Global BOT Security Report 2025

The state of bot attack & AI traffic readiness across 17,000 websites worldwide

DataDome

Table of contents

About this report

DataDome's mission is to free the web of fraudulent traffic.

The annual Global Bot Security Report is one way we advance that mission, providing organizations with clear, data-backed insights into how well websites around the world stand up against automated attacks.

For the 2025 edition, we tested slightly under 17,000 popular websites across industries and regions using our Vulnerability Scan, examining their resilience against common automated threats. This year's scope also expanded to assess how websites handle AI crawler activity, a fast-growing concern for businesses seeking to protect proprietary content from large-scale scraping and theft.

We use the findings from this report to identify trends, highlight gaps in protection, and inform the public about the evolving bot, cyberfraud, and AI-driven threat landscape. Many of this year's results continue a concerning trend: **most websites still cannot block even the simplest automated attacks.**

The goal of this report is to give businesses a clear picture of their risk, the attack vectors they should be aware of, and where vulnerabilities remain, so they can take informed action to safeguard their users, data, and revenue.

Executive summary

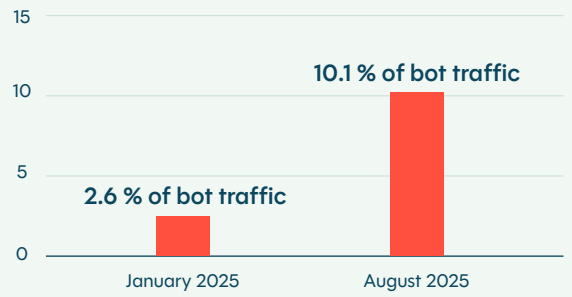
Bots remain one of the most relentless threats to online businesses and a primary driver of digital fraud. Over the past year, bot attacks have not only increased in volume but in sophistication. DataDome alone now blocks over **400 billion bot attacks annually**, a **14% year-over-year increase**.

The reality is that in 2025, the majority of online businesses remain fully exposed to even **simple, script-based bots** that carry out scraping, credential stuffing, carding, fake account creations, and more.

Our 2025 report reveals that across **16,900+ popular domains**, only **2.8%** are fully protected against the threat vectors we tested—down sharply from **8.4% in 2024**. Even more concerning: **over 61%** of domains failed to detect any of our test bots, meaning they are completely exposed. This widespread exposure to even very simple bots means businesses are extremely vulnerable to the far more sophisticated, AI-powered bot attacks we're seeing in 2025.

So far this year, **large language model (LLM) crawler traffic has more than 4X'd** across DataDome's global customer base, rising from **2.6% of all verified bot traffic in January to over 10.1% by August of 2025**. These crawlers, often tied to generative AI training efforts, scrape web content at unprecedented scale, usually without consent or oversight. In August, DataDome detected nearly **1.7 billion requests from OpenAI crawlers alone**.

x4 increase in LLM traffic between January and August 2025



LLM crawler traffic as % of overall bot traffic

1.7B
Requests from OpenAI
(Aug 2025)

Meanwhile, threat actors are increasingly blending basic automation with more advanced tactics such as leveraging **agentic AI tools** that simulate human behavior, bypass CAPTCHAs, manipulate TLS fingerprints, and adapt in real time. These bots aren't just executing instructions; they're making decisions in real time.

The result: attackers are combining both simple and sophisticated bots, often within a single attack strategy.

This year's Global Bot Security Report explores:

- **Why simple bots continue to pose a serious threat, and why most businesses are still vulnerable**
- **How resilience against bot attacks has declined year-over-year, leaving even large enterprises at risk**
- **The rapid rise of AI agents and LLM crawlers harvesting data at an unprecedented scale**

The takeaway is clear: if your defenses can't stop simple bots, then you're underprepared for what's now becoming the norm.

> **As AI agents become more common, bot protection must evolve. It's no longer enough to ask "is this a bot or a human?" — we must ask "what is the intent behind this action?"**

To stay secure in an AI-driven world, businesses need adaptive, intent-based bot and cyberfraud protection that can analyze behavior in real time, not just identity or static signals. The future of security depends on understanding not just who is behind the request, but why they're there.

– **DataDome Advanced Threat Research**

Key findings from the 2025 report

In testing nearly 17,000 popular domains with the DataDome Vulnerability Scan, we found that the majority of websites remain vulnerable to bot attacks. In many cases, vulnerability rates have actually worsened since 2024.

Most sites still fail to detect and block simple bots

In 2025, 61.2% of 16,900+ tested websites were unprotected, 36% were partially protected, and only 2.8% were fully protected. This means 3 out of 5 domains remain fully unprotected against simple bot attacks.

Full protection rates have worsened since 2024

Compared to 2024, the percentage of fully protected websites dropped sharply, from 8.4% in 2024 to 2.8% in 2025. However, there has been an increase in domains that scored as partially protected (26.4% in 2024 and now 36% in 2025). Partial protection means that at least one type of bot was detected, but not all of them.

AI-driven bots now frequently target high-value endpoints

In 2025, 64% of AI bot traffic touched forms, 23% touched login pages, and 5% reached checkout flows, introducing new fraud, compliance, and security risks. The rise of agentic AI and LLMs makes it essential to assess intent, since not all AI traffic is bad. The outdated 'bot or not' model no longer applies.

LLM crawlers surged in volume

LLM crawler traffic more than 4X'd between January and August of 2025 across DataDome's customer base. OpenAI's GPTBot alone made over 1.7 billion requests to DataDome customer websites in August 2025.

Businesses are moving to block AI crawlers

88.9% of robots.txt files in our dataset explicitly disallow GPTBot, making it the most-referenced AI crawler by far. This points to a broader trend: businesses are increasingly blocking AI traffic amid growing concerns over content theft and data misuse.

High-risk industries remain exposed

In 2025, the weakest sectors were Government, Non-Profit, and Telecoms, showing the lowest levels of bot protection. Meanwhile, Travel & Hospitality, Gambling, and Real Estate led the way with the highest combined rates of full and partial protection.

Bot protection tools from popular vendors are inconsistent

Across widely used vendors, detection rates for our test bots ranged from just 6% to 42%—revealing major gaps in real-world effectiveness, even among providers claiming bot mitigation as a core capability. (DataDome was excluded from this comparison.)

The most-trafficked websites are among the least protected

Only 2% of domains with over 30M monthly visits were fully protected, suggesting that scale alone does not translate to better protection.

Bigger doesn't mean safer

Even among the largest organizations (10,001+ employees), just 2.2% of their domains were fully protected, with 61% unprotected.

Some bots are far harder to detect than others

Detection remained weakest for advanced, anti-fingerprinting bots, which were only blocked by ~7% of targets—leaving most organizations highly vulnerable to account takeover, carding, and advanced scraping attacks. Fake Chrome and Curl bots were detected just 21% of the time, while 79% evaded defenses.

The rise of AI-powered automation is exacerbating the risks of inadequate bot protection. As organizations grapple with rapidly increasing volumes of traffic from AI-powered bots, AI agents, and LLM crawlers, gaps in basic bot protection leave them even more vulnerable to attacks and fraud.

Interestingly, our research shows that **most websites are now actively looking to control AI crawler traffic**. The vast majority of the sites we tested disallow AI bots via their robots.txt files, underscoring growing concern over AI-driven scraping, content theft, and data leaks. Unfortunately, robots.txt directions do not reliably protect against unwanted AI traffic.

Methodology

The DataDome Vulnerability Scan is a safe, non-intrusive testing solution designed to identify vulnerabilities in websites without disrupting normal operations or causing any harm. The tool is accessible and free to the public to use and can be found here.

To test a website's defenses, the Vulnerability Scan sends a controlled, low-volume of traffic using a range of bot profiles used in real-world attacks and fraud schemes. These include:

- **Curl command bots:** Basic bots simulating simple scripted requests
- **Fake Googlebot:** Bots that spoof the identity of Google's crawler. This helps us to test whether the site properly verifies user-agent claims
- **Fake Chrome bots:** Bots that mimic real browser behavior (e.g., Chrome). This helps us to test for superficial bot detection based on headers or JavaScript execution
- **Anti-fingerprinting headless browsers:** Advanced headless bots that evade detection by masking common fingerprinting signals (e.g., headless flags, WebGL spoofing, etc.).

These profiles are designed to simulate both basic and advanced attack techniques, helping IT, security, and fraud teams assess how their site responds to everything from obvious bots to stealthier threats. All tests were performed ethically, in line with DataDome's commitment to safe and responsible security research. Protection status is determined based on how a site responds to each bot profile:

- **Unprotected:** The site fails to block even a single bot profile
- **Partially protected:** The site blocks some bot profiles, but not all
- **Protected:** The site successfully blocks all bot profiles used in the test *

Note: The DataDome Vulnerability Scan does not determine whether a website is fully protected against all possible bot attacks. A site that successfully passes each test in this scan may still be vulnerable to sophisticated adversaries who use techniques such as heavily modified automated browsers, native JavaScript execution, forged browser fingerprinting signals, or AI-assisted evasion to bypass basic detection.

Expanded testing for LLM-readiness & AI traffic volumes

Since our last report, the rapid growth of large language model (LLM) crawlers and agentic AI has introduced a new class of automated threats. To reflect this shift, we've added two additional testing methodologies to this year's analysis:

- A review of robots.txt directives for all 16,900+ tested domains, to assess how sites are managing LLM crawlers like GPTBot, CCBot, and ChatGPT-User agents.
- A separate analysis of anonymized traffic across DataDome's customer base*, providing insight into which AI bots are visiting customers' websites and which endpoints receive the most AI traffic.

It's important to note that this test reflects a different sample than the 16,900+ domains tested through the Vulnerability Scan, which excludes all DataDome customers.

A note on the ethical use of bots

As a cybersecurity company, ethical and safe online behavior is of paramount importance to DataDome. Our mission is to eliminate cyberfraud from the web, without causing disruption or harm to the websites we test or their users.

The Vulnerability Scan has been carefully designed to cause zero impact on site infrastructure or user experience. It sends only a very small number of bot requests per test, ensuring no degradation in performance.

These requests are specifically engineered to assess a site's ability to identify automated traffic, not to overwhelm it. The Vulnerability Scan does not evaluate resilience against volumetric attacks such as DDoS or brute-force campaigns.

Data privacy is also a top priority. Our test bots do not collect, store, or process any private or sensitive data from the websites being tested or from their end users.

