DataDome

ANNUAL REPORT

State of Bot & Agent Security Report *2026*

Table of contents

About this report

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

The annual State of Bot & Agent Security Report — formerly named the Global Bot Security Report — is one way we advance that mission. The new name reflects a shift in the threat landscape. Bots and AI traffic now represent two distinct and growing challenges that organizations need to understand together.

We aim to give organizations a clear, data-backed view of where bot and AI threats stand today, how the landscape is evolving, and where the most critical gaps in protection remain.

For the 2026 edition, we combined multiple research methods, two primary datasets, and a focused analysis of the AI traffic within our customer-traffic dataset to build the most complete picture we've published to date. We tracked how real attacks are changing in volume, type, and technique. We examined the rapid rise of AI agents and large language model (LLM) crawlers, including who's generating the traffic, where it's landing, and what it means for security teams. And we looked at how organizations across industries and regions are prepared to defend against automated threats.

Together, this report provides a holistic view of bot and agent traffic readiness across the internet and a broader view of how the landscape is shifting and what that demands of organizations in 2026 and beyond.

Executive *summary*

As we analyzed this report's data, one thing became clear: automated traffic is changing faster than ever. At DataDome, we're seeing not only more bots and AI agents, but also more sophisticated activity and more attempts to reach the high-value parts of the customer journey. **Bad bot traffic grew approximately 124%** between July 2025 and June 2026 — over **9x faster than human traffic growth,** and AI traffic grew by 82.3% over the same time period. This means that traffic is becoming harder to evaluate using identity-based controls alone.

This report combines an analysis of over a trillion requests across 75,000+ customer sites, as well as a test of more than 20,000 popular websites. Across these datasets, the findings point to a clear shift: **automated traffic is becoming more valuable and more active, while most businesses are not adapting their defenses quickly enough.**

The clearest signal is the continued growth of scraping. **Scraping accounted for 70.9% of bad bot traffic across DataDome's customer base, the largest of any attack vector, and increased 185.2%** during the 12-month study period. One possible explanation is the growth of an AI data supply chain. Third-party data resellers and applications building AI agents may be collecting web data at scale to support model training and other AI services. This activity does not always identify itself as an AI crawler. It can operate through conventional scraping infrastructure, which means the impact of AI may be larger than the traffic labeled as an AI bot alone.

AI is reshaping the traffic landscape on both sides of the security equation. DataDome detected **52.7 billion AI agent and LLM crawler requests** across its customer base over the study period, with a **total AI traffic increase of 82.3%.** Among identified AI crawlers, Meta-affiliated bots generated the largest share at 46.3%, followed by OpenAI-affiliated bots at 34.6%. These AI agents are also reaching high-risk endpoints: **within H1 2026, login page targeting by AI bots grew more than 8x.**

The gap between the threats businesses face and the protection they have in place is widening. An analysis of 20,000+ websites found that **65.3% were completely unprotected against the 10 bot types in our test. Only 2.4% achieved full protection.** Full protection has now declined for the second consecutive year, from 8.4% in 2024 to 2.8% in 2025 and 2.4% in 2026. The decline is happening as attackers gain access to more capable tools, including anti-fingerprinting frameworks, automated browsers, and low-cost bot-as-a-service platforms.

What we are seeing across these datasets reinforces the need for intent-based detection. Businesses that hard-block all AI traffic will lose revenue, and businesses that allow everything through will face fraud and abuse.

An AI bot that hits a login page repeatedly to test stolen credentials is not the same as one that reads your website in order to provide the end-user with information about your product. The question your detection engine should be asking in 2026 is no longer "Is this a bot or a human?" It is: "What is this visitor trying to do, and is it beneficial to the business?"

Key findings

70.9%

Scraping is the fastest-growing attack

Scraping is the dominant and fastest-growing attack vector. Scraping accounted for 70.9% of bad bot traffic and increased 185.2% during the study period. One possible explanation is the growing AI data supply chain, in which third-party data resellers and applications building AI agents collect web data at scale for training, agent responses, and other AI services.

124%

Malicious automated traffic is growing much faster than human traffic

Bad bot traffic increased 124% over 12 months, over 9x the growth rate of human traffic over the same period.

290.7%

Scalping activity increased

with median daily volume nearly quadrupling. The sustained rise puts continued pressure on businesses to protect high-demand inventory and purchase flows from automated

5.5%

The protection gap is visible even against basic automation

Only 5.5% of tested websites stopped every basic bot, and just 3.1% stopped every real-browser bot. These results show that the challenge is not limited to advanced attackers. Many websites remain exposed to the simplest forms of automation.

>8x

Login-page targeting by AI agents increased more than 8x within H1 2026

Monthly AI bot requests to login pages rose from 11.9 million in January 2026 to 99.7 million in June 2026 across DataDome customer sites. Some of this activity may come from legitimate agents acting on behalf of users, but the same endpoint behavior can also resemble credential testing, account takeover reconnaissance, and form abuse.

34.5%

Account-related abuse is increasing across multiple attack paths

Fake account creation grew 34.5%, while credential stuffing remained cyclical rather than declining. Credential stuffing activity surged, dropped by nearly 90%, and later recovered to new single-day highs, showing how attackers can pause campaigns while refreshing credential lists or rotating infrastructure.

49%

Attackers continue to succeed with both simple and sophisticated tools

Simple bots accounted for 49% of sampled bad bot traffic, even though they made no effort to hide their automated nature. At the same time, 21.5% of sampled bad bot traffic representing "payload forgers" was caught using more advanced signals from WebAssembly (WASM) and our Virtual Machine.

52.7 billion

AI traffic is moving from passive crawling into active user journeys

DataDome detected 52.7 billion AI agent and LLM crawler requests across its customer base over 12 months, with total AI traffic increasing 82.3% during the study period. In H1 2026, AI agents generated 605.6 million requests to high-risk endpoints, including login pages, forms, carts, payment flows, and account-creation pages, with login pages accounting for 51.7% of that traffic.

65.3%

20,000+ popular websites are completely unprotected against bots and AI agents

In the expanded 2026 test, 65.3% of tested websites stopped none of the 10 bot types, while only 2.4% stopped all of them. The directional trend in full protection has declined from 8.4% in 2024 to 2.8% in 2025 and 2.4% in 2026, although the 2026 test suite included additional spoofed AI agent and advanced bot types.

45%

AI agent spoofing increased 45% between February and July 2026

User-prompted agents showed the sharpest category-level increase, rising from approximately 1.4% to 5.6%. This makes live AI assistants an increasingly important type of traffic to verify.

Methodology

Customer total traffic analysis

Our primary dataset includes anonymized, de-identified traffic data across DataDome's global customer base. For the 2026 report, we analyzed **trillions of requests across 75,000+ customer sites** between July 2025 and June 2026.

The data covers total traffic composition, attack vector volumes and trends, bot behavior patterns, and attack technique evolution. No end-user data is collected or stored. All figures represent aggregated, anonymized traffic signals.

HOW TO INTERPRET ATTACK VECTORS DATA

Throughout this report, an attack vector refers to an automated request or sequence of requests that our systems classify as an attempt to carry out a specific activity, such as scraping, credential stuffing, distributed denial-of-service (DDoS), or fake account creation. It does not indicate that the activity was completed successfully or that a business impact occurred.

Attack vector classifications are based on signals observed in the request and session, including the endpoint targeted, request velocity and volume, behavioral patterns, browser and device fingerprints, network characteristics, claimed identity, and known attack signatures. A request may therefore be classified as an attempted attack even when DataDome blocks it before the action is completed. The figures in this section represent detected attempts, not confirmed successful attacks.

Customer AI traffic analysis

For our AI traffic analysis, we filtered our first dataset, looking specifically at AI traffic across DataDome's global customer base. It classifies AI agents, agentic browsers, and LLM crawler traffic by bot type, endpoint, and referral source. The AI traffic covers **52.7 billion requests** across DataDome's 75,000+ customer websites.

The AI endpoint and AI referrer analyses both cover H1 2026 (January–June 2026).

Test of 20k+ popular websites

For the third portion of the report, we created a dataset by testing **20,000+ of the world's most popular websites** across 15 industries and several major geographic regions (excluding those without a large enough sample size). Testing took place during June 2026. Domains were excluded if they were inaccessible during the testing window. The final valid sample was **21,491 domains**, up from 16,948 in 2025.

