

L'Observatoire du GESTE sur l'IA & les crawlers



L'Observatoire du GESTE sur l'IA & les crawlers

- Dans le cadre de l'observatoire du GESTE, les éditeurs fournissent leurs logs de connexions.
- Botscorner croise ces logs avec sa base de bots, et fournit des informations B2B sur les bots : IA, agents-IA, datasets, veille des médias, ad-techs, SEO, RSS, etc.

GESTE
Les éditeurs de contenus et services en ligne

ouest
france 


leParisien.fr

Les Echos

Challenge^s.fr
L'économie en temps réel

mind
MEDIA

Europe 1

Capital

LE FIGARO

L'ÉQUIPE

Le Télégramme

TF1
NEWS

BFM
TV.

RTL

TV5MONDE

Business Map

Indexation Extraction

Contenus sous droits d'auteur

- Presse,
- Livres,
- Bases de données,
- Vidéos,
- Photos,
- Blogs,

...

Crawlers: RSS managers

Scrapers:

Proxies rotatifs

Brokers d'IPs

services:

- Scrapy
- Inoreader
- Feeder.co
- Oculus
- Ipxo
- Vultr
- Rayobyte
- Oxylabs
- Smartproxy...

Datasets

- Webshare
- Webzio
- Brightdata
- Adatapro
- NewsApi
- Diffbot
- Archive.org
- CommonCrawl
- Newsapi.ai

Bases de données standardisées

exploitation

valorisation Ad-Tech SEO

Media Monitoring

IA-LLM

Services à base d'IA:

- Analyses,
- Tendances, news monitoring
- e-reputation
- assistant business
- Modération
- Traduction,
- creation : articles

Images, vidéos

etc



Les éditeurs de contenus et services en ligne

Marchés

- B2B
- B2B2C
- B2C



cartographie détaillée du trafic B2B

Les BOTS

Les BOTS représentent près de 50% du trafic

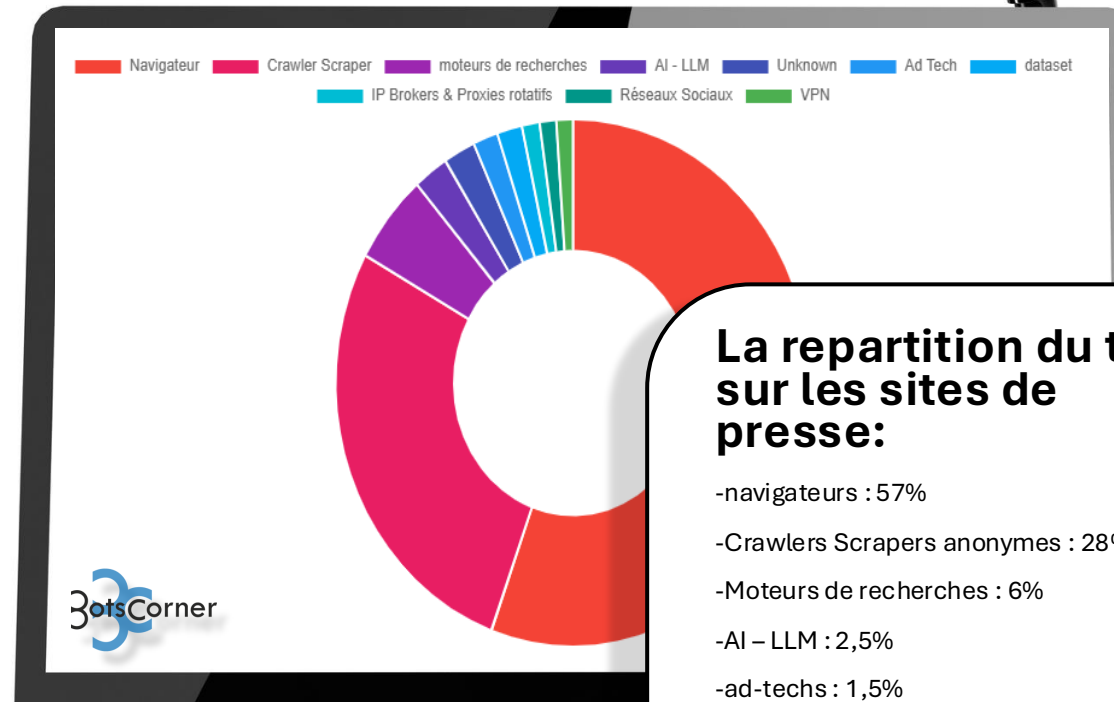
12% consultent le fichier robots.txt (consulter n'est pas respecter)

