1 % de leurs requêtes visent des pages qui n'existent pas, comportement typique du balayage de failles, pas de la lecture de contenu.

Et le total, seule lecture qu'un outil non vérifiant puisse produire, affiche 53,7 % d'erreurs. Un chiffre entièrement fabriqué par les imposteurs.

La comparaison externe s'impose. Vercel, dans l'étude de référence de décembre 2024, publie 34,82 % de pages inexistantes pour ChatGPT et 34,16 % pour Claude, contre 8,22 % pour Googlebot. Ces taux, cités partout comme la preuve que les crawlers IA gaspillent, sont de l'ordre de notre total non filtré, jamais de l'ordre de notre trafic authentifié.

Nous formulons donc une hypothèse, en la donnant pour ce qu'elle est : le taux d'erreur attribué aux crawlers IA mesure peut-être, pour une part substantielle, le comportement de ceux qui se font passer pour eux. Nous ne connaissons pas le détail de la méthode de Vercel et n'affirmons rien sur leurs chiffres. Nous montrons que sur nos serveurs, avec une méthode publique et rejouable, l'écart entre la lecture déclarative et la lecture vérifiée atteint 53,7 points.

SECTION 06

POURQUOI DEUX SITES DONNENT-ILS DES TAUX DIFFÉRENTS ?

Un lecteur attentif du premier tableau aura noté que les moteurs classiques sont vérifiés à 78,9 % sur un site et 96,5 % sur l'autre. Dix-huit points d'écart, même méthode, même résolveur. Anomalie ?

Nous avons résolu les adresses une par une. 67 adresses distinctes se déclaraient Googlebot sur fast-growth.fr : 7 avec une résolution Google authentique, 3 pointant vers des tiers, 57 sans résolution du tout. Sur nomo-ia.com, 20 adresses déclarées, dont les 7 mêmes authentiques.

Les vrais Googlebot sont identiques des deux côtés. Ce sont les faux qui font la différence. Parmi les usurpateurs, des machines hébergées chez des fournisseurs où n'importe qui loue une machine à l'heure.

L'écart n'est donc pas un défaut de méthode. C'est une mesure de l'exposition de chaque site aux imposteurs, une grandeur qu'aucun outil déclaratif ne peut produire, puisqu'il ne sait pas qu'ils sont des imposteurs. Un site plus visible, plus ancien ou plus attaqué porte plus de faux Googlebot. Le chiffre de contrôle contrôle bien quelque chose.

SECTION 07

LES CANAUX D'ACQUISITION MENTENT-ILS AUSSI ?

La vérification d'identité corrige le comptage des robots. Elle ne corrige pas tout, et deux angles morts méritent d'être documentés, parce qu'ils touchent tous les outils du marché, le nôtre compris.

Le canal referral d'abord. Sur la même fenêtre, 12 087 requêtes de nos journaux portaient un site référent, définition classique du trafic venu d'ailleurs. 10 740 d'entre elles, soit 88,9 %, étaient de la navigation interne : un visiteur déjà sur le site cliquant vers une autre page du même site, que la chaîne de classement rangeait en acquisition faute de règle pour ce cas. Le quatrième canal de notre écran d'acquisition était faux à près de 90 %. Nous avons corrigé la règle. Combien d'écrans d'acquisition, ailleurs, comptent encore leur propre navigation comme des visites référées ?

Les referrals IA ensuite. Sur la fenêtre, quatre requêtes référées par une interface d'IA. Ce chiffre est un plancher, pas une mesure : les applications natives de ChatGPT ou de Claude n'envoient pas d'en-tête de provenance, donc une part inconnaissable de ce trafic arrive en direct. Tout « zéro referral IA », le nôtre compris, se lit « zéro referral porteur d'en-tête », jamais « zéro visiteur envoyé ».

À ce propos, les ratios d'aspiration qui circulent.

SOURCE	RATIO	FENÊTRE	PÉRIMÈTRE
Cloudflare	70 900 : 1	juin 2025	robots d'Anthropic
Cloudflare	≈ 4 580 : 1	28 jours, fin juin 2026	réseau Cloudflare
Microsoft Clarity	≈ 6 000 : 1	août 2026	exemple publié

Trois ratios entre pages aspirées et visiteurs renvoyés. Tous réels, aucun comparable aux autres.