The study was conducted by testing websites' homepages against multiple bot and AI traffic types by simulating real attack scenarios using residential proxies in three geographic locations: the United States, Canada, and France. Requests are routed through residential internet protocol (IP) addresses to avoid detection based purely on IP reputation.

THE TEST SENDS REQUESTS FROM 10 BOT TYPES ACROSS FOUR TIERS OF DIFFICULTY:

- **TIER 1: BASIC BOTS** (2 TEST TYPES IN TIER)
  These bots do not attempt to disguise themselves. They send raw, unadorned requests with minimal or generic headers. They are the lowest-effort attack type and the easiest to detect.

- **TIER 2: SPOOFED AI AGENTS AND CRAWLERS** (4 TEST TYPES IN TIER)
  These bots forge the Hypertext Transfer Protocol (HTTP) headers of known, trusted identities, including Google's crawler and AI agents from OpenAI and Anthropic. This includes spoofed versions of ChatGPT, GPTBot, and ClaudeBot.

- **TIER 3: DISGUISED BOTS** (1 TEST TYPE IN TIER)
  The bot used in Tier 3 goes beyond headers and forges the network-level fingerprint of a real browser — including Transport Layer Security (TLS) handshake patterns and HTTP/2 behavior — without actually running one. Catching them requires looking past the visitor's claimed identity and analyzing the connection itself against known browser fingerprint profiles.

- **TIER 4: REAL BROWSERS** (3 TEST TYPES IN TIER)
  These bots run a genuine browser engine, such as Chrome or Firefox, and are fully automated and equipped with anti-detection techniques designed to mimic human behavior. These are the most sophisticated bots and the hardest to detect.

Scoring framework: Each domain receives a score based on how many bot types were challenged or blocked. A site is scored **fully protected** if it challenged or blocked all tested bot types, **partially protected** if it challenged or blocked some — but not all — test requests, and **unprotected** if it let all 10 test bot requests through without a challenge.

Fully protected: Successfully blocked or challenged 100% of our 10 bot types.

Partially protected: Successfully blocked or challenged a portion of our test bots.

Unprotected: Failed to block or challenge any test bot.

A note on the ethical use of bots

Our test is designed to have zero impact on sites. Tests are rate-limited, non-destructive, and designed to simulate realistic bot and AI behavior, not to stress or disrupt infrastructure. No end-user data is accessed or retained. The test does not exploit vulnerabilities; it only tests whether automated traffic triggers a detection response.

Part 1: Bot traffic & threat trends in 2026

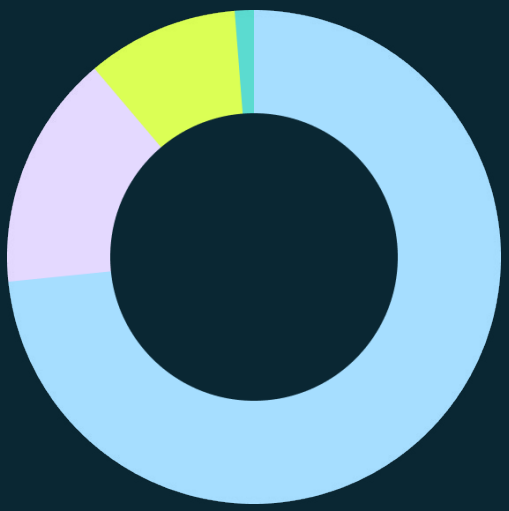
1.0 Key findings

- **Bad bot traffic grew 124% in 12 months,** over 9x the growth rate of human traffic.
- **AI traffic grew by 82.3%,** over 6x the rate of human traffic growth.
- **Scraping is the dominant attack vector at 70.9% of bad bot traffic,** and grew 185.2% year-over-year.
- **Scalping activity increased 290.7%,** with median daily volume nearly quadrupling.
- **Credential stuffing is cyclical.** Net volume was flat year-over-year, but attacks surged through summer 2025, collapsed by nearly 90%, then recovered to new single-day highs by April 2026.
- **DDoS attacks grew 39.9%,** reaching a peak of more than 2 billion requests in a single day in April 2026.

1.1 How much web traffic is automated?

DataDome analyzed **trillions of requests across 75,000+ customer sites** over the 12-month study period. Over 1 in 4 web requests were automated. Bots and AI agents collectively generated **~26.5% of all traffic in this dataset.***

* Approximately 60% of DataDome customers operate behind a content delivery network (CDN) with basic bot filtering enabled. Because DataDome is deployed downstream of the CDN layer, it analyzes only the traffic that passes through, not the full origin request volume. The automated traffic figures reported in this study therefore represent a conservative lower bound. Real-world bot traffic rates across the internet are likely higher than what this data reflects.

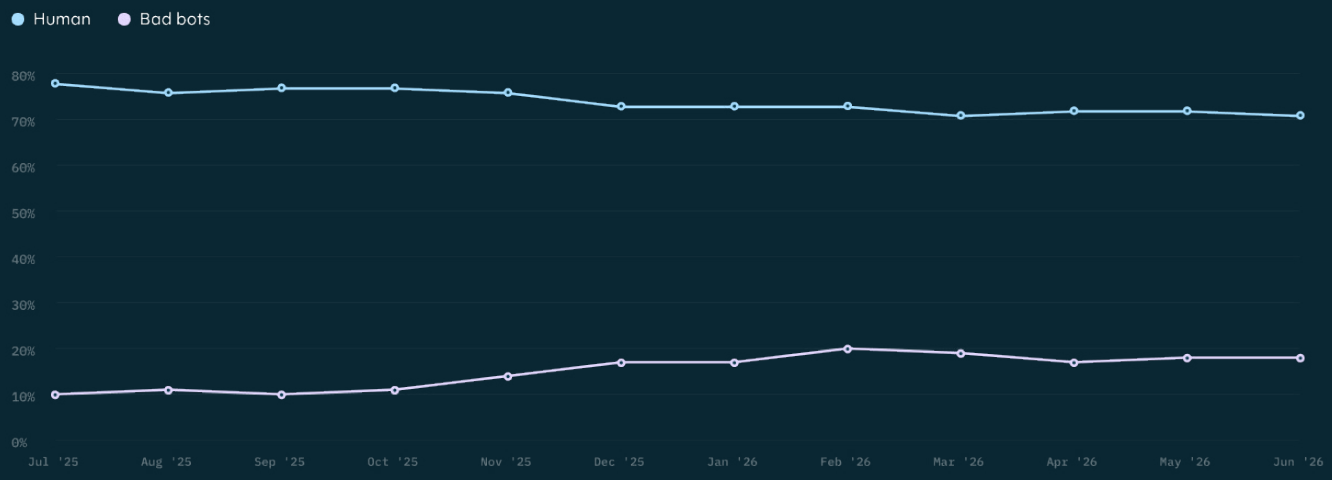


Breakdown of DataDome total traffic analyzed by type

- Human 73.41%
- Bad bots 15.42%
- Good bots 9.93%
- AI agents 1.25%

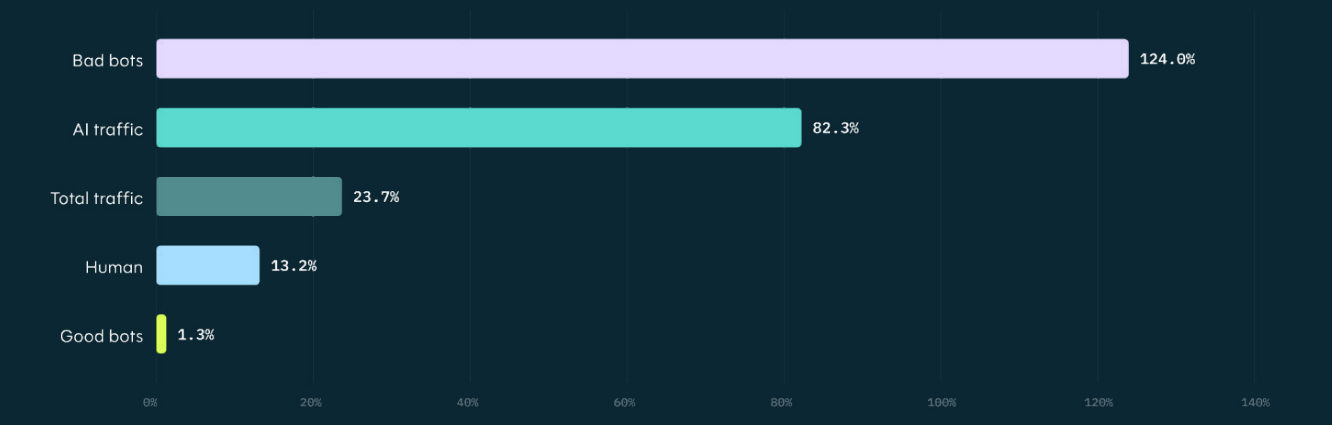
Human traffic's share of total requests declined from 77.9% in July 2025 to a low of 70.7% in February 2026, before partially recovering to settle around 71.3–71.8% through the first half of 2026. That share was absorbed almost entirely by bad bots, which climbed from 9.9% in July 2025 to a peak of 20.3% in February 2026 before settling at approximately 17.6% through the second quarter of 2026.