Study sample

This report draws on two distinct data sets to evaluate both bot protection maturity and AI traffic trends:

Vulnerability Scan sample (16,948 domains)

For our primary analysis, we tested 16,948 high-traffic websites (each averaging at least 5 million monthly visitors) across all major regions and industries using DataDome's Vulnerability Scan. This included bot detection assessments and a review of each domain's robots.txt file to determine how sites manage access for AI bots.

Importantly, this sample **excludes** all DataDome customers. Data collection occurred over multiple weeks and concluded on July 31, 2025. Sites were categorized by industry, geography, company size, and volume of monthly traffic.

AI traffic sample (DataDome customer traffic)

To supplement this, we conducted a separate analysis of anonymized traffic across hundreds of DataDome customer environments.

This allowed us to observe real-world AI crawler activity, identifying which types of endpoints are experiencing the highest volumes of AI-driven traffic. This sample is entirely separate from the 16,948 domains included in the Vulnerability Scan.

Bot attack vectors used in testing

In 2025, our Vulnerability Scan tested each domain against a representative set of automated threats designed to mimic the bots most commonly used in real-world fraud and abuse. While not exhaustive, these vectors cover a wide spectrum, from rudimentary scripts to advanced automation, revealing how well sites detect and block the tools bad actors rely on.

1.

Curl Command Bot

A simple command-line request tool that can fetch, upload, and manipulate web resources. Curl is often used in low-sophistication attacks to probe site structure, scrape content, or attempt bulk actions. Its straightforward nature means many protection systems should be able to detect it, yet unprotected endpoints can still allow Curl-based attacks to succeed.

2.

Fake Googlebot

Malicious crawlers that impersonate Google's official indexing bot to bypass access controls. These bots often scrape proprietary data, harvest product information, or monitor prices under the guise of legitimate search engine traffic. Detecting them requires verifying bot identity against authoritative Google IP ranges.

3.

Fake Chrome Bot

A bot that forges the HTTP headers, user-agent strings, and other indicators of Google Chrome. This allows it to blend in with normal user traffic while executing automated tasks at scale, such as credential stuffing, inventory denial, or content scraping. Fake Chrome bots exploit detection systems that rely solely on static browser fingerprinting.

4.

Advanced Bot (including Anti-Fingerprinting Headless Browser)

A sophisticated automation framework designed to evade traditional detection. Advanced bots can execute full JavaScript, maintain sessions, and mimic human-

like interactions such as mouse movements and scrolling. The anti-fingerprinting headless browser variant takes this a step further by randomizing or masking identifying signals like canvas fingerprints, screen size, and timezone, making it far harder to detect using standard behavioral or fingerprinting techniques. These bots are frequently used in high-value fraud scenarios, including account takeover, card testing, and targeted content theft.

Evaluating resilience against AI traffic

The DataDome Vulnerability Scan does not currently simulate LLM crawlers or AI agents. To understand how websites are handling these emerging threats, we analyzed the robots.txt files across our dataset using a custom parsing script.

While robots.txt is designed to guide compliant crawlers, it's important to note that **it is not a security mechanism**. LLMs and AI crawlers—especially those used for scraping or unauthorized data harvesting—can easily ignore or bypass these directives.

During our analysis, we found that:

- **6% of websites responded to robots.txt requests with a CAPTCHA challenge**, indicating attempts to guard access to this file, even though it is meant to be publicly accessible to bots.
- A large number of sites had **syntax errors or malformed robots.txt files**, which can result in unintended access permissions or ineffective controls.

Taken together, these findings highlight the growing complexity of managing AI-driven traffic and the limitations of relying solely on protocol-based approaches like robots.txt.

Geographic locations

This year's Global Bot Security Report analyzed domains worldwide.

Several regions had large enough samples to allow for a side-by-side comparison across global regions: Asia Pacific, Europe, Latin America, and North America.

The Vulnerability Scan tests were performed on domains in each region, covering a representative mix of industries, company sizes, and traffic volumes. This geographic diversity ensures that the findings reflect both regional differences in bot protection practices and global trends.

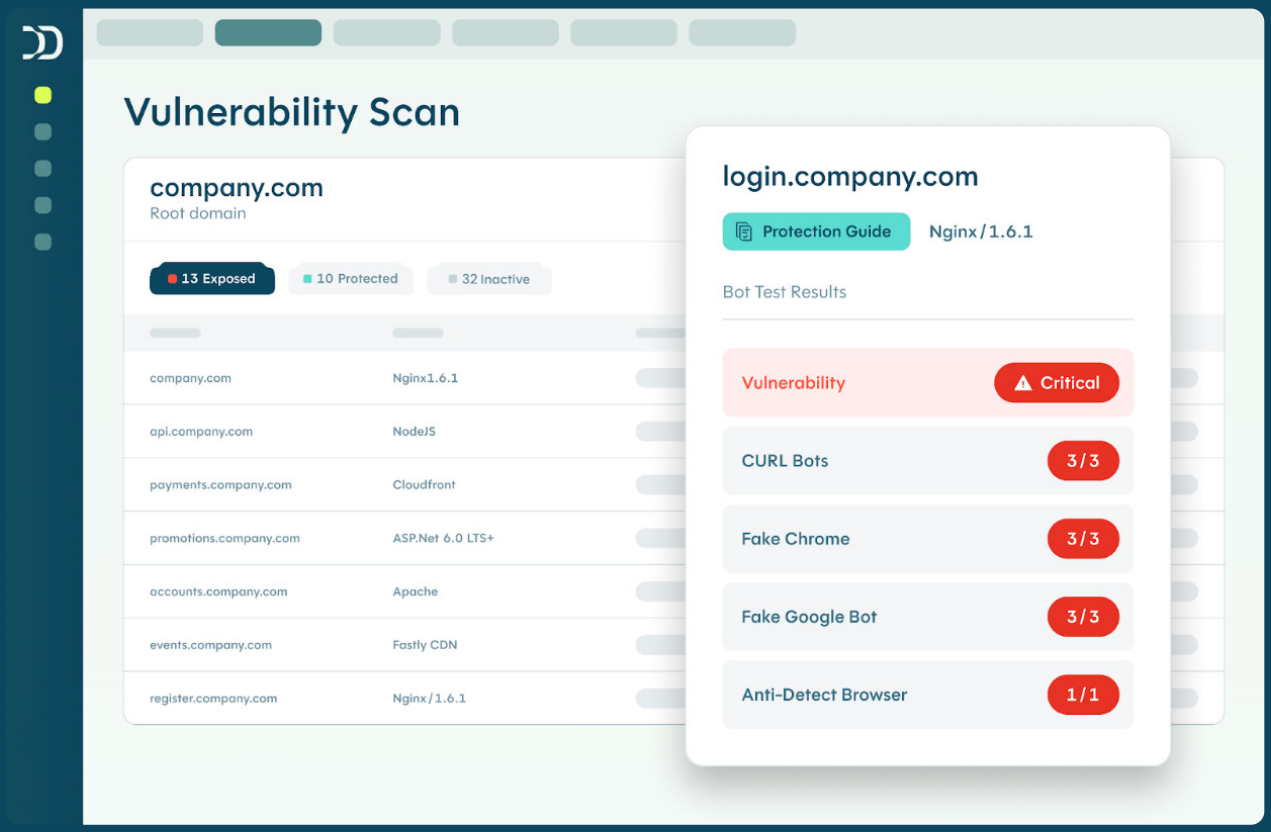
In some cases, the geographic origin of a request can impact detection results. For each type of bot attack vector, the Vulnerability Scan therefore uses residential proxies to send requests from three distinct locations: USA, Canada, and France.

Test results

Each domain was tested multiple times to validate findings, and the tests were designed to be non-intrusive, ensuring no disruption to the sites evaluated.

The Vulnerability Scan answers a simple question: Did the target website correctly identify the request as coming from a bot? If the answer is yes, and the bot requests are either blocked or challenged by a JavaScript proof of work/interstitial page (e.g., a CAPTCHA challenge), we consider the website to be protected against that type of bot attack. If not, the website is vulnerable to that category of attack vector.

A look at the DataDome Vulnerability Scan results



The state of bot attack preparedness

The following is an overview of our findings from testing 16,948 of some of the most popular global websites for bot attack preparedness.

These findings shed light on the prevailing state of bot protection, including:

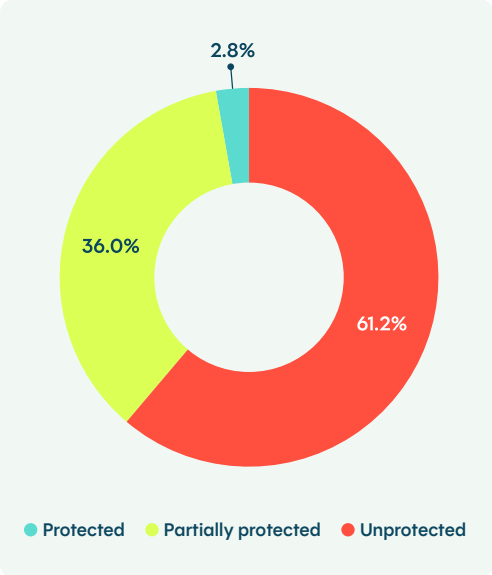
- **Protection across regions, industries, and business**
- **Variations in the performance of different bot detection systems**
- **The most commonly effective bot attack vectors**

3 out of 5

global domains remain unprotected against simple bot attacks

Of all the domains we tested:

- **Only 2.8% were fully protected**, successfully blocking all bot requests.
- **36.0% were partially protected**, detecting and blocking some, but not all, bot requests.
- **61.2%—roughly three in five organizations—were completely unprotected**, letting through all bot requests without detection.



Were websites protected against bots?

These results mark a notable change from our findings in 2024, where 8.44% of websites blocked all requests, 26.4% blocked some requests, and 65.2% blocked none.

The drop in full protection likely reflects the continuous updates we make to our Vulnerability Scan tool to better mirror today's bot landscape. Recent improvements—like support for advanced fingerprinting and residential proxy usage—make even basic test bots harder to detect. As a result, while some companies might have been slightly better prepared to stop last year's threats, they're now struggling to keep up with today's more advanced (but commonly used) tactics.

The sample size did grow (from 14,000 in 2024 to 16,900+ in 2025), but that alone doesn't explain the drop. What it reveals is more concerning: basic bot mitigation remains inefficient or absent across a majority of digital businesses worldwide, even against simple, low-cost attacks that are easy and cheap to deploy.