À part les principaux moteurs de recherches , réseaux sociaux et adtechs et quelques LLMs sur certaines activités,

très peu de bots affichent un user agent explicite.

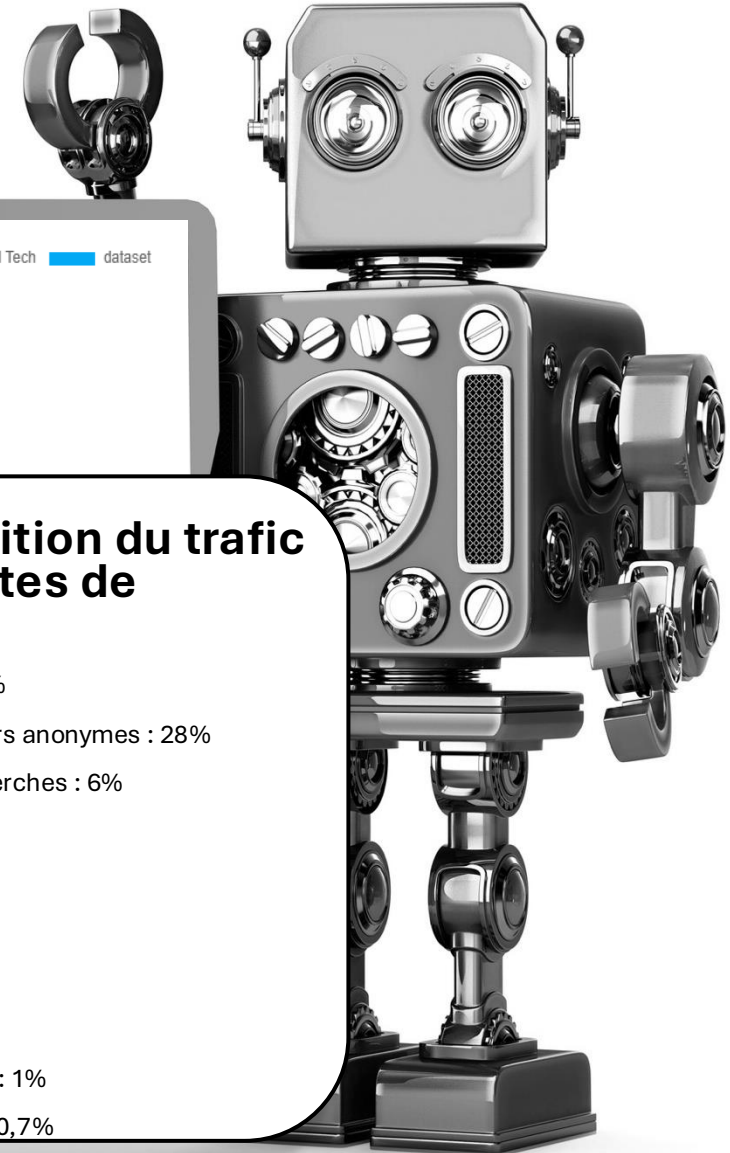
GESTE

Les éditeurs de contenus et services en ligne



La repartition du trafic sur les sites de presse:

- navigateurs : 57%
- Crawlers Scrapers anonymes : 28%
- Moteurs de recherches : 6%
- AI - LLM : 2,5%
- ad-techs : 1,5%
- datasets : 1,2%
- IP Brokers : 1,1%
- VPN : 1%
- Réseaux Sociaux : 1%
- B2B data users : 0,7%

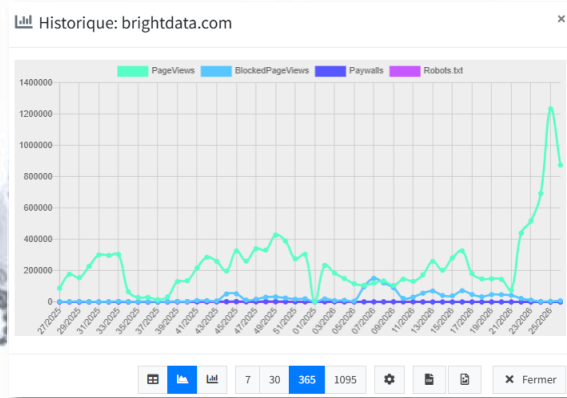


cartographie détaillée du trafic
B2B

Les IA Les agents

L'activité des IA

- les IA américaines qui posent leur user agent, déclarent des ips de crawls
- Où sont les Chinois, les Français, ... ?
- Quelques agents , du rag, ...



Infomédiaire	Client	Opérateur	Sessions	Hits	Pages	Volumes	Robots.txt
	Amazon Searc...		289.44 k	1.37 mn	196.98 k	59.66 Go	1
Alibaba	scrap Alibaba ...		159.56 k	539.08 k	93.10 k	8.98 Go	6
apple	AppleBot		304.66 k	337.61 k	230.45 k	19.11 Go	121
Tencent	Tencent (Hun...		193.95 k	273.15 k	134.52 k	11.53 Go	25
AMAZON	Amazonbot (R...		151.30 k	179.34 k	166.25 k	22.59 Go	6
OpenAI	OpenAI ChatG...		742	82.73 k	82.61 k	17.55 Go	2
OpenAI	OpenAI Searc...		17.10 k	61.13 k	60.77 k	9.83 Go	278
Cloudflare platfor...	training LLM f...		8.19 k	53.75 k	46.59 k	3.74 Go	0
Anthropic	Claude-Searc...		1.16 k	37.15 k	37.11 k	2.85 Go	35
Yandex	Yandex (GPT, ...		43.81 k	35.56 k	28.48 k	3.16 Go	1.51 k

IA LLMs

Affichage de 1 à 10 sur 64 entrées

Afficher 10 entrées

Infomédiaire	Client	Opérateur	Sessions	Hits	Pages	Volumes	Robots.txt
OpenAI	OpenAI ChatGP...		101.16 k	146.92 k	146.84 k	14.71 Go	0
	OpenAI ChatGP...		4.33 k	102.36 k	26.06 k	2.33 Go	0
	Glance AI, Inc		1.46 k	78.36 k	7.49 k	2.11 Go	0
AMAZON	Amazon Quick		237	48.01 k	2.20 k	321.12 Mo	25
	AI Agents on He...		4.23 k	30.84 k	30.00 k	3.96 Go	13
	easyreponses.c...		163	28.88 k	28.87 k	5.73 Go	2
	aranet.ai		3.67 k	15.33 k	15.11 k	1.13 Go	241
Cloudflare platfor...	training Agents f...		14.45 k	14.91 k	5.66 k	772.03 Mo	1
	linkup.so (rag a...		654	6.33 k	6.33 k	276.32 Mo	0
	naget.com (AI a...		1.86 k	5.97 k	5.96 k	472.84 Mo	0