Percentage of total traffic by human and bad bots, July 2025 to June 2026



Between July 2025 and June 2026, total traffic across all DataDome customers grew by 23.7%, while human traffic grew by 13.2%. Both figures were significantly outpaced by automated categories. **Bad bot traffic grew by 124.0%, more than 9x the rate of human traffic growth. AI traffic grew by 82.3%, over 6x the rate of human traffic growth.**

Percentage growth in traffic type over a 12-month period



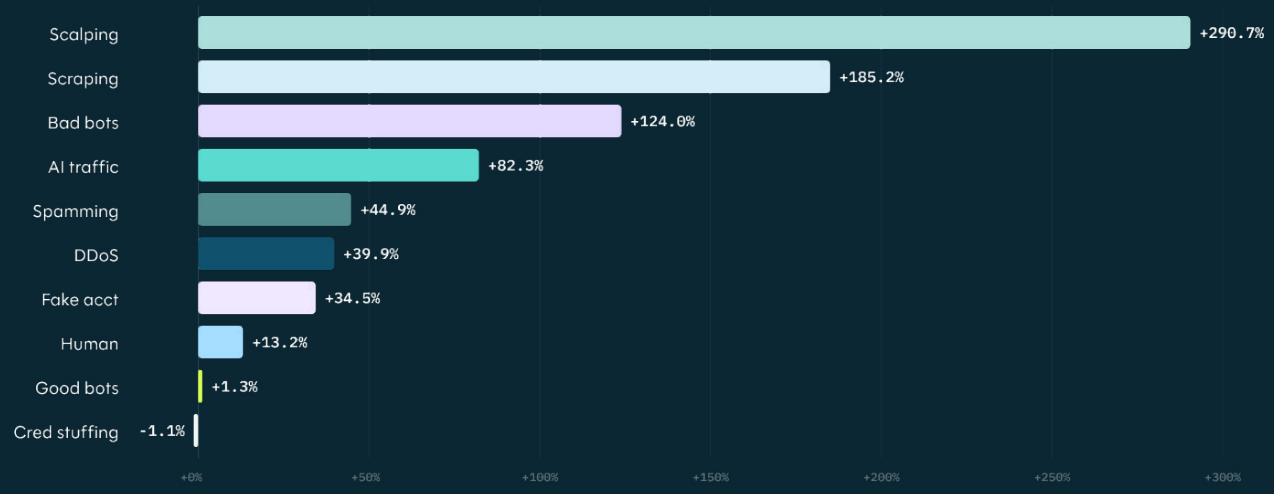
1.2 Which bot attack vectors are growing fastest?

Bad bot traffic grew by 124% year-over-year, but that aggregate figure masks significant divergence across attack types. Scalping, scraping, spamming, and DDoS expanded significantly. Credential stuffing held flat on an annual basis while cycling through many drops and then resurgences.

The table below shows total change for major threat types across DataDome's customer base, between July 2025 and June 2026.

Note: Attack vector data measures detected attempts, not successful attacks. Classifications are based on the signals and behavior observed in each request or session.

Change in traffic by attack type, July 2025 to June 2026



Web scraping is the dominant attack type by far, accounting for 70.9% of all bad bot traffic observed. The volume of scraping attempts saw a 185.2% increase over 12 months. This growth reflects both the proliferation of automated scraping tools and the rising demand for data harvested at scale, increasingly to feed AI model training pipelines.

DDoS attacks grew 39.9% and peaked at 2 billion requests in a single day in April 2026. Unlike scraping, which surged and partially corrected, DDoS showed a more consistent upward trajectory, suggesting structural growth rather than campaign-driven spikes. DDoS traffic made up 12.7% of all bad bot traffic that DataDome saw over the 12-month study period, the second largest observed attack type.

Spamming saw a 44.9% increase over the 12-month study period, but the overall series is better characterized as a sequence of high-volume waves than as a steady increase. Spamming made up 7.9% of all bad bot traffic across DataDome's network.

Scalping saw a 290.7% increase. Median daily volume nearly quadrupled across the 12-month period. The bulk of scalping was not limited to a single event. Scalping activity became more consistently elevated during the second half of the study period, while still producing periodic surges.

Total bad bot traffic across DataDome customers by attack type

- Scraping 70.9%
- DDoS 12.7%
- Spamming 7.9%
- Cred stuffing 6.7%
- Fake account creation 0.6%
- Payment fraud 0.5%
- Vulnerability scanning 0.5%
- Scalping 0.2%

Fake account creation grew 34.5% over the 12-month period, adding to the pressure on businesses to protect registration flows as well as login pages. Attackers commonly use browser automation frameworks such as Playwright and Chrome DevTools Protocol for fake account creation campaigns, making this activity harder to distinguish from legitimate users.

The cost of account fraud: *a multibillion-dollar problem*

Account abuse can create costs for both consumers and businesses through unauthorized transactions, chargebacks, refunds, customer support, remediation, and lost trust. Consumers reported more than $12.5 billion in fraud losses in 2024, according to the Federal Trade Commission. During the same year, the FBI's Internet Crime Complaint Center recorded $16.6 billion in reported internet-crime losses. Together, these figures show the scale of the financial problem surrounding online fraud and cybercrime, even though neither figure is limited to account abuse alone.

Sources: Federal Trade Commission, Consumer Sentinel Network Data Book 2024; Federal Bureau of Investigation, 2024 Internet Crime Report

Credential stuffing is cyclical, not declining. Net growth was essentially flat, but that figure obscures a pattern of extreme spikes followed by sharp collapses. Volume surged through summer 2025, then dropped by nearly 90% over the following three months before fully recovering by early 2026 and reaching new single-day highs in April. This boom-bust cycle repeats: attackers run high-intensity campaigns, go dark, likely to refresh credential lists and rotate infrastructure, then return at full scale. Credential stuffing attempts made up 6.7% of all bad bot traffic that DataDome detected.

1.3 What is the sophistication level of bot attacks?

Most bot defenses are built to stop the obvious: simple automated scripts that therefore do not attempt to hide their identity. The question is how much of today's attack traffic is actually obvious and how much has moved beyond simple detection.

We classified a sample of bad bot traffic by sophistication level, based on the detection signals each request triggered. The analysis covers over 1.5 billion bad bot requests over 30 days in 2026. Because this classification methodology is new this year, there is no prior-year comparison.

Bot traffic sophistication level over 30 days

- Simple bots 49%
- Average bots 29.5%
- Advanced bots 21.5%

Simple bots still dominate. The largest share of bot traffic, 49%, is generated by bots that do not attempt to hide their automated nature. These bots trigger WebDriver signals, a fingerprint exposed by Selenium and basic ChromeDriver automation. They are among the easiest types to detect. Yet as Part 3 of this report shows, simple bots evade detection on an average of 88.6% of websites. Attackers continue using them because they continue to work.