Key takeaway

A significant majority of global digital businesses remain vulnerable to simple bot attacks, leaving them exposed to scraping, account takeover (ATO), carding, and other forms of cyberfraud.

How AI is raising the stakes

Generative AI and automation frameworks are accelerating this threat curve:

- **Credential stuffing** now uses AI-driven password mutation to increase hit rates.
- **ATO campaigns** leverage LLM-powered tools to navigate login flows or interact with customer service portals like a human.
- **Carding** models identify which transaction patterns bypass scoring systems in real time.
- **Scraping** tools can structure and enrich stolen data instantly, making it market-ready in minutes.

AI also democratizes advanced bot capabilities, meaning even low-skill actors can deploy high-efficacy attacks using pre-packaged, AI-augmented bot kits purchased on underground markets.

Bot protection by region

Our 2025 data shows some differences in bot protection rates across the regions with the largest tested sample sizes: Asia Pacific, Europe, Latin America, and North America. Other regions were excluded from this regional comparison due to insufficient sample sizes.

Protection by region

Latin America stands out as the most protected region, with 3.49% of websites fully protected and 38.45% partially protected. This means that just over 41% of domains had at least some level of defense in place, leaving nearly 6 in 10 completely exposed.

North America and Europe follow a similar pattern. In North America, 3.25% of domains were fully protected and 36.99% partially protected, while in Europe, 2.52% were fully protected and 36.03% partially protected. In both regions, around 60% of websites had no bot protection at all.

Asia Pacific lags slightly, with the lowest rate of full protection at 1.59%, though nearly 30% of websites had partial defenses in place. Still, more than two-thirds of domains in the region allowed all bot requests through without challenge.

These results suggest that attackers face few regional barriers

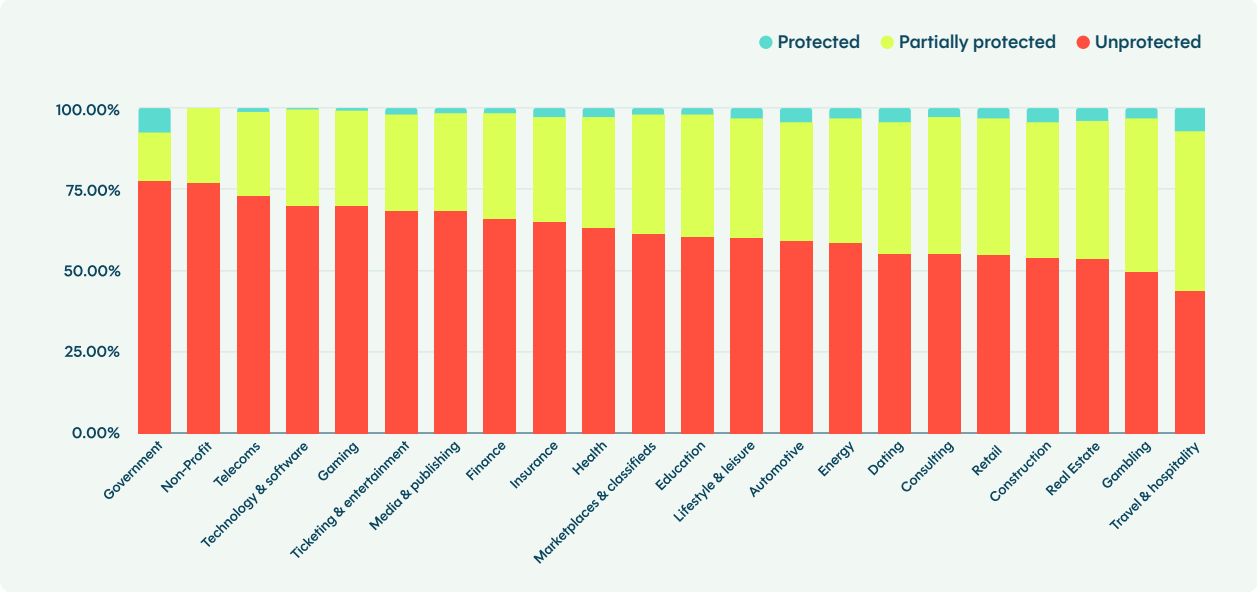
Weak bot defenses are a global constant. This uniformity works in fraudsters' favor, allowing them to reuse the same tools and tactics across markets with little need for adaptation. The result is lower costs, higher efficiency, and faster scaling for global cyberfraud campaigns. In short, the lack of regional variation in protection doesn't just leave businesses exposed but actively enables attackers to operate worldwide with ease.

Bot protection by industry

Bot attack preparedness continues to differ widely across industries, with some sectors making progress in partial defenses while others remain heavily exposed.

Last year, our tests included 18 main industries; this year, we've tested domains from 22 industries. Some new industries added to the list include **Automotive, Government, Non-profit, and Real Estate**.

We define the least protected industries as those with the highest percentage of websites reporting no protection. We define the most protected industries as those with the highest combined rate of full and partial protection.



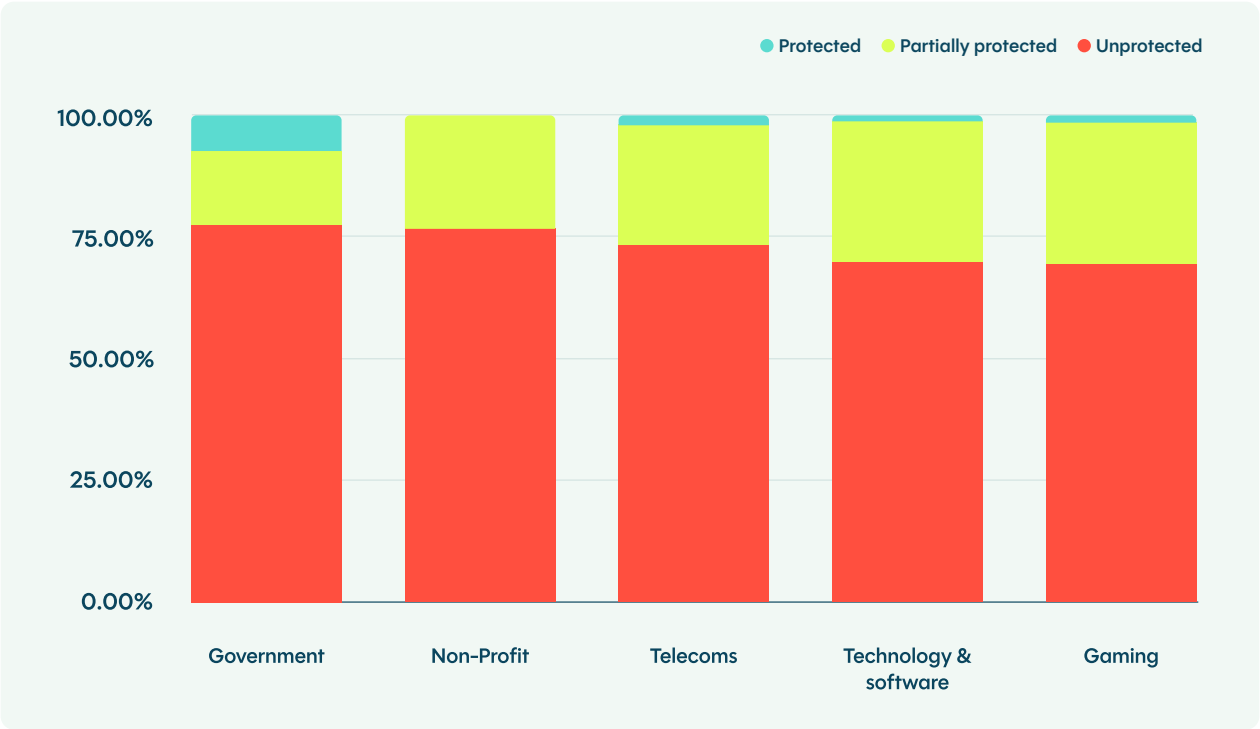
Protection by industry

The 5 least protected industries

(by % of unprotected websites)

These industries have the greatest proportion of websites without any bot protection in place:

- Government – **77.5% unprotected**
- Non-Profit – **76.9% unprotected**
- Telecoms – **73.0% unprotected**
- Technology & software – **69.9% unprotected**
- Gaming – **69.8% unprotected**



The 5 least protected industries (by % of unprotected websites)

These industries—where digital services, customer data, and transactions are core to operations—remain highly exposed to automated abuse.

Government

With nearly 4 in 5 domains unprotected, the public sector is dangerously exposed. These websites often handle citizen data, process transactions for fines and taxes, deliver essential digital services, and provide access to infrastructure, making them high-value targets for disruption, DDoS attacks, surveillance, and data theft.

Non-Profit

These organizations manage sensitive donor data and facilitate online giving, yet often lack the budget or staff for robust security. That leaves them highly vulnerable to scraping, phishing infrastructure, and fraud campaigns that exploit trust-based relationships.

Telecoms

Telecom companies are frequent targets for account fraud, credential stuffing, and SIM swap attacks—all of which rely on automation. With 73% unprotected, this sector is alarmingly exposed to scalable abuse.

Technology & software

Nearly 70% of tech companies failed to block even basic bots—a red flag for an industry built on digital infrastructure. These businesses operate APIs, cloud services, and developer tools that bots can probe, reverse engineer, or exploit at scale.