Rag, agents

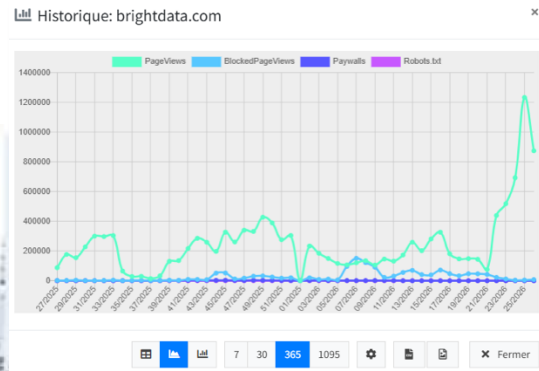
Affichage de 1 à 10 sur 233 entrées

cartographie détaillée du trafic B2B

Les datasets Les brokers Les scrapers

L'alternative pour l'approvisionnement : les datasets et plateformes, les brokers d'ips , les scrapers

- des services connectés
- un prix bas
- une base standardisée
- l'anonymat de l'utilisateur



Informédaire	Client	Opérateur	Sessions	Hits	Pages	Volumes	Robots.txt
fake browser (scr...	Fake Browser ...		217.82 k	3.45 mn	292.89 k	55.80 Go	302
	bc.googleuse...		107.87 k	1.25 mn	1.16 mn	32.49 Go	102
fake browser (scr...	fake browser ...		39.75 k	361.25 k	335.42 k	13.93 Go	0
	Meltwater? W...		121.23 k	164.79 k	164.79 k	10.43 Go	0
Diffbot	DiffBot.com		6.36 k	142.29 k	86.72 k	4.14 Go	728
	Scrap google...		158.85 k	132.53 k	119.03 k	10.14 Go	52
supabase	Supabase (Po...	à bloquer	61.26 k	120.42 k	77.28 k	6.92 Go	477
Infomédia-Retrie...	Opoint.com	Licencié CFC	20.99 k	102.85 k	71.23 k	7.41 Go	0
	brightdata.com	Bloqué	93.71 k	92.40 k	92.40 k	57.55 Go	0
Emeigh Investme...	sprious.com ...	Bloqué	13.55 k	92.28 k	24.81 k	4.21 Go	2

Affichage de 1 à 10 sur 205 entrées

datasets

Client	Opérateur	Hits	Pages	Volumes	Robots.txt
[MOZILLA/5.0] ...		2.71 mn	18.44 k	173.37 Go	3.49 k
Headless Chro...		953.87 k	154.47 k	10.20 Go	35
ASN-CXA		575.12 k	113.31 k	18.39 Go	200
Axios (used by ...		302.03 k	301.28 k	3.05 Go	0
scraping huaw...		201.00 k	74.21 k	7.31 Go	13
SQL Injection ...		187.61 k	187.61 k	8.46 Go	0
Scrap VNPPT Vi...		182.49 k	142.25 k	9.50 Go	1
OVH 2001:41d...		145.62 k	136.50 k	33.90 Go	55
Ktor		132.57 k	101.00 k	2.33 Go	0
scraper micros...		128.13 k	79.55 k	7.65 Go	

r 940 entrées

scrapers

Client	Opérateur	Hits	Pages	Volumes	Robots.txt
Various Brokers		2.28 mn	712.05 k	88.08 Go	289
IPXO 2, Lituan...		1.23 mn	33.93 k	11.76 Go	7
IPXO 1 (Gnowit...		196.41 k	80.09 k	6.72 Go	6
Prefixbroker		162.54 k	99.94 k	3.63 Go	1
IPXO 3 , Bulgar...		90.73 k	30.31 k	1.76 Go	3
Latitude		42.05 k	16.92 k	1.39 Go	4
CODE200 (bro...		33.42 k	8.02 k	1.38 Go	2
LogicWeb		32.17 k	15.32 k	787.14 Mo	0
FineGroupServ...		13.86 k	11.44 k	2.39 Go	0
scrap Selectel ...		13.71 k	10.94 k	1.38 Go	0

sur 40 entrées

IP brokers

cartographie détaillée
du trafic B2B

Le
marché

Les datasets existent
depuis 25 ans.