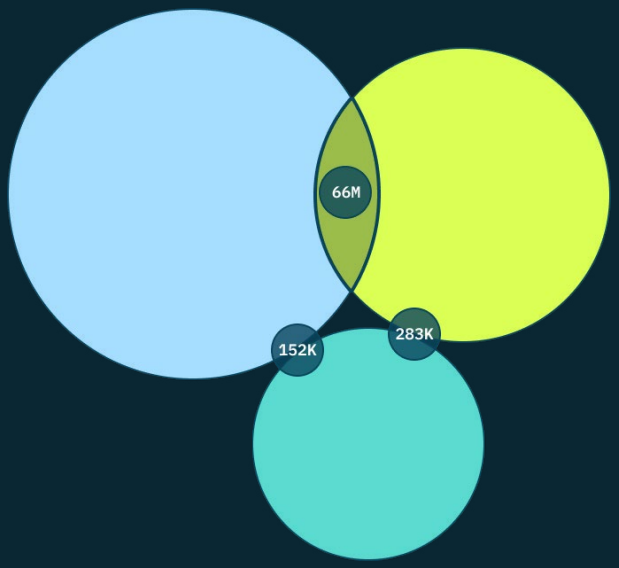
Average-complexity bots account for nearly 1 in 3 bad bot requests. This tier uses browser automation frameworks — primarily Playwright and Chrome DevTools Protocol (CDP) — that more convincingly simulate legitimate browser environments. These tools expose subtle fingerprints in the JavaScript environment that are detectable, but require more sophisticated analysis than WebDriver detection alone. Playwright-based bots are widely used in scraping campaigns, fake account creation, and credential testing.

21.5%
SAMPLED BAD BOT TRAFFIC

21.5% of sampled bad bot traffic was advanced enough to engage with active challenges. This segment represents payload forger activity: automated requests designed to bypass security challenges and evade anti-detection controls. We detected this behavior using signals from WebAssembly (WASM) and our Virtual Machine.

1.4 How are attackers combining techniques?

A key question in bot defense is whether attackers layer techniques, combining simple and advanced methods in a single campaign to overwhelm detection. The data suggests they do not. Simple and advanced techniques almost never appear together within the same campaign window. The overlap between simple and advanced signals is only 0.01% of total volume. Simple campaigns and advanced campaigns are run by different actors using different toolchains for different reasons.



Bot traffic by sophistication level overlap

- Simple bots 735.2M
- Average bots 443M
- Advanced bots 322.2M

But simple and average techniques overlap more often, with 4.4% of the sample showing both WebDriver and Playwright signals in the same campaign window. This suggests attackers are deliberately layering techniques: starting a session with basic automation and escalating to more convincing browser simulation when detection is likely. This hybridization makes attribution and blocking harder; a rule that stops WebDriver bots may miss the Playwright component of the same campaign.

However, attackers using advanced bots rarely combine them with simple or average bots within the same attack. This is most likely due to the fact that a sophisticated attacker will not want to fail by using a simple signal used by basic bots. It is also still true that attackers will most likely still leverage simple bots for vulnerability scanning and reconnaissance prior to launching an attack with more advanced bots. An attacker might also leverage simple bots for easier targets in order to avoid wasting resources.

The implication: Bot defenses that operate on single-signal rules are increasingly inadequate. A model tuned to catch WebDriver still leaves much of the 29.5% of traffic using Playwright exposed, and since the two signals sometimes layer within the same campaign, a defense that stops checking once it spots the first one can still miss the second. A defense built around CAPTCHA challenges might miss the 21.5% of traffic that has already solved those challenges at scale.

Client-side signals are valuable, but they should be one layer in a broader detection strategy. Attackers can suppress or imitate browser-side signals, and some automated requests never execute client-side code because they target APIs, mobile endpoints, or direct server-side requests. Server-side analysis adds visibility into network characteristics, request sequencing, endpoint sensitivity, and session behavior. The strongest defenses combine client-side and server-side signals with behavioral and intent analysis in real time.

Part 2: AI traffic trends in 2026

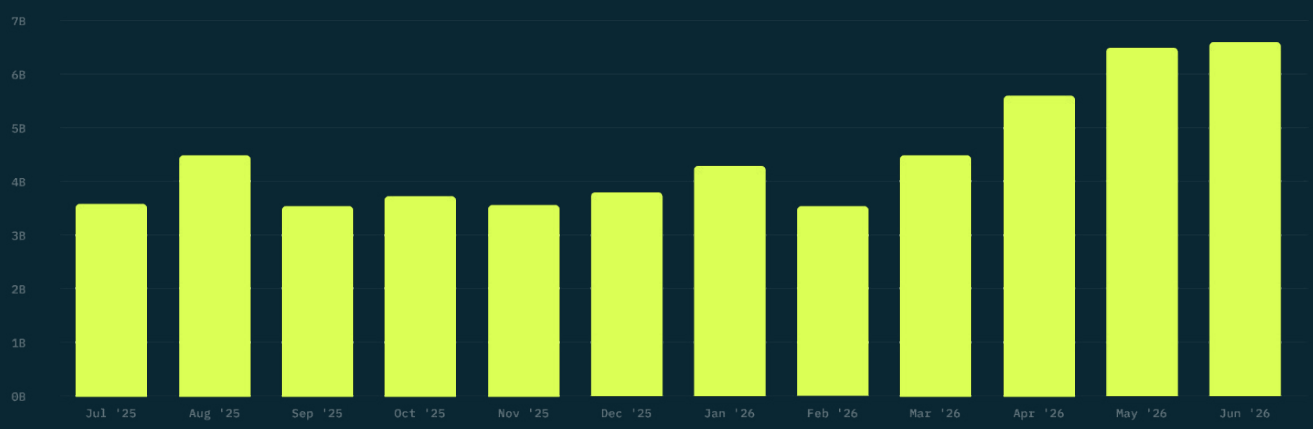
2.0 Key findings

- **AI traffic grew 82.3% from July 2025 to June 2026,** with AI traffic making up 1.7% of total web traffic analyzed by June 2026.
- **Meta is the largest source of AI traffic at 46.3%** followed by OpenAI at 34.6% and Huawei at 5.3%.
- **AI agent spoofing increased by 45% between February and July 2026.**
- **Login endpoint targeting by AI agents increased more than 8x (735.8%) across H1 2026,** signaling a shift from passive crawling to active endpoint probing.
- **ChatGPT.com drives 83.4% of all AI-referred traffic** to websites in our dataset.
- **Smaller AI traffic sources are scaling quickly.** DuckDuckGo recorded a 423% increase in H1 2026, while Comet Browser accounted for 12.2% of Perplexity-attributed AI traffic.

2.1 How fast is AI agent traffic growing?

Looking at all AI traffic between July 2025 and June 2026, DataDome detected 52.7 billion AI agent and LLM crawler requests across its customer base. Monthly volume was volatile, with two months seeing double-digit pullbacks (September and February), but the trend was net upward overall, with the sharpest acceleration concentrated in the final four months, from March through June 2026. If the current pace of growth continues, **agentic traffic is on pace to exceed 79.2 billion annual requests by 2027.**

Monthly AI bot traffic between July 2025 and June 2026



From August through December 2025, traffic averaged 3.89 billion requests per month. Volume rose to 4.11 billion requests in January 2026, dipped to 3.45 billion in February, and then climbed steadily through April. Traffic reached 6.36 billion requests in May and climbed further to 6.6 billion in June, making June the single highest-volume month in the period and showing the trend still accelerating rather than leveling off.

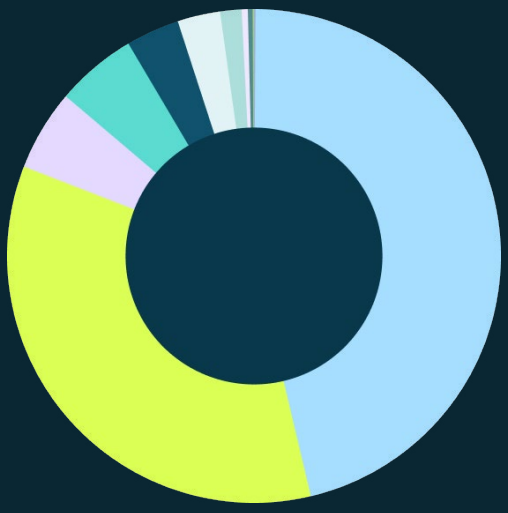
The first half of 2026 generated 29.65 billion AI agent requests, compared with 23.06 billion in the second half of 2025. **Total AI bot traffic grew substantially during the study period, increasing by 82.3% from July 2025 to June 2026.**

This growth coincides with the broader adoption of AI shopping assistants, research agents, autonomous browsing tools, and LLM-powered integrations. These tools are becoming embedded in consumer and enterprise products at scale. As a result, interactions that once required a human click are increasingly generated programmatically across the industries represented in DataDome's customer base.

AI traffic's share of total web requests grew by nearly 50% over the study period, rising from 1.2% in July 2025 to 1.7% by June 2026. That growth reflects AI agents becoming embedded in consumer and enterprise products, generating automated web requests continuously as a side effect of normal use.