Gaming

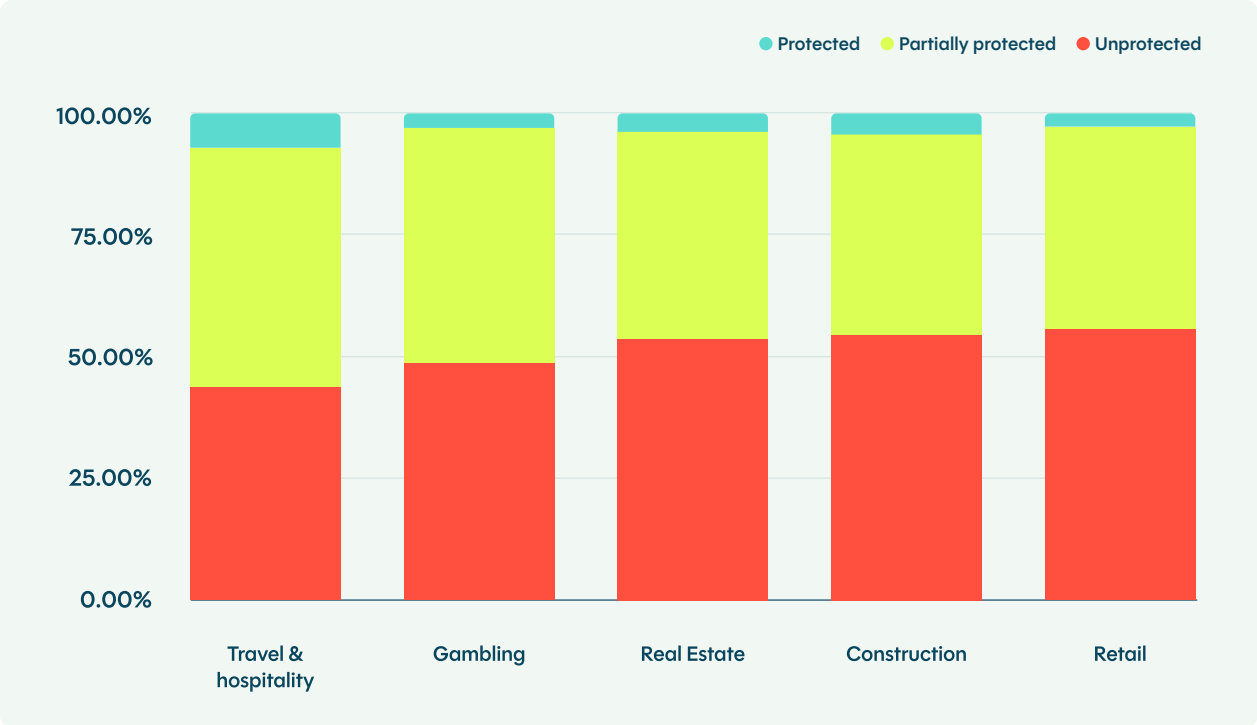
In a fully digital environment, bots undermine everything from in-game economies to real-money transactions and stored value. Account fraud is rampant when protections are weak, which can directly impact user trust, platform integrity, and revenue.

The 5 most protected industries

(by combined % of full + partial protection)

These industries show the strongest levels of protection adoption, combining full and partial implementations:

- Travel & hospitality – **56.2%** (7.1% fully protected, 49.1% partially protected)
- Gambling – **50.6%** (3.3% fully protected, 47.3% partially protected)
- Real Estate – **46.3%** (4.0% fully protected, 42.2% partially protected)
- Construction – **46.1%** (4.6% fully protected, 41.6% partially protected)
- Retail – **45.3%** (3.2% fully protected, 42.1% partially protected)



The 5 most protected industries (by % of full + partial protection)

These sectors are making meaningful progress in defending against automated threats, but gaps remain.

Travel & hospitality

This industry leads the pack, with more than half of the domains showing at least some protection. It's a promising shift for a sector long plagued by scalping, inventory hoarding, and loyalty point fraud, all of which bots execute at scale. However, with just **7.1% fully protected**, critical endpoints like booking engines remain at risk.

Gambling

A perennial target for fraud, abuse, and scraping, this industry shows that nearly half of its domains are partially protected. This is essential in an environment where real-money transactions and digital assets are core to the business model. Still, only **3.3% of domains fully block all bot traffic**, leaving room for sophisticated threats to get through.

Real Estate

With increasing digital listings and virtual showings, the real estate industry has become more exposed to scraping and fake lead generation. Encouragingly, over **46% of domains have some protection in place**, suggesting greater security awareness as the sector digitizes.

Construction

Often overlooked in discussions of cybersecurity, the construction industry showed surprising strength, ranking fourth overall. This likely reflects increased use of digital bidding systems and connected project management tools, which are ripe targets for automation if left unprotected.

Retail

As one of the most attacked industries globally, retail continues to invest in defense. But with more than half of sites still fully unprotected, the sector remains highly exposed to carding, account takeover, payment fraud, and pricing intelligence scraping, all driven by bots.

The gap between partial and full protection

Even among the top-performing industries, **full protection remains rare**, and partial protection alone isn't enough to stop today's more sophisticated bots. Partial defenses can detect basic bots, but they often fail to stop fraud, scraping, and targeted abuse.

To truly reduce risk, organizations must move beyond visibility and detection and implement real-time mitigation across all endpoints.

How has protection changed from last year?

Data from 2024 and 2025 shows a worrying trend: **most industries saw a decline in the percentage of websites fully protected** from bot attacks. Below are the standout shifts, including both negative and positive changes.

More movement toward partial protection, but lower rates of full protection. Just like last year, we see many industries catching some, but not all, bots. Retail and Marketplaces & Classifieds, for instance, show lower rates of unprotected domains, but their full-protection rates remain low.

Travel & Hospitality is now leading the pack. Last year, Gambling and Media topped the charts in preparedness. In 2025, Travel & Hospitality is among the leaders, with higher percentages of at least partial protection. In 2024, the least-protected industries were Health, Luxury, and E-commerce, with failure rates above 68%. This year, Technology & Software, Gaming, and Telecoms stand out as the least protected, with over two-thirds of domains in each sector failing every test.

Key takeaway

The level of bot protection differs significantly across industries. Some sectors—like Travel, Gambling, and Retail—are moving toward broader adoption of efficient defenses, while others, including Government, Telecoms, and Gaming, remain disproportionately exposed.

But even in the better-performing sectors, full protection is rare. The uneven distribution of defenses highlights where risk is most concentrated and where stronger protection is most urgently needed.

How AI is raising the stakes for these industries

A lack of protection against simple bot attacks already leaves gaps. But now, with AI-powered bots and agentic AI tools easily accessible, those gaps are even easier to exploit. Here's how AI is reshaping the risk profile for key industries:

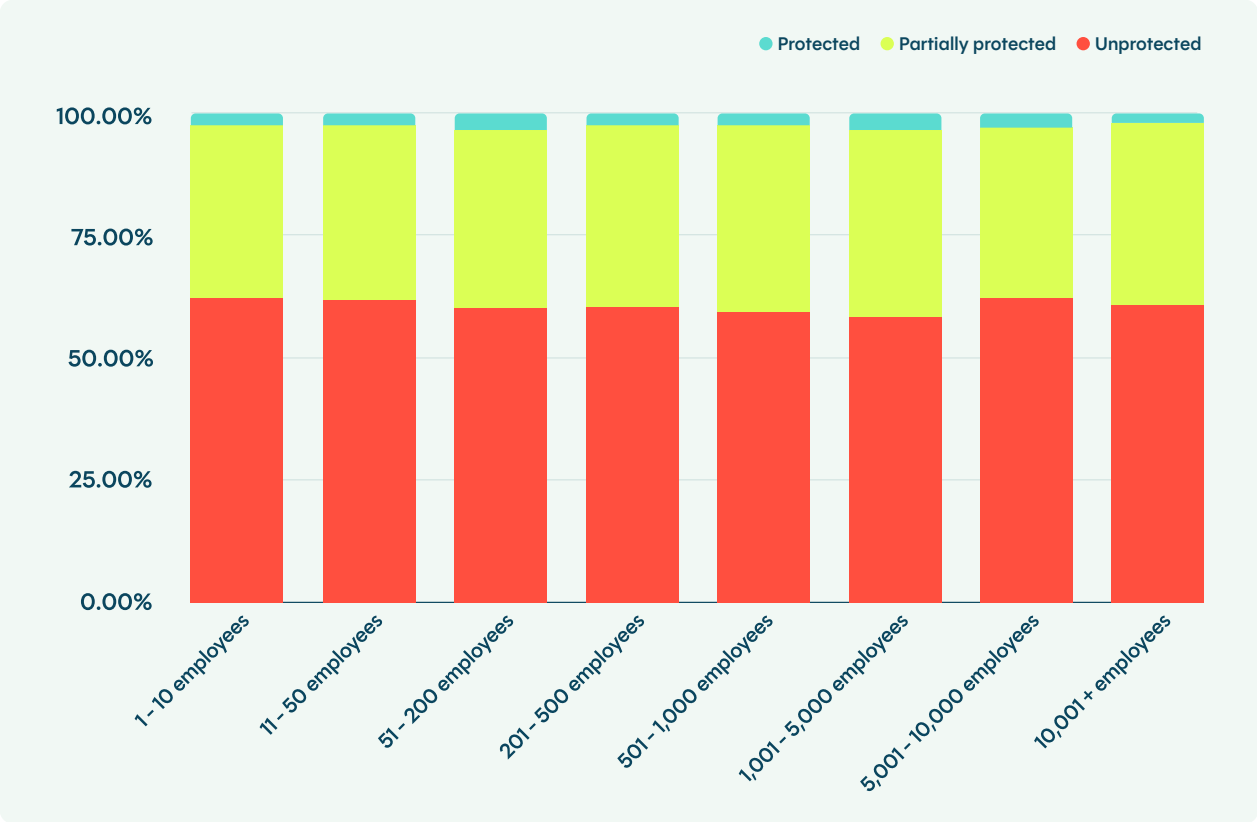
- **Retail:** Agentic commerce is gaining traction, with AI agents managing carts, checking out, and making real-time purchasing decisions. This creates new vectors for price scraping, inventory fraud, and automated undercutting.
- **Travel:** AI agents now search, compare, and book flights or hotels autonomously. This opens an entirely new attack surface, including the risk of AI agent takeover and manipulation.
- **Media & Publishing:** LLM crawlers are harvesting content at scale for training and redistribution. The result: IP theft, SEO dilution, and lost revenue.
- **Finance, Gambling & Insurance:** AI-driven fraud bots simulate human behavior, bypass CAPTCHAs, and adapt in real time, making it nearly impossible for legacy tools to detect or stop them.

If your website handles sensitive data or high-value transactions, ensuring efficient protection against automated threats is now a mission-critical task.

Bot protection by company size & monthly volume of requests

Our 2025 findings reinforce a familiar trend: larger companies tend to have stronger bot protection than their smaller counterparts, but the gap remains narrower than one might expect.

At the smaller end of the spectrum, around 62% of businesses with 50 or fewer employees failed to detect any of our test bots, with less than 3% fully protected. The very smallest companies (10 or fewer employees) did slightly better than last year, with 2.55% achieving full protection. Even among companies with fewer than 1,000 employees, the average rate of full protection was approximately 3%.