Les fenêtres, les périmètres et les méthodes diffèrent. Citer un ratio sans sa fenêtre, c'est citer un titre, pas une mesure. Et comme ces ratios se calculent sur des referrals sous-comptés, ils surestiment mécaniquement, ce que leurs propres auteurs signalent.

SECTION 08

QUI PASSE VRAIMENT, UNE FOIS LES PAPIERS VÉRIFIÉS ?

Une fois le trafic filtré sur les identités confirmées, le paysage réel surprend. Ces ordres de grandeur ont été mesurés sur dix-sept jours, deux sites, avant la fenêtre principale, et restent stables depuis.

Sur les fetchs déclenchés par un prompt réel, OpenAI représente la totalité du trafic vérifié. Sur l'indexation IA, le premier robot n'est ni OpenAI ni Anthropic : c'est PetalBot, le robot de Huawei, à 55 et 68 % selon le site. Sur la collecte d'entraînement, amazonbot fait environ quarante fois le volume de GPTBot.

Et le plus important est invisible par construction. Gemini ne visite jamais votre serveur : il lit l'index de Google, constitué par Googlebot pour la recherche classique. Un tableau de bord de fetchs, aussi vérifié soit-il, est structurellement aveugle à l'un des principaux moteurs de réponse du marché.

Vos outils vous montrent qui passe. Ils ne peuvent pas vous montrer qui vous lit sans passer.

SECTION 09

QUELLES SONT LES LIMITES DE NOTRE MESURE ?

Un document qui reproche au marché son déclaratif se doit d'être irréprochable sur ses propres trous. Les voici.

Les 792 requêtes non tranchées de fast-growth.fr, 12,3 % du volume d'agents IA, concernent un seul robot, meta-externalagent, dont la vérification aboutit sur un site et pas sur l'autre. Ce n'est ni une usurpation ni une impossibilité : c'est une lacune de notre instrument, non élucidée à ce jour, comptée à part et déclarée ici plutôt que corrigée en silence.

Les volumes sont ceux de deux sites de petite taille : 154 fetchs authentifiés en treize jours sur le site principal, 24 sur le second. Trop peu pour porter un chiffre seuls. Nous ne décrivons pas le web, nous décrivons une méthode et son effet sur les chiffres.

La fenêtre est courte, treize jours, et sera étendue. Enfin, la classification des catégories elles-mêmes, index, entraînement, fetch, reprend les déclarations des éditeurs sur le rôle de leurs robots, déclarations que nous ne pouvons pas vérifier.

SECTION 10

QUELLE QUESTION POSER À SON OUTIL DE MESURE ?

Pour un dirigeant, tout ce document se replie en une question, à poser à son outil ou à son prestataire : sur quoi repose l'identification des robots dans ce rapport ?

Si la réponse est le user-agent, les volumes affichés mélangent robots et imposteurs dans des proportions que ce document permet d'imaginer.

Le standard, lui, avance. Les signatures cryptographiques de Web Bot Auth, décrites par la RFC 9421, donneraient une vérification forte sans résolution inverse, et OpenAI publie déjà ses plages d'adresses à des adresses stables. Le jour où signer sera la norme, la question changera de camp : elle se posera aux robots qui ne signent pas, et aux éditeurs qui ne publient rien.

En attendant, la vérification dans les deux sens reste le meilleur outil disponible, et elle est à la portée de quiconque lit ses journaux serveur. Compter juste est possible dès aujourd'hui, et c'est le préalable à toute stratégie de visibilité IA : avant d'optimiser pour des robots, s'assurer que ce sont des robots.

SECTION 11

QUESTIONS FRÉQUENTES

Comment vérifier l'identité d'un robot qui visite mon site ?

Trois mécanismes existent, du plus fort au plus faible : la signature cryptographique définie par la RFC 9421, la plage d'adresses publiée par l'éditeur, et la résolution inverse confirmée dans les deux sens. Cette dernière consiste à demander le nom associé à l'adresse, puis à vérifier que ce nom pointe bien vers cette adresse. L'aller simple ne suffit pas : un imposteur peut configurer le retour.