2.2 Which AI bots are behind the traffic surge?

DataDome's bot taxonomy classifies AI traffic across more than 100 named AI agents and LLM crawlers. Here's how H1 2026 AI traffic breaks down by attributed company.



AI bot traffic by company during H1 2026

- Meta (Facebook) 46.3%
- OpenAI (ChatGPT) 34.6%
- Huawei (Petal Bot) 5.3%
- Amazon 5.3%
- Anthropic (Claude) 3.5%
- Perplexity 2.8%
- ByteSpider (TikTok) 1.4%
- Google (Gemini & Vertex) 0.4%
- DuckDuckGo 0.3%
- DeepSeek <0.1%
- Manus <0.1%
- Mistral AI <0.1%
- Microsoft Copilot <0.1%
- Apple <0.1%

Meta is the dominant source of AI traffic on the web, generating 46.3% of tracked AI traffic. Its two primary crawlers, Meta-ExternalAgent and Meta-WebIndexer, generated 8.54 billion and 5.30 billion requests, respectively. **Meta's traffic also increased 164.9% from January to June 2026, reaching 2.65 times its January level.**

OpenAI is a distant second at 34.6%. The portfolio includes ChatGPT-User (browsers and agents accessing sites through ChatGPT interfaces), ChatGPTBot (OpenAI's web crawler), and OpenAI-SearchBot.

Huawei's PetalBot is the third-largest AI traffic source by company at 5.3%, with 1.59 billion H1 2026 requests across its crawler fleet. For organizations with significant Asia Pacific traffic, it is a material and growing automated presence that warrants classification and policy.

Google shows a smaller AI traffic source than its scale might suggest, at 0.4%. This reflects Google's strategy of routing AI agent traffic through existing Googlebot infrastructure rather than deploying separate LLM crawlers, meaning Google's actual AI footprint on the web is likely much larger. Gemini can retrieve content from Google's existing search index and cached content rather than sending a separate request to every website.

DuckDuckGo recorded the largest company-level increase in H1 2026 AI bot traffic, with a 423% increase. It reached roughly five times its earlier level, highlighting how quickly company-specific AI activity can scale.

Comet Browser, Perplexity's own AI browser, accounts for 12.2% of Perplexity's total bot traffic, alongside PerplexityBot's 78.2% crawler share.

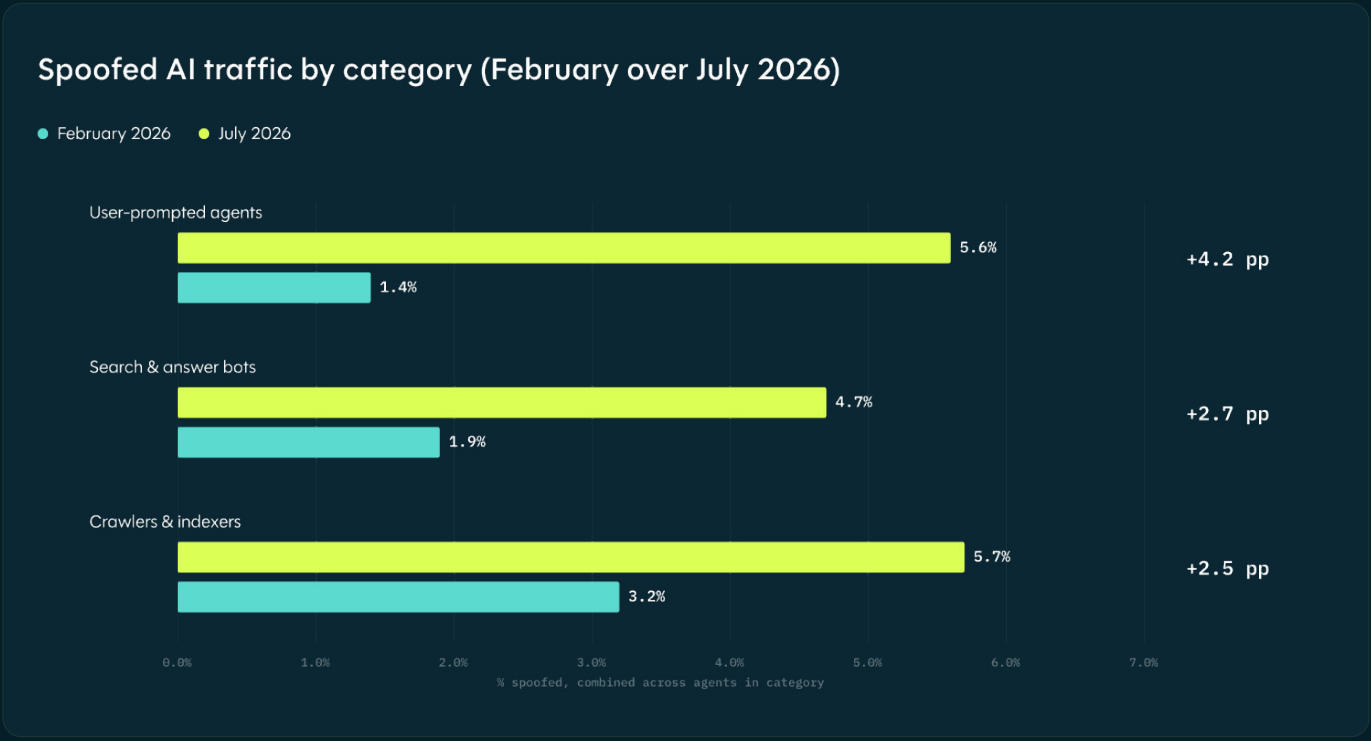
The remaining platforms make up a very small share of tracked AI bot traffic. DeepSeek, Manus, Mistral AI, and Microsoft Copilot each accounted for less than 0.1% of H1 2026 traffic, showing how concentrated the current AI bot ecosystem remains.

2.3 Is AI agent spoofing increasing?

You cannot assume that AI agents are who they claim to be. In our 2026 Future of Search and Discovery Report, **DataDome found that 80% of AI agents do not properly identify themselves when visiting websites,** which creates a visibility problem for sites trying to distinguish between legitimate indexers, training data scrapers, and bad actors spoofing trusted agent identities.

We grouped the tracked agents into three categories based on how they operate. **User-prompted agents** act live on behalf of a human user. This category includes Perplexity-User and ChatGPT-User. **Search and answer bots** fetch content to help answer a query. This category includes OAI-SearchBot and Meta-ExternalAgent. **Crawlers and indexers** collect or process content at scale for search, training, or indexing. This category includes ClaudeBot, PerplexityBot, GPTBot, Meta-WebIndexer, and Google-CloudVertexBot.

These category figures represent the **combined rates of the agents shown in each group**, rather than a traffic-weighted pooled rate. Taken together, the results show that organizations should verify agent behavior and infrastructure instead of relying on a claimed user-agent string alone.



Across the three categories, **AI agent spoofing increased 45% between February and July 2026 across all tracked agents.** User-prompted agents rose from approximately 1.4% to 5.6%. Search and answer bots increased from approximately 1.9% to 4.7%. Crawlers and indexers rose from approximately 3.2% to 5.7%.

User-prompted agents saw the largest increase, suggesting that live AI assistants are becoming a rapidly growing spoofing surface. Search and answer bots also experienced a substantial increase, showing that AI search infrastructure is attracting more impersonation attempts.

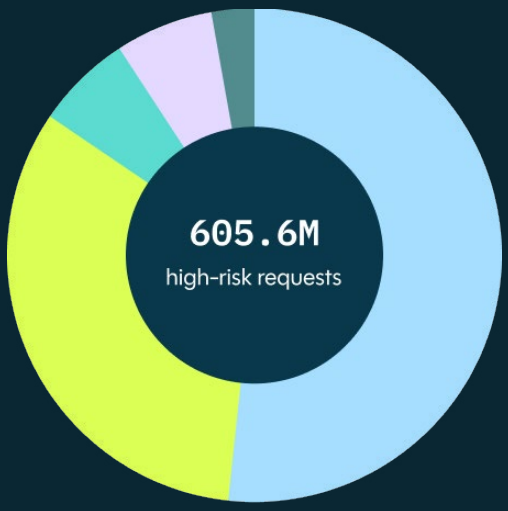
By volume, Meta-ExternalAgent remained the most impersonated identity. It accounted for **16.4 million spoofed requests in February** and **12.7 million in July**. ChatGPT-User ranked second in February, with **8 million spoofed requests**, while ClaudeBot ranked second in July, with **4.1 million.**

At the individual-agent level, PerplexityBot had the highest spoofing rate in February, at **2.4%**. Perplexity-User had the highest rate in July, at approximately **5.1%.**

2.4 Are AI agents targeting high-risk endpoints?

AI agents are not only crawling public pages. They are also reaching login pages, forms, carts, payment flows, and account creation endpoints. DataDome analyzed 29.02 billion AI bot requests across its customer base between January and June 2026. Of that total, 97.9% went to general content such as homepages and informational pages.