Protection by company size (by # of employees)

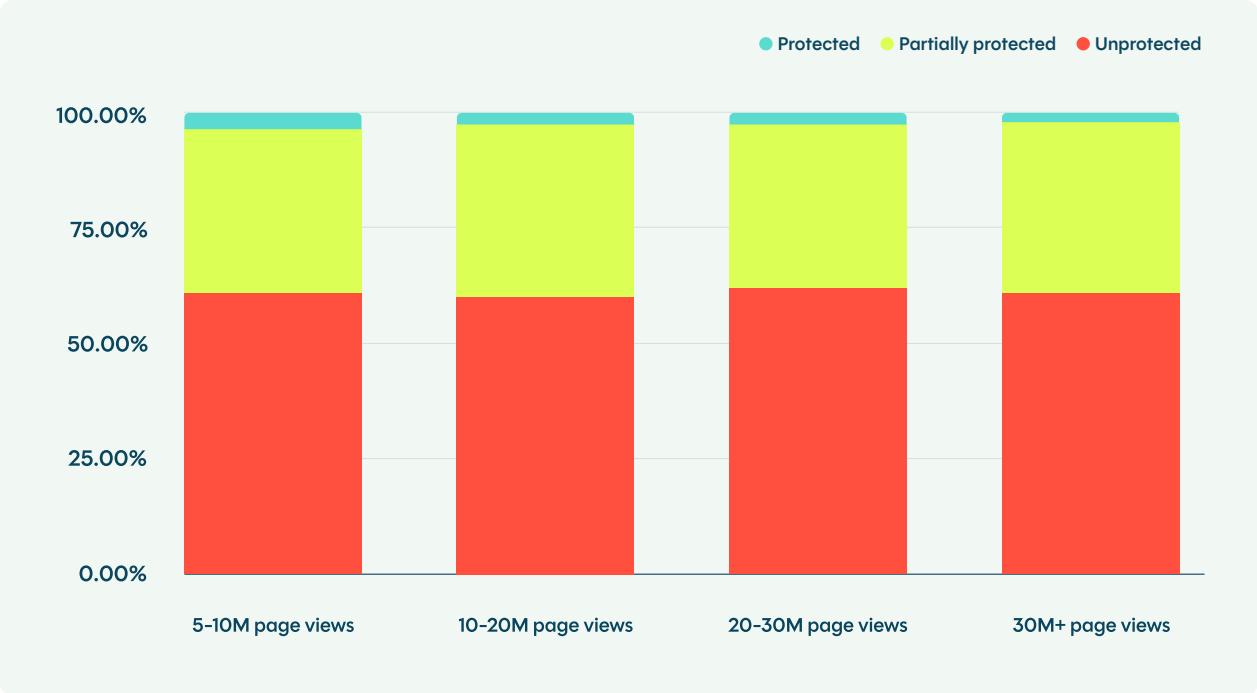
While the largest organizations fared better, the results still reflect substantial gaps in protection across companies of all sizes. Among businesses with more than 10,000 employees, nearly 61% of domains let through every bot request, and just 2.16% were fully protected. This means that even at enterprise scale, where resources and security budgets are significantly higher, a meaningful portion of companies remain fully exposed to simple bot attacks.

The data underscores a critical reality: size alone does not guarantee effective protection. Smaller organizations often face resource constraints, while larger ones may contend with the complexity of protecting vast and diverse digital infrastructures. In both cases, gaps in simple bot defenses can leave the door wide open to more advanced and costly attacks.

Year-over-year shifts in protection across company size (by # of employees)

Compared to last year's data, the protection gap between small and large companies has narrowed significantly. While larger enterprises still maintain a slight edge, the margin is smaller. In 2024, large companies were over 10x more likely to have full protection than the smallest businesses; in 2025, that gap shrank to just a single percentage point in many cases.

This shift isn't entirely new; many of the underlying factors have been developing over the past few years. One likely contributor is that large enterprises still rely heavily on legacy defenses like WAFs and CDNs. While this reliance isn't new to 2025, it continues to create a false sense of security that may be slowing their response to today's bot threats and widening protection gaps in unexpected places.



Protection by domain traffic volume (Monthly website traffic)

We also looked at domains based on their monthly traffic volumes. Our analysis confirmed that protection levels against bot attacks remain low across all categories of traffic volumes, with only small variations.

The largest domains, those with over 30 million monthly page views, actually had the lowest full protection rate in our dataset at just 2.04%, while 36.87% were partially protected and 61.09% were completely unprotected.

This is a counterintuitive and concerning finding: one might expect the highest-traffic websites, often backed by significant security budgets, to have the most mature defenses.

Instead, their scale makes them attractive targets, and their weaker-than-expected defenses leave them exposed to simple bots that can disrupt operations, steal data, scrape, or serve as an entry point for more advanced fraud.

In comparison, segments with lower traffic showed slightly higher rates of protection, but still struggled with full coverage. Domains with up to 10 million monthly page views, for example, showed 3.46% full protection and 35.23% partial protection, leaving over 61% fully exposed.

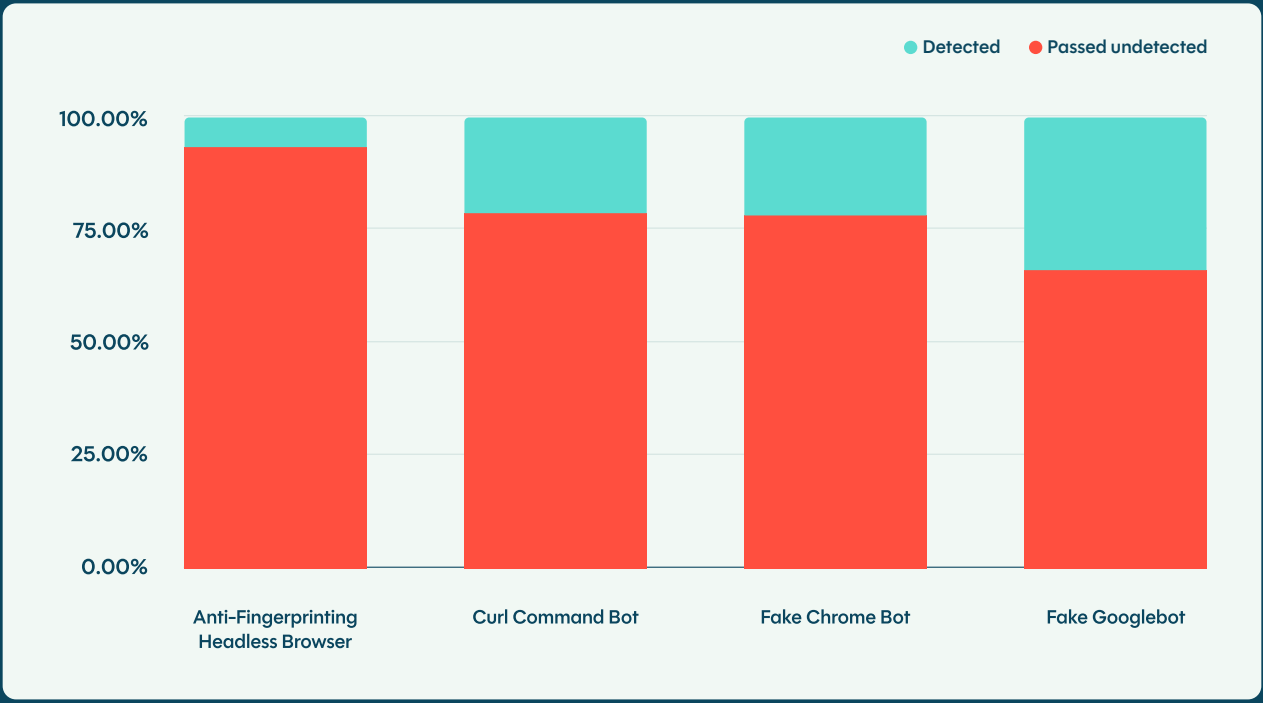
Key takeaway

Traffic volume alone is not a strong indicator of bot protection maturity in 2025. In fact, the lowest levels of protection were found among the largest, highest-traffic websites—those with over 30 million monthly visitors.

This is especially concerning: these sites attract the most attention from attackers, yet their defenses against even the simplest bots are often weaker than expected. Regardless of audience size, the vast majority of websites remain vulnerable to basic bot attacks, underscoring the urgent need for proactive protection strategies.

Success rate of bot attack vectors

All the attack vectors we tested are common and well-known, yet detection rates still vary widely. The overall patterns we observed last year persist, showing no significant year-over-year improvement in detection rates for any of the four bot tests we sent.



Protection by bot type (attack vector)

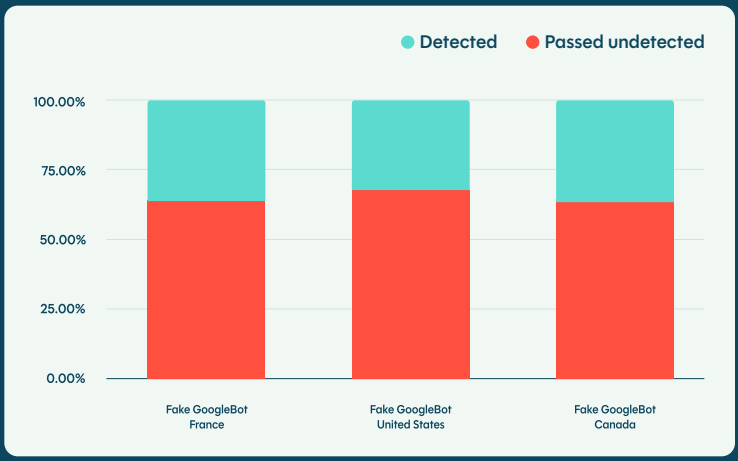
On average, the anti-fingerprinting headless browser—the only advanced bot in our study—was the hardest to stop, **evading detection in 93% of cases.**

Among basic bots, Curl command bots and Fake Chrome bots followed closely, **slipping through in roughly 79% of attempts**.

Fake Googlebots were caught slightly more often, but still **passed undetected over 66% of the time**.