Pourquoi le taux d'erreur des crawlers IA est-il si élevé dans les études ?

Parce qu'il est calculé sur des déclarations. Sur nos journaux, les requêtes authentifiées affichent zéro page inexistante sur 154 requêtes, tandis que les usurpations prouvées en affichent 69,1 %. Le total non filtré, seule lecture possible sans vérification d'identité, ressort à 53,7 %. Un outil qui ne vérifie pas attribue donc aux moteurs le comportement de ceux qui les imitent.

ClaudeBot est-il un robot suspect puisqu'il n'est jamais vérifié ?

Non. Il est invérifiable, ce qui est différent d'une usurpation. Anthropic ne publie ni plage d'adresses ni enregistrement inverse permettant le test, donc aucun verdict ne peut être rendu. Confondre les deux situations reviendrait à accuser un robot faute de pouvoir l'identifier.

Peut-on comparer les ratios d'aspiration publiés par les différents acteurs ?

Non, sauf à comparer leurs fenêtres et leurs périmètres. Les trois ratios qui circulent, 70 900 pour 1, environ 4 580 pour 1 et environ 6 000 pour 1, sont tous réels mais portent sur des périodes, des réseaux et des méthodes différents. Un ratio cité sans sa fenêtre est un titre, pas une mesure.

Mon outil de visibilité IA est-il fiable ?

La question à lui poser tient en une phrase : sur quoi repose l'identification des robots dans ce rapport ? Si la réponse est le user-agent, les volumes mélangent robots authentiques et imposteurs. Si la réponse mentionne des plages d'adresses publiées ou une résolution inverse dans les deux sens, l'outil sait de quoi il parle.

SECTION 12

GLOSSAIRE

User-agent

Ligne de texte qu'un visiteur automatique écrit lui-même pour annoncer son identité. Champ libre, donc falsifiable : le compter revient à compter des déclarations.

Résolution inverse dans les deux sens

Vérification consistant à demander le nom associé à une adresse, puis à confirmer que ce nom pointe vers cette même adresse. Seul l'aller-retour est probant, l'aller simple étant configurable par un imposteur.

Usurpation prouvée

Verdict attribué quand la vérification était possible et a échoué : l'adresse n'appartient pas à l'éditeur déclaré. À distinguer de l'invérifiable.

Invérifiable

Verdict attribué quand l'éditeur ne publie ni plage d'adresses ni enregistrement inverse. Aucun test n'est possible, aucune faute n'est établie.

Fetch utilisateur

Requête d'un assistant qui va chercher une page à la demande d'un humain, par opposition à l'indexation et à la collecte d'entraînement.

Web Bot Auth

Mécanisme de signature cryptographique des requêtes de robots, appuyé sur la RFC 9421. Il donnerait une vérification forte sans dépendre de la résolution inverse.

Ratio d'aspiration

Rapport entre pages aspirées par les robots d'un éditeur et visiteurs renvoyés par lui. Sensible à la fenêtre et au périmètre, donc non comparable d'une source à l'autre.

SECTION 13

SOURCES

1. Vercel et MERJ (cabinet d'analyse d'audience). *The rise of the AI crawler*, 17 décembre 2024. Taux de pages inexistantes de 34,82 % pour ChatGPT, 34,16 % pour Claude et 8,22 % pour Googlebot. vercel.com
2. OpenAI, plages d'adresses publiées pour ses robots, au format JSON, à des adresses stables. [gptbot.json](#), [searchbot.json](#), [chatgpt-user.json](#)
3. RFC 9421, *HTTP Message Signatures*, IETF (Internet Engineering Task Force), février 2024. Base technique de Web Bot Auth. rfc-editor.org
4. Cloudflare. *From Googlebot to GPTBot: who's crawling your site in 2025*. Ratios entre pages aspirées et visiteurs renvoyés. blog.cloudflare.com
5. Microsoft Clarity. *Scrape-to-Referral insights*, 13 août 2026. Ratio d'environ 6 000 pour 1 dans l'exemple publié. clarity.microsoft.com
6. Mesures Fast Growth Advisors, journaux serveur de fast-growth.fr et nomo-ia.com, fenêtre du 11 au 23 août 2026. Bases indiquées avec chaque chiffre de ce document.