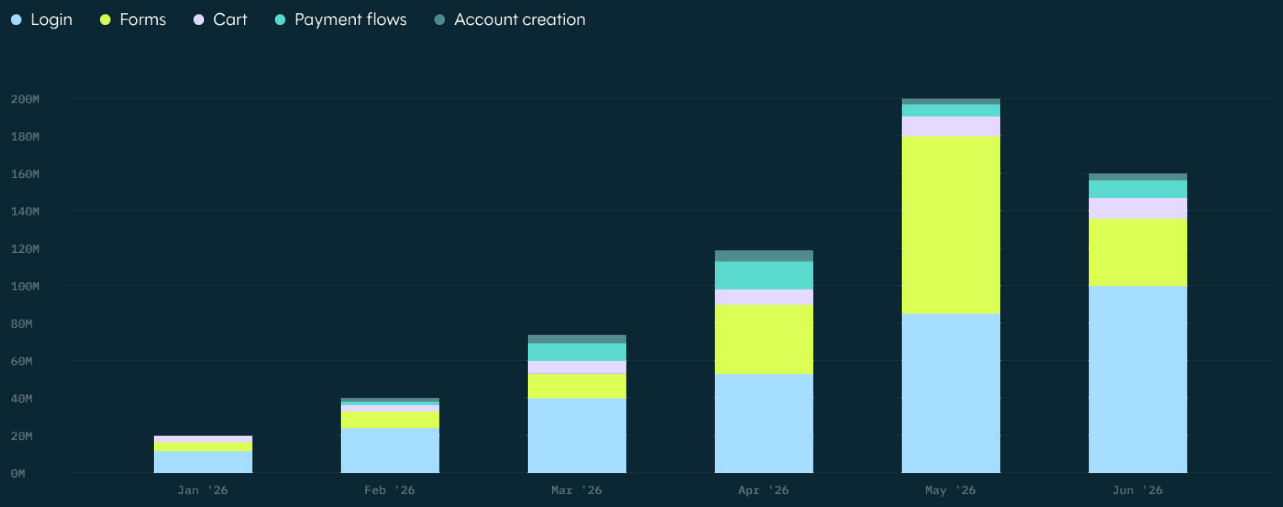
The remaining 605.6 million requests reached endpoints connected to account access, lead capture, commerce, and transactions. Because general content represents a different type of activity, the breakdown below focuses only on this high-risk endpoint traffic.



AI traffic share by high-risk endpoint (H1 2026)

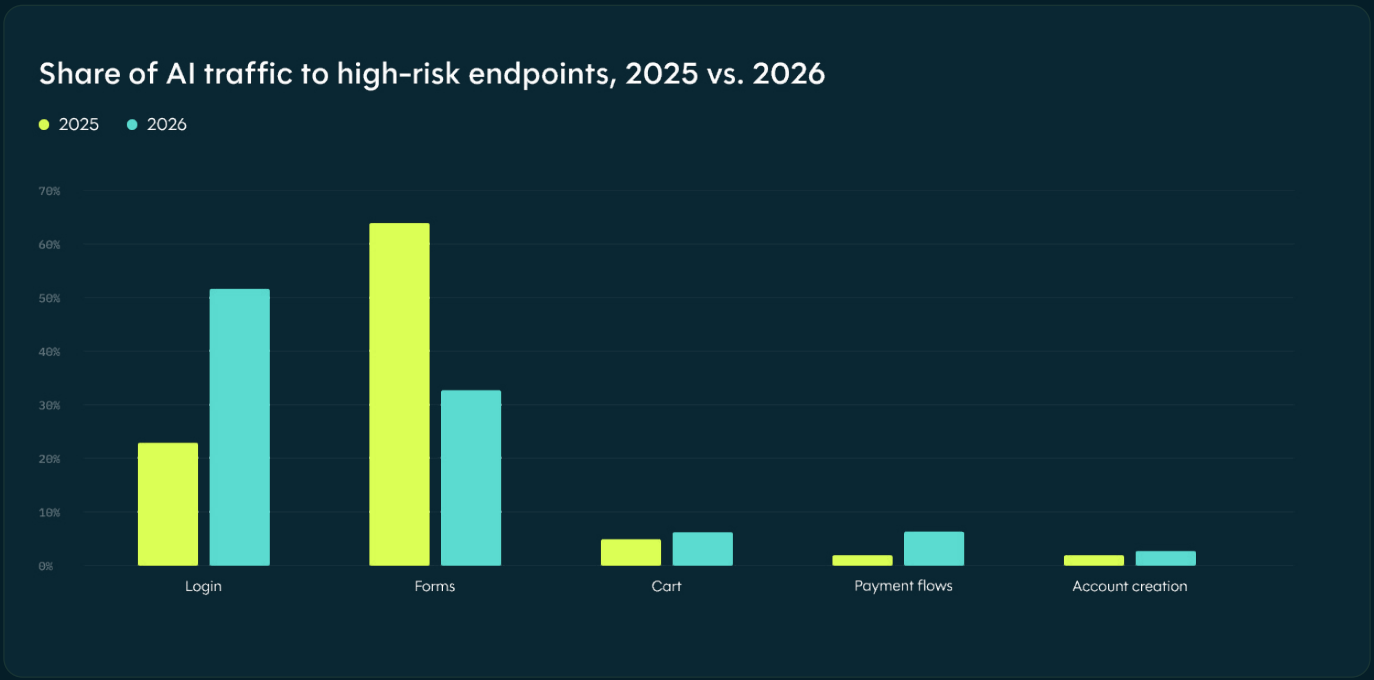
- Login pages 51.7%
- Forms 32.8%
- Payment flows 6.4%
- Cart 6.3%
- Account creation 2.8%

Monthly volume of AI traffic to high-risk endpoints



Login pages received the largest share of high-risk endpoint traffic, with 313.0 million requests, or 51.7% of the total. Monthly login-page targeting by AI bots also increased more than 8x (735.8%) across H1 2026, from 11.9 million requests in January 2026 to 99.7 million in June 2026. Login activity also saw a meaningful increase in share of high-risk endpoint traffic year-over-year, increasing from 23% in 2025 to 51.7% in 2026.

The rise in login activity suggests that AI traffic is moving beyond broad content retrieval toward more active, account-based workflows. Login is often the gateway to personalized information and accounts. Legitimate AI agents may use it to retrieve order information, check loyalty balances, or manage account data on behalf of an authorized user. The same endpoint is also valuable to attackers conducting credential testing, account takeover reconnaissance, and session abuse.



Forms remained the second-largest category, with 198.9 million requests, or 32.8% of high-risk endpoint traffic. Their share fell from 64% in 2025, a decline of 31.2 percentage points. This does not mean that the total volume of traffic to forms decreased, but that form traffic grew more slowly than traffic reaching login, cart, and payment endpoints during the 2026 study period.

The shift may reflect a change in how AI agents interact with websites. Forms often support one-time actions, such as lead submission, registration, or data entry. Login pages provide access to reusable account context and a wider range of personalized actions. As AI agents become more capable of navigating authenticated experiences, they may have more reason to log in than to complete isolated forms.

Forms remain an important abuse surface. They support lead capture, customer service, registration, and other business processes. They are also common targets for input testing, lead fraud, form abuse, and automated data collection. Businesses should therefore evaluate not only how much form traffic they receive, but also whether each request reflects an authorized user action or an automated attempt to manipulate the workflow.

Other high-risk endpoints gained share in 2026. **Cart traffic rose from 5% to 6.3%, payment-flow traffic increased from 2% to 6.4%, and account creation traffic increased from 2% to 2.8%.** Together, cart, payment, and account-creation endpoints received 93.9 million requests, or 15.5% of high-risk endpoint traffic.