Detection by bot geolocation & type

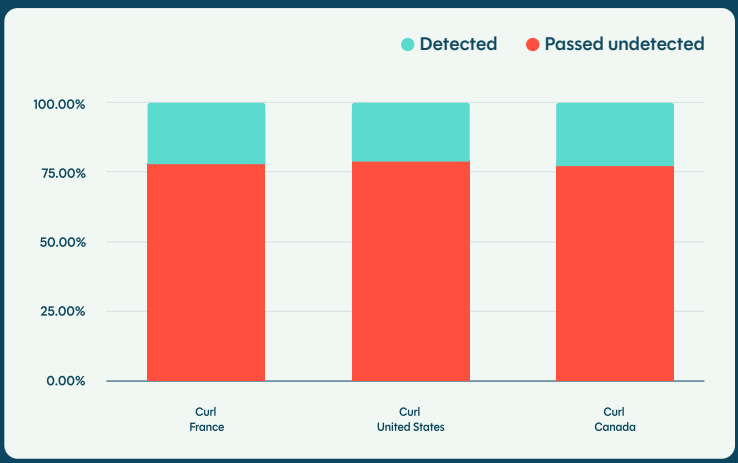
In some cases, the geographic origin of a request can impact detection results. To account for this, the Vulnerability Scan uses residential proxies to send requests from three distinct locations—France, USA, and Canada—to check whether defenses were simply blocking based on IP geography.



Detection by Bot Geolocation: Fake GoogleBot

Fake Googlebots

Fake Googlebots were detected more often than other bot types, but detection rates remain far from adequate. Nearly 33% of our fake Googlebot requests were correctly identified and blocked. That still means about two-thirds passed undetected, leaving websites open to risks like content spamming, click fraud, and vulnerability scanning.



Detection by Bot Geolocation: Curl Command Bots

Curl Command Bots

Simple Curl-like bots, which are commonly used for data scraping, product price harvesting, fake account creation, and denial of inventory attacks, evaded detection in 79% of cases. Only 21% were blocked, allowing the majority to freely extract valuable site data or disrupt inventory systems.

Detection by Bot Geolocation: Fake Chrome Bot

Fake Chrome Bots

By mimicking legitimate Chrome browser headers and routing through residential proxies, fake Chrome bots remain a dangerous threat. Detection rates were low at just 21%, with 79% undetected. Their ability to blend in with real traffic makes them well-suited for fraud campaigns such as Layer 7 DDoS and account takeover (ATO).

Advanced bots

The most sophisticated bot in this year's study, an anti-fingerprinting headless browser, proved to be the hardest to detect. Only 7% of requests were blocked, leaving 93% undetected.
These bots mimic legitimate Chrome fingerprints so effectively that most protections fail to distinguish them from real users.

This gap highlights a growing risk:

> If defenses cannot stop even a small-scale test with these advanced tools, they are likely ill-prepared for full-scale fraud campaigns powered by AI-enhanced automation.

Key takeaway

While detection rates vary by bot type and only slightly based on the origin of bot requests, the reality is that most websites still cannot reliably identify even the simplest automated threats.

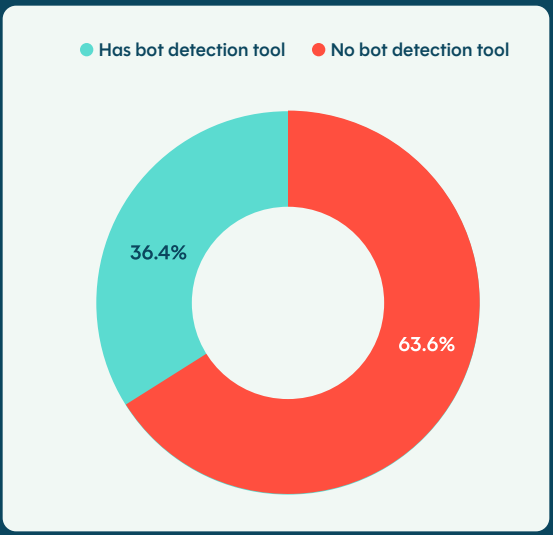
Bot protection vendors show wide variation in effectiveness

To assess the prevalence and effectiveness of bot management solutions, we analyzed whether the domains in our sample had a third-party bot management tool in place, and then measured how well each tool detected our test bots.

Of the 16,948 websites we tested, 6,177 domains (36%) had a vendor solution in place, up significantly from last year (~25% in 2024). However, this still means 10,771 websites—nearly two-thirds of the sample—had no bot protection tool at all.

We then analyzed detection performance by vendor type, using anonymized identifiers to ensure neutrality.

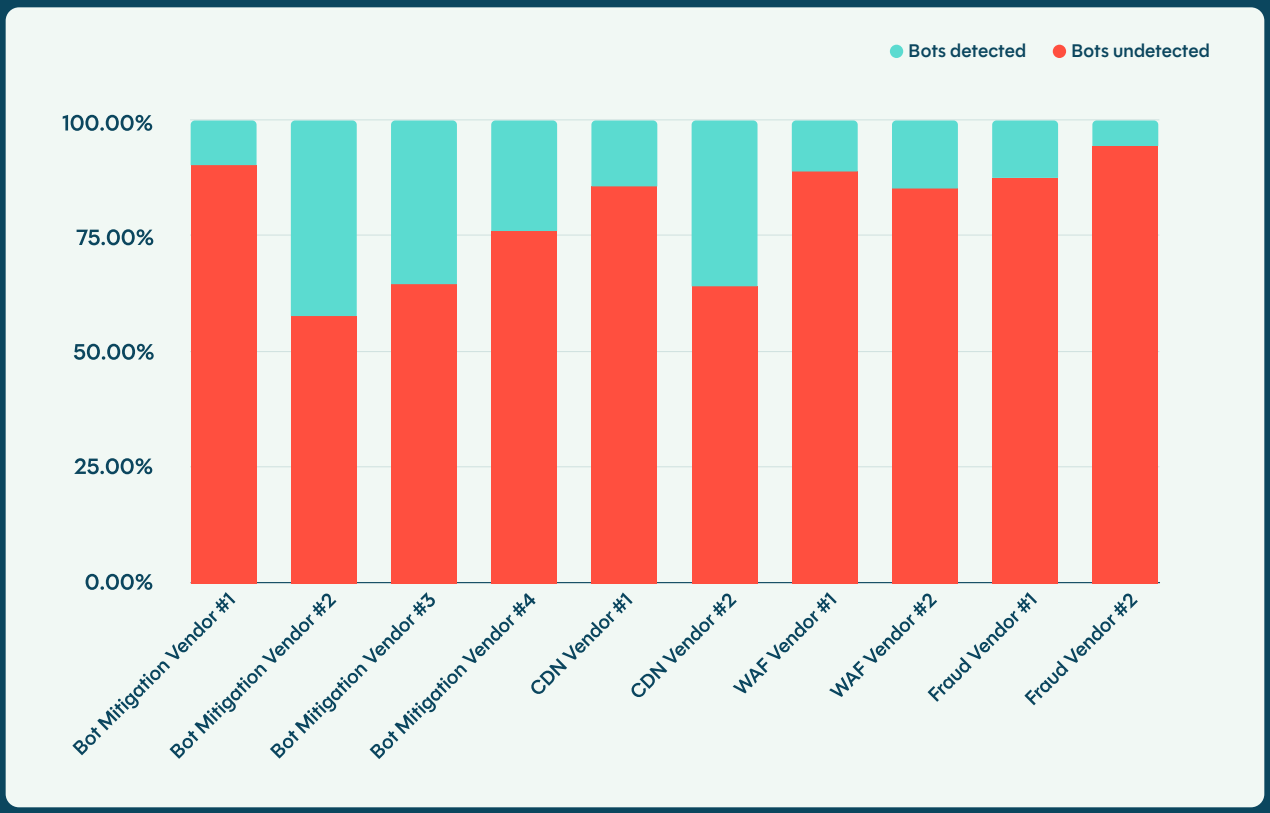
Our goal: highlight how well different categories of vendors detect and block bot traffic in real-world scenarios.



% of websites with a bot detection tool in place

Categories of vendors analyzed

- **CDNs** – Content delivery networks offering bot protection as an add-on service
- **WAFs** – Web application firewalls with built-in bot detection
- **Bot Mitigation** – Vendors focused primarily on stopping bots
- **Fraud** – Tools aimed at detecting fraud, some with limited bot coverage



How did other vendors perform?

Bot Mitigation Vendor #2 was the top performer, detecting 42% of all bot traffic—significantly higher than the average. Despite widespread deployment, CDN Vendor #1 (14% detection rate) and WAF Vendor #1 (11% detection rate) ranked near the bottom in detection efficacy. This aligns with what we've reported in the past: CDNs and WAFs often rely on outdated detection techniques or only cover surface-level signals.

Fraud-only tools performed worst in bot detection. Fraud Vendor #2 detected just 6% of bots. These tools typically analyze user behavior post-login or during a transaction, but miss upstream automation at the perimeter. CDN Vendor #2 appeared on the most domains in the sample, yet its detection rate was a modest 36%. Adoption doesn't correlate with accuracy.

The numbers speak clearly: **having bot protection is not the same as having effective bot protection.**

Most tested domains with a bot mitigation system still allowed a majority of basic bots through. This highlights the continued gap between vendor claims and real-world performance, particularly for solutions not purpose-built for bot and cyberfraud defense.

Even more concerning: these results come from basic bots and a single anti-fingerprinting headless browser (advanced bot). These are all attack vectors that modern protection systems should be capable of stopping. If many tools are unable to reliably block these simple tests, they will be even less effective against AI-powered bots and more sophisticated automated fraud attempts.

Key takeaway

The presence of a bot protection vendor is no guarantee of safety. Real resilience requires proven effectiveness, continuous testing, and adaptation to new tactics like AI-driven automation.

If your business relies on WAF, CDN, or fraud tools alone to stop bots, you are almost certainly exposed. To understand your true risk, conduct a Vulnerability Scan. Basic bots shouldn't get through. And if they do, it's time to reevaluate.

AI traffic trends & impacts on the fraud landscape

The rise of AI is reshaping the automated threat landscape in ways that businesses can't afford to ignore. In recent years, AI has been democratized, meaning both well-meaning users and malicious actors can more easily access it.

It's no longer enough for bot detection systems to ask *"Is this a bot?"* Instead, we need to ask *"What is the intent behind this traffic?"* Intent-based detection and increased controls for website users are becoming mission-critical.