Exemple de tarif:

Diffbot:

1,000,000 Credits (\$0.0009 per
Credit)

1 credit = Extracting a webpage

zyte

Get website pricing. Instantly.
From \$0.06 per 1,000 successful responses

Zyte API Zyte Data Scrapy Cloud

DATA PRO **ng We're hiring** **MENU**

Training Data for AI
Get high-quality training data and build reliable AI models.

DataSet

Unlimited Value Starts Here.
Experience the full potential of your data at a fraction of the price. No gotchas.

\$0.99
Per GB
Per day. Billed annually.*

DIFFBOT Data Solutions Products Pricing Resources Company Login Signup

Plans & Pricing
Plans starting at Free. No contracts required.

Get Started for Free
No credit card required. Full API access.

FREE \$0/mo	STARTUP \$299/mo	PLUS \$899/mo	ENTERPRISE Custom
Jumpstart hobby projects. No credit card required. Free forever. See rate limits.	Small teams looking for easy, plug-and-play scraping and Knowledge Graph access.	Everything in Startup, plus access to Crawl and discounted credit pricing.	Bespoke plans and managed solutions with the full works, including premium SLA support.
Get Started for Free	Get Started	Get Started	Schedule a Demo
Product Access <ul style="list-style-type: none">✓ Extract✗ Bulk Extract✗ Crawl✓ Natural Language✓ Knowledge Graph Search✓ Knowledge Graph Enhance	Products Included <ul style="list-style-type: none">✓ Extract✗ Bulk Extract✗ Crawl✓ Natural Language✓ Knowledge Graph Search✓ Knowledge Graph Enhance	Product Access <ul style="list-style-type: none">✓ Extract✓ Bulk Extract✓ 25 Active Crawls✓ Natural Language✓ Knowledge Graph Search✓ Knowledge Graph Enhance	Product Access <ul style="list-style-type: none">✓ Extract✓ Bulk Extract✓ 100+ Active Crawls✓ Natural Language✓ Knowledge Graph Search✓ Knowledge Graph Enhance
Usage & Features <ul style="list-style-type: none">✓ 10,000 Credits (\$0 per credit)	Usage & Features <ul style="list-style-type: none">✓ 250,000 Credits (\$0.01 per Credit)	Usage & Features <ul style="list-style-type: none">✓ 1,000,000 Credits (\$0.0009 per Credit)	Usage & Features <ul style="list-style-type: none">✓ Custom Credit Allotment & Rate

coresignal Agentic search Datasets Data APIs Use cases Resources Pricing Contact us Log in Start free

API solutions: get direct access to AI-ready data
✓ Free trial ✓ JSON, JSONL, and CSV formats ✓ Updates in real time

Free	Starter	Pro	Premium
Starting from \$0 No credit card required	Starting from \$49 /month \$0.196-\$0.133/record	RECOMMENDED Starting from \$800 /month \$0.080-\$0.050/record	Starting from \$1,500 /month \$0.030-\$0.005/record
Start free	Sign up	Sign up	Sign up
✓ Includes 200 Collect, 400 Search credits	✓ Includes at least 250 Collect, 500 Search credits	✓ Includes at least 10,000 Collect, 20,000 Search credits	✓ Includes at least 50,000 Collect, 150,000 Search credits

bright data Produits Data pour AI Tarifs Resources Docs

5K page loads	\$1.5 /1K page loads	\$499 /mois
✓ 5K page loads per month	✓ Payez uniquement en cas de succès	✓ 383K page loads included
✓ Aucune carte de crédit requise	✓ Fixez des limites de dépense mensuelles	✓ \$1.3/1K additional page loads
✓ Support d'experts	✓ Concurrence illimitée	✓ Concurrence illimitée
	✓ Support d'experts	✓ Annuler à tout moment
		✓ Support d'experts

GESTE

Les éditeurs de contenus et services en ligne

Merci !



contact@botscorner.com