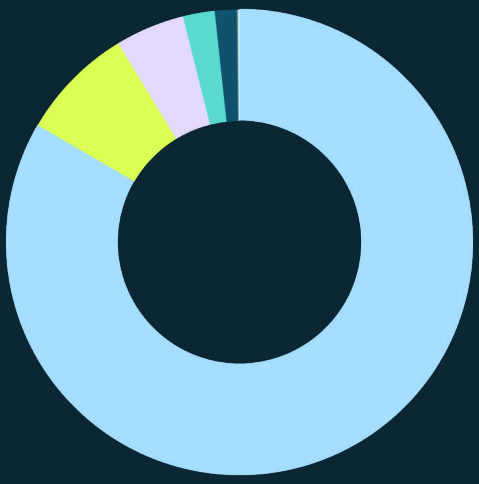
These endpoints connect directly to purchasing, stored value, promotions, and customer accounts. The increase in cart and payment activity is consistent with AI agents becoming more active in commerce-related journeys. It also creates additional opportunities for card testing, inventory abuse, promotion abuse, and fraudulent transactions.

The shift from forms toward login, cart, and payment endpoints shows why endpoint type alone is not enough to determine intent. Businesses need behavioral analysis that evaluates the sequence, speed, context, and purpose of each interaction across the full customer journey.

2.5 How much traffic does AI chat send to websites?

In addition to direct AI crawler traffic, DataDome tracked inbound visits to customer sites referred from AI chat platforms. This is traffic generated when a user in ChatGPT, Gemini, Perplexity, or another AI assistant clicks through to a website from an AI-generated response. Between January and June 2026, DataDome detected 89.4 million AI-referred visits to customer sites.

Between January and June 2026, DataDome detected 89.4 million AI-referred visits to customer sites.



AI referrer traffic share by company

- ChatGPT 83.4%
- Gemini 7.9%
- Perplexity 4.8%
- Copilot 2.2%
- Claude 1.5%
- Grok 0.1%
- Other 0.1%

ChatGPT is the dominant driver of AI-referred web traffic by a wide margin, generating 83.4% of all referrals. Monthly referral volume from AI chat platforms peaked at 19.1 million visits in March 2026 and has sustained at approximately 13 to 14 million per month through mid-2026.

These referrals represent a new traffic acquisition channel, and one with different behavioral characteristics than search or social referrals. AI-referred visitors tend to arrive with high intent, having been pre-qualified by an AI answer. They also arrive with session context that may be invisible to standard analytics: the AI may have already retrieved, summarized, or synthesized content from the destination site before the user clicks through.

Part 3: Bot & AI agent protection benchmark in 2026

3.0 Key findings

- **65.3% of tested websites are completely unprotected,** up from 61.2% in 2025. Nearly two in three popular websites fail to detect any of our 10 test bots.
- **Full protection has declined for the second consecutive year:** 8.4% (2024) → 2.8% (2025) → 2.4% (2026).
- **Spoofed AI bots are blocked or challenged more often than any other bot type, but this isn't necessarily a sign of more sophisticated defenses.** 29.3% of sites stopped at least one spoofed AI bot, likely because blocklist rules added for GPTBot, ClaudeBot, and similar user-agent strings catch spoofed versions too. Any attacker using a custom or unknown agent string passes straight through.
- **More than 7 in 10 sites allow a spoofed AI agent or crawler through without any challenge,** simply because the request claims to be GPTBot, ClaudeBot, or a similar trusted identity.
- **Telecommunications is the least protected industry, with 82.9% of sites completely unprotected.** Tech platforms followed at 77%, and financial services at 73.8%.
- **Marketplaces and classifieds were the strongest-performing industry, with 47.7% of sites showing full or partial protection.** However, more than half — 52.3% — remained completely unprotected.
- **Traffic volume is not a predictor of protection.** Sites in the top 1,000 by traffic had a 64.4% unprotected rate, virtually the same as the lowest-traffic tier.

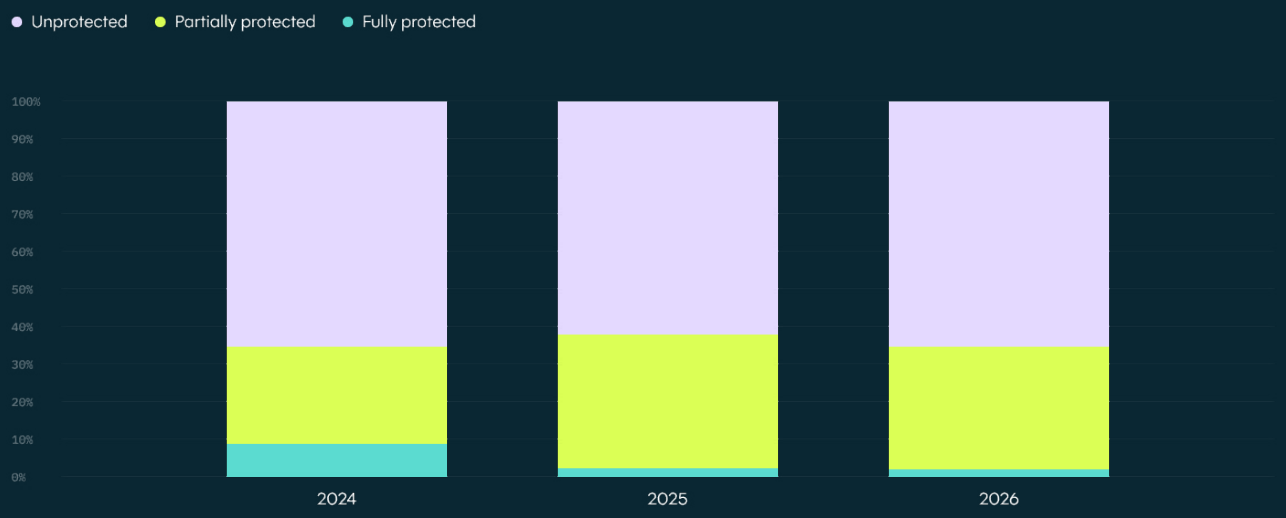
3.1 What percentage of websites are protected against bad bots & AI agents?

In 2026, **65.3% of the 21,491 websites we tested were fully unprotected**, meaning they failed to detect any of the bot types we deployed. Just 2.4% of sites achieved full protection.

Overall bot and AI traffic protection levels

- Unprotected 65.3%
- Partially protected 32.3%
- Fully protected 2.4%

Protection levels: 2024, 2025, 2026



The decline in partial protection (from 36% in 2025 to 32.3% in 2026), combined with an increase in the fully unprotected rate, suggests that incremental or passive defenses are eroding in effectiveness. Sites that previously caught at least one bot type are now failing more consistently across the board.

The most direct explanation: bot tooling is getting better, faster than defenses are keeping up. The anti-fingerprinting bots we tested with in 2026 are meaningfully more capable than those deployed in prior years. As these tools proliferate, increasingly available through low-cost, low-code bot-as-a-service platforms, protection thresholds that held in 2025 are no longer sufficient.

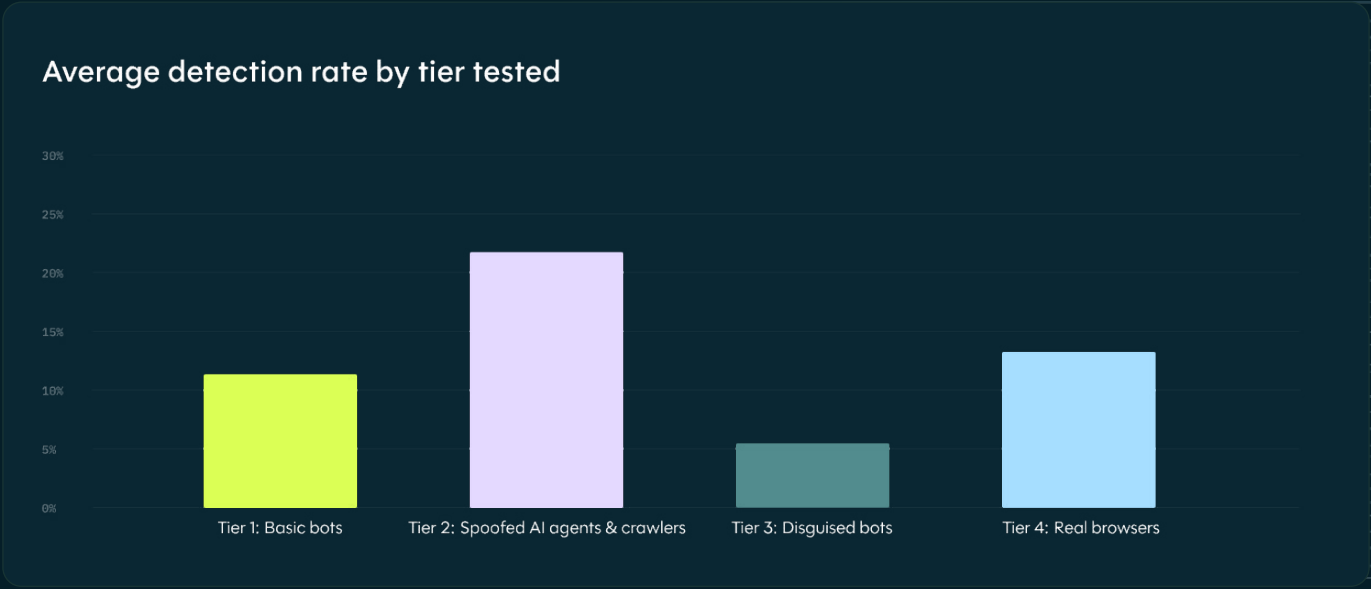
This year's methodology reflects that shift. We expanded our test suite to include spoofed AI agents and more sophisticated bot types, ensuring our findings accurately represent the threat landscape businesses face today.

KEY TAKEAWAY

Full protection has now declined two years in a row. The gap between what websites deploy for defense and what attackers actually use is widening, not closing.

3.2 Which bot types do websites detect best and worst?

Detection rates vary across bot tiers, but remain low across the board. Across all 10 bot types, **65.3% of sites did not stop any test bot.** Among those that did defend, protection is uneven and drops sharply as bot sophistication increases.



A few findings stand out.

Even the simplest bots get through on most sites. Only 5.5% of sites detected every Tier 1 bot — the category requiring the least sophistication to detect. Tier 3, which involves forged browser fingerprints at the network level, is stopped by just 5.5% of sites. Standard defenses — IP reputation, user-agent filtering, simple behavioral rules — are failing against even low-sophistication automation.

More sites are blocking spoofed AI bots than simple bots. Tier 2 has the highest detection rate of any tier: 29.3% of sites stopped at least one spoofed AI agent, and 14.1% stopped all of them. But this is not a sign of more sophisticated defenses. The more likely explanation is that publishers and content platforms added explicit rules for GPTBot, ClaudeBot, and similar user-agent strings, and spoofed versions of those bots get caught by the same rules, not because sites can detect the impersonation, but because the identity being spoofed was already on a blocklist.

These controls do not address attackers using unknown or custom agent strings. As more businesses begin allowing legitimate AI agents to access their sites, this weakness becomes more consequential. Static allowlists and blocklists cannot distinguish an authorized AI agent from a malicious bot impersonating one. Businesses need to verify agent identity and evaluate behavior and intent rather than rely on user-agent strings alone.

Identity-based trust is a widespread vulnerability. More than 7 in 10 sites allow a spoofed AI agent or crawler through without any challenge, simply because the request claims to be GPTBot, ClaudeBot, or a similar trusted identity. Tier 2 does have the highest "stopped all" rate of any tier at 14.1% — reflecting the growing number of sites that have added blanket AI-crawler blocks, typically through robots.txt enforcement or user-agent rules. But that protection has a critical ceiling: it works only against bots that announce themselves honestly. Any bad actor spoofing one of these identities could pass straight through.

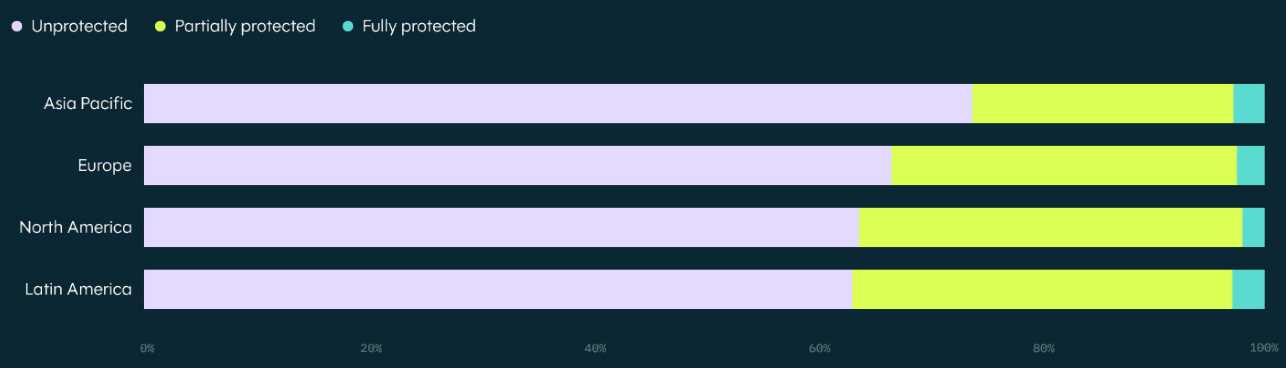
Real-browser bots are the hardest to detect, and most sites blocking them may not be doing so for the right reasons. Only 3.1% of sites stopped every Tier 4 bot tested. Critically, sites stopped the cloaked and uncloaked real-browser builds at nearly identical rates, a difference of just 0.05 percentage points. This suggests most are applying an indiscriminate challenge to any browser-like request, rather than genuinely reading the automation signal.

3.3 Which region has the weakest bot protection?

The test distributes requests from three proxy locations — the United States, France, and Canada — to ensure results are not driven by IP-based geographic filtering. Detection results were broadly consistent across locations, confirming that the protection gaps identified are structural, not geography-specific.

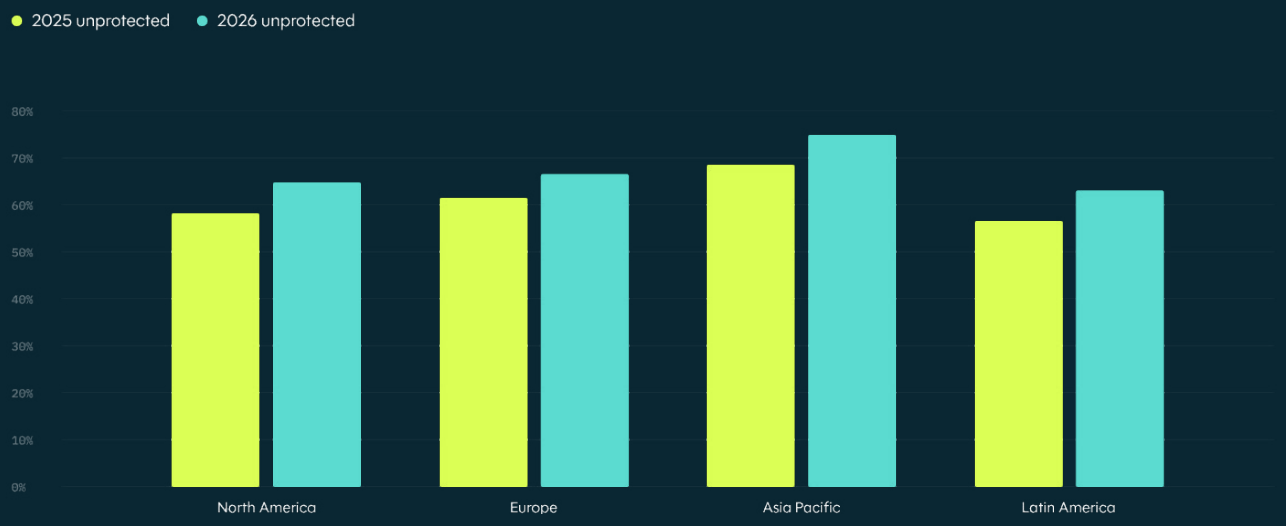
Asia Pacific has the highest unprotected rate at 73.9%, followed by Europe at 66.7% and North America at 63.8%.

Bot protection by region in 2026



All major regions declined year-over-year. North America's unprotected rate moved from 59.8% in 2025 to 63.8% in 2026. Europe went from 61.5% to 66.7%. Asia Pacific grew from 68.4% to 73.9%. Latin America's unprotected rate also moved from 58.1% to 63.2%.

Change in regional protection, 2025 vs. 2026



KEY TAKEAWAY