Emerging trends in AI-driven traffic

AI is amplifying both the volume and complexity of bot traffic. Here's how:

LLM crawlers & Agentic AI

Large Language Model (LLM) crawlers and AI agents are now regularly scanning websites, often without consent, to gather training data. While these crawlers and agents may be helpful in some cases, they can:

- **Ingest proprietary content and customer data**
- **Bypass site policies and terms of service**
- **Introduce compliance and IP exposure risks**
- **Consume excessive bandwidth and degrade site performance**
- **Divert users from content producers to AI companies, reducing engagement and monetization**
- **Lower opportunity for cross-sell and upsells in the checkout process**

More recently, Agentic AI (AI agents capable of reasoning and taking autonomous action) are beginning to interact with sites programmatically, making decisions, navigating flows, and mimicking legitimate users. These agents blur the line between human and machine activity and introduce a new class of vulnerabilities.

AI-powered fraud bots

In addition to LLMs and AI agents, fraudsters now use AI to

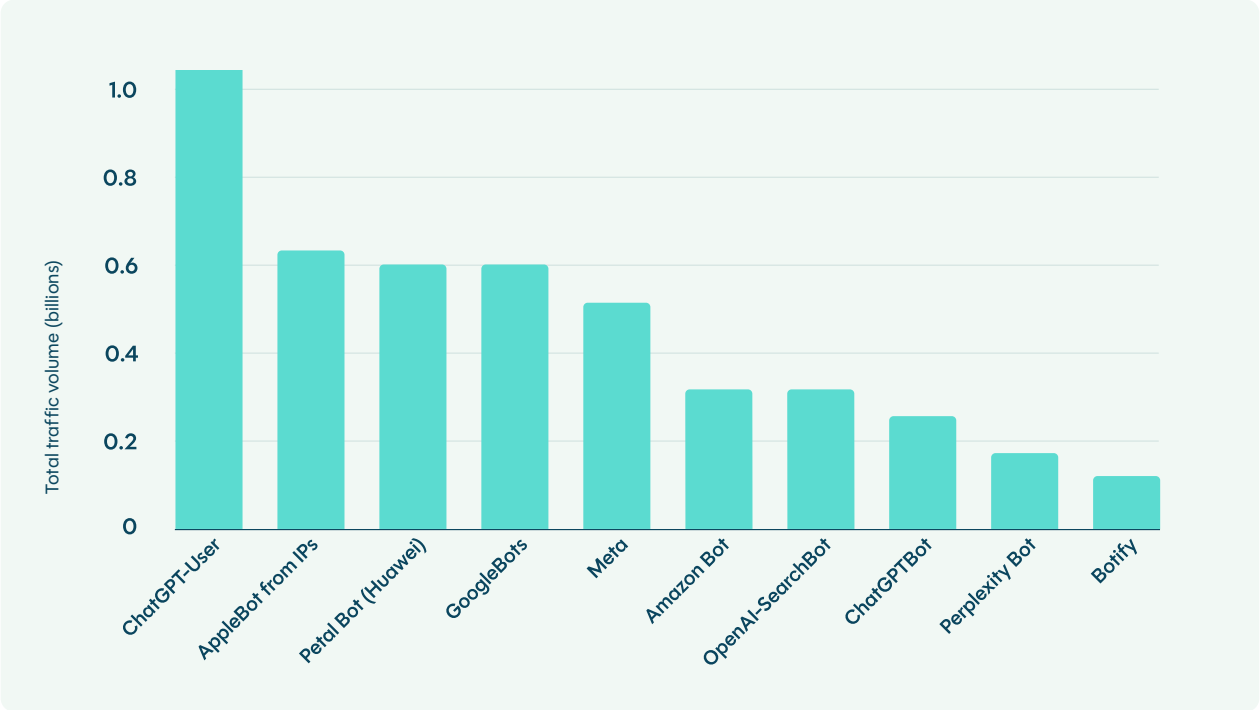
- **Automatically optimize credential stuffing campaigns**
- **Learn and adapt to defenses in real time**
- **Simulate human interaction patterns**
- **Evade detection with sophisticated fingerprint spoofing**

These bots make it easier than ever for bad actors to launch large-scale attacks, even with limited technical skill, thanks to Bots-as-a-Service platforms and low-code attack kits.

Which AI bots are really behind the traffic surge?

Looking at all DataDome customer traffic over a recent 30-day window, we observed significant activity from a variety of AI bots and agents, including LLM crawlers, AI search engines, and proprietary crawlers. The top contributor by volume was ChatGPT-User traffic, with consistently high levels across the entire period. **In August, DataDome detected nearly 1.7 billion requests from OpenAI crawlers alone**. This traffic likely represents browsing-mode sessions and AI agents acting on user prompts, navigating multiple endpoints with autonomy.

But ChatGPT is just part of the picture.



Total volume of AI traffic by type

These AI bots vary in intent and sophistication. Some are search-oriented (e.g., OpenAI-SearchBot, Petal, Perplexity), crawling to power AI-driven search results or LLM grounding. Others are user-triggered (e.g., ChatGPT-User), acting on behalf of individuals using browsing tools, plugins, or agents.

What's important is that these AI bots aren't limited to static scraping. As observed across other data in the report, many are engaging interactively—touching forms, login pages, and even transactional flows.

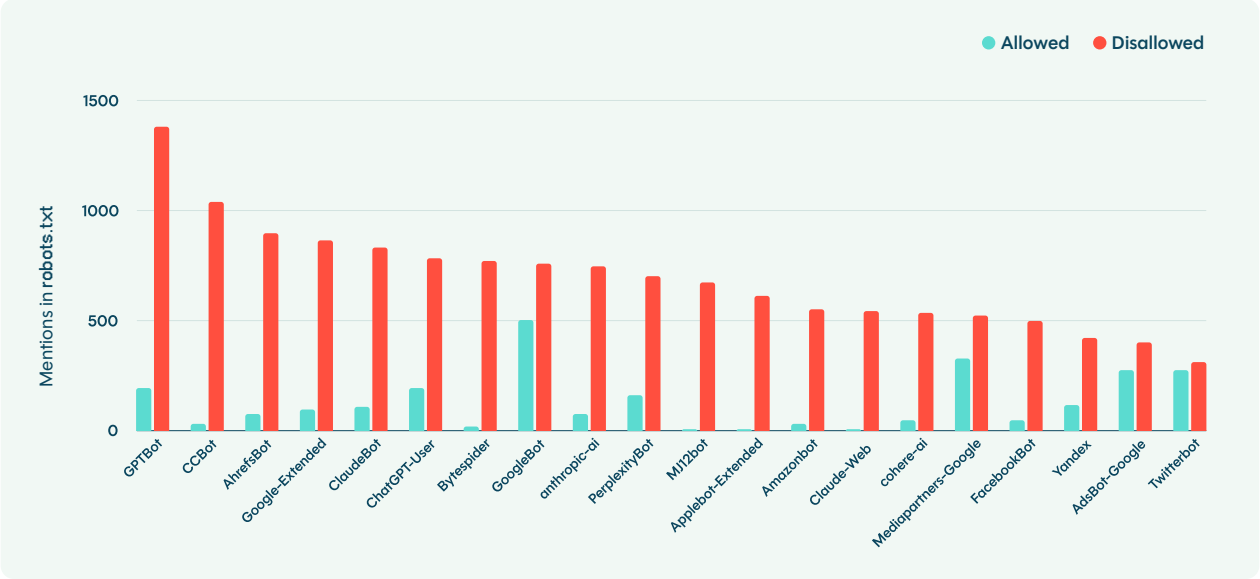
Trends in AI crawler management

A very interesting finding is that the majority of domains we analyzed explicitly disallow GPTBot in their robots.txt. This means businesses are attempting to stop AI crawlers from harvesting content, prioritizing control of digital assets.

Which AI/LLM bots are most commonly referenced in robots.txt (and how they're treated)

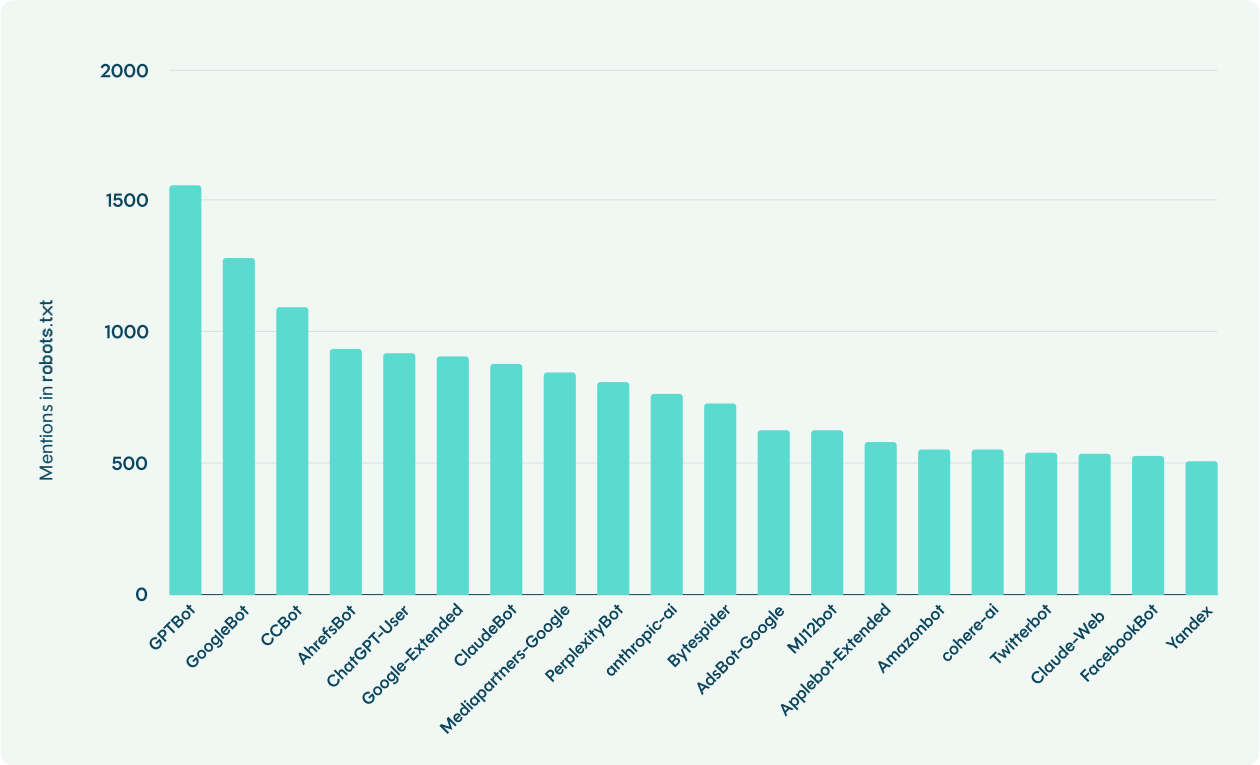
Of the websites that allowed us to access their robots.txt, we found the following:

- **GPTBot** — mentioned in 1,529 files; 11.12% allowed vs 88.88% disallowed
- **ChatGPT-User** — mentioned in 924 files; 18.72% allowed vs 81.28% disallowed
- **ClaudeBot** — mentioned in 892 files; 9.30% allowed vs 90.70% disallowed
- **PerplexityBot** — mentioned in 813 files; 14.88% allowed vs 85.12% disallowed
- **Google-Extended** — mentioned in 918 files; 8.71% allowed vs 91.29% disallowed
- **anthropic-ai** — mentioned in 784 files; 6.63% allowed vs 93.37% disallowed
- **cohere-ai** — mentioned in 572 files; 6.99% allowed vs 93.01% disallowed



Bots allowed vs. disallowed in robots.txt

Disallow directives dominate across top AI crawlers; typically, 80–90%+ of directions indicate that AI crawlers are disallowed. Notably, ChatGPT-User has the highest relative allow rate (~19%), but still skews heavily toward disallow.



Volume of bots mentioned in robots.txt

GPTBot is now the single most-referenced bot in robots.txt in our dataset, ahead of Googlebot and traditional SEO crawlers like AhrefsBot. That shift underscores how central AI crawlers have become to site policy decisions.

Important note regarding robots.txt

While robots.txt is designed to guide compliant crawlers, it's important to note that **it is not a security mechanism**. LLMs and AI crawlers—especially those used for scraping or unauthorized data harvesting—can easily ignore or bypass these directives.

As such, these configurations don't reflect protection levels, but rather a growing desire among website owners to assert control over how their content is used by AI systems.

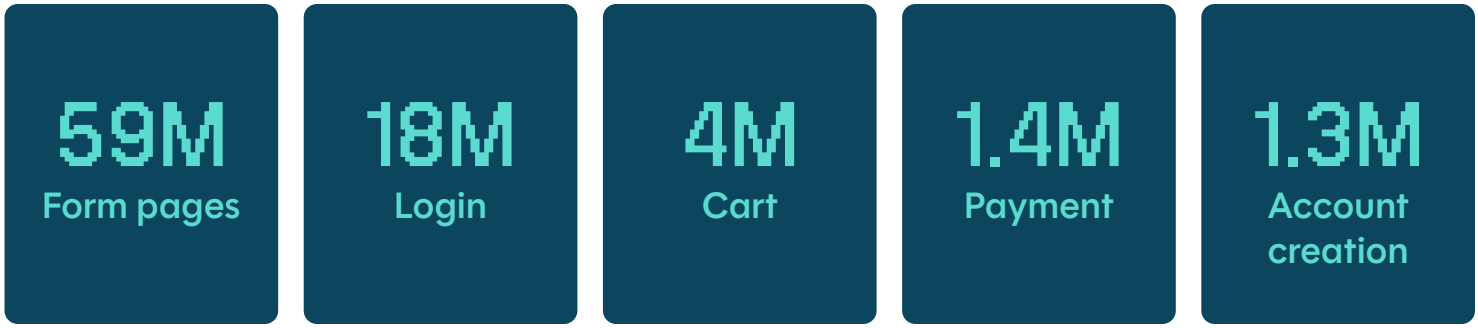
Key takeaway

AI is accelerating both sides of the bot arms race. While many businesses remain open to AI crawlers, either through inaction or strategic choice, others are already restricting access. Without active enforcement beyond robots.txt, organizations risk exposing their content, data, and infrastructure to the next generation of automated threats.

AI traffic isn't just crawling content. They're visiting high-value endpoints too.

Many assume that LLM crawlers and AI agents are only interested in content. But the data shows otherwise. These autonomous tools are interacting across forms, logins, carts, payments, and account creation flows, not just your blog or homepage.

Over a recent 30-day period, the highest share of AI traffic on DataDome customer sites was observed on the following endpoints: **forms (64%)**, followed by **login (23%)**, **cart (5%)**, **payment (2%)**, and **account creation (2%)**. This data excludes traffic to pages without these endpoints, such as home pages and purely informational pages.



Why does this matter?

- Forms are being used for lead fraud, field discovery, and input testing.
- Logins and accounts are prime targets for AI-powered credential stuffing and fake account creation attacks.
- Carts and payment interfaces are seeing simulated transactions, scalping, and card testing.

This reinforces the need for **real-time, intent-based detection across all endpoints.** Agentic AI isn't coming for just one part of your site; it's exploring it all.

Key takeaway

AI bots aren't just indexing your content, they're interacting with your business. Whether for training, search, or goal-oriented automation, these bots are becoming increasingly dynamic and difficult to distinguish from humans.

You need visibility not just into *who* is visiting, but how they're behaving and especially *why*. That requires real-time AI detection that can detect intent, not just static user-agent filtering.

Why intent matters more than ever

As AI continues to evolve, the old binary model of bot vs. not-bot is breaking down. Legitimate use cases (like SEO, uptime monitoring, or a customer making a legitimate purchase via an agent) can appear automated, just like malicious ones. To protect digital ecosystems without overblocking, security systems must assess:

- **Behavioral context:** What actions is the visitor taking?
- **Traffic patterns:** Is the request part of a known fraud scheme?
- **Endpoint sensitivity:** What part of the site is being accessed and how?

At DataDome, we're constantly evolving our AI Detection Engine to adapt to this new landscape, using real-time multi-layered AI to analyze not just identity, but *intent*.

We want our customers to be protected against modern, advanced fraud, while still being able to adapt their business to the opportunities created by emerging automation and AI traffic.

Takeaways & recommandations

From our tests of nearly 17,000 major domains around the world, widespread vulnerability to simple bot attacks remains the rule, not the exception.

While partial protection has grown slightly year-over-year, fully protected domains have dropped sharply.

At the same time, malicious traffic is becoming harder to distinguish from human behavior. Even legitimate users are starting to rely on AI agents to browse, compare, and transact. Staying protected today requires more than just blocking obvious bots.

Weak bot defenses invite high-ROI fraud

With the majority of domains failing to detect even the simplest attack vectors like fake Googlebot or Curl requests, they become prime entry points for more sophisticated attacks:

- If basic bot traffic is undetected, residential-proxy-based ATO campaigns are even less likely to be caught.
- If rudimentary scraping passes, AI-enhanced scraping bots will do so undeterred.
- If simple requests bypass fraud prevention, distributed carding operations can slip through using the same blind spots.

Vulnerabilities open the door to escalation. Attackers go where they face the least resistance. **Weak "bot hygiene" becomes a signal to fraudsters that your stack is ripe for more aggressive, high-ROI exploitation.**

Advanced attacks don't always start off advanced. They escalate.

And if your detection stack is failing at the basics, **you're not just missing small threats, you're inviting bigger ones.**

Key recommendations

1. Intent-based detection is no longer optional

As AI-powered bots get better at mimicking humans and skirting defense systems, security systems must move beyond binary logic. Instead of asking "Is this a bot?", teams must ask "What is this visitor trying to do?" Detection based on behavior, context, and endpoint sensitivity is critical to separating good automation from fraud.

2. Protection must extend to your most targeted endpoints

AI bots aren't just crawling content; they're targeting logins, carts, forms, and payment flows. These high-value areas need layered protection that detects threats before they convert into fraud.

3. Bot protection must be validated, not assumed

Many businesses are relying on vendors that fail to stop even basic bot attacks. Regular testing is essential to ensure your tools work as expected under real-world conditions, not just in sales demos.

Recognized as a Leader in the 2024 Forrester Wave™ for Bot Management

Trusted by over 300 leading brands worldwide

Etsy

The New York Times

SOUNDCLOUD

ASUS

Tripadvisor

petco

PayPal

Why DataDome?

DataDome is the leading Cyberfraud Protection Platform, built to stop today's most advanced automated threats in real time.

Trusted by leading global brands, DataDome protects web applications, mobile apps, and APIs against bots, account fraud, scraping, DDoS, and AI-powered abuse without compromising performance or user experience.

DataDome is fully automated and integrates seamlessly into any tech stack. Backed by a 24/7 SOC team of advanced threat researchers, DataDome stops over 400 billion attacks annually. Experience protection that outperforms with DataDome.

Real-time AI, built for scale

We analyze 100% of requests in under 2 milliseconds, learning from 5 trillion+ signals per day to detect and block fraud with unmatched speed and accuracy.

Protection that outperforms

Our industry-leading <0.01% false positive rate ensures legitimate customers aren't blocked, even as attacks evolve.

Full-journey coverage

From signup to login to checkout, DataDome protects every step of the user journey, across web, mobile, and APIs.

Intent-based detection

In an era of AI agents, DataDome goes beyond "bot vs. human" to assess behavioral intent in real time, distinguishing fraud from legitimate activity.

Additional resources

For readers looking to go deeper into bot detection, cyberfraud prevention, and the evolving role of AI in online threats, we recommend the following resources from DataDome:

- Vulnerability Scan Check your website with our self-service Vulnerability Scan to quickly discover if you're vulnerable to simple bot attacks and cyberfraud.
- The Forrester Wave™: Bot Management Software, Q3 2024 An essential resource for decision-makers which scores leading bot management solutions on detection accuracy, integration capabilities, customer service, and more.
- A CISO's Guide to Cyberfraud Protection A comprehensive framework for understanding and addressing modern fraud tactics across the digital attack lifecycle.
- Bots Unmasked: The Book A paperback book all about exposing cyberfraud in the era of AI, including attacker techniques, case studies, and defense strategies.
- Bot Management & Online Fraud Prevention Buyer's Guide A free toolkit including key questions for potential bot protection vendors, a step-by-step evaluation checklist, and a PowerPoint template for internal selling to key stakeholders.
- Public AI Crawlers Threat Intelligence Database An up-to-date resource on AI and LLM crawlers, including which bots are most frequently allowed, disallowed, and emerging in robots.txt files worldwide.
- Watch a Demo See DataDome in action and learn how our real-time detection engine protects against billions of automated threats every day.
- ROI Calculator Estimate the financial impact of bot and fraud protection on your business and quantify the return you can expect from investing in DataDome.

DataDome

datadome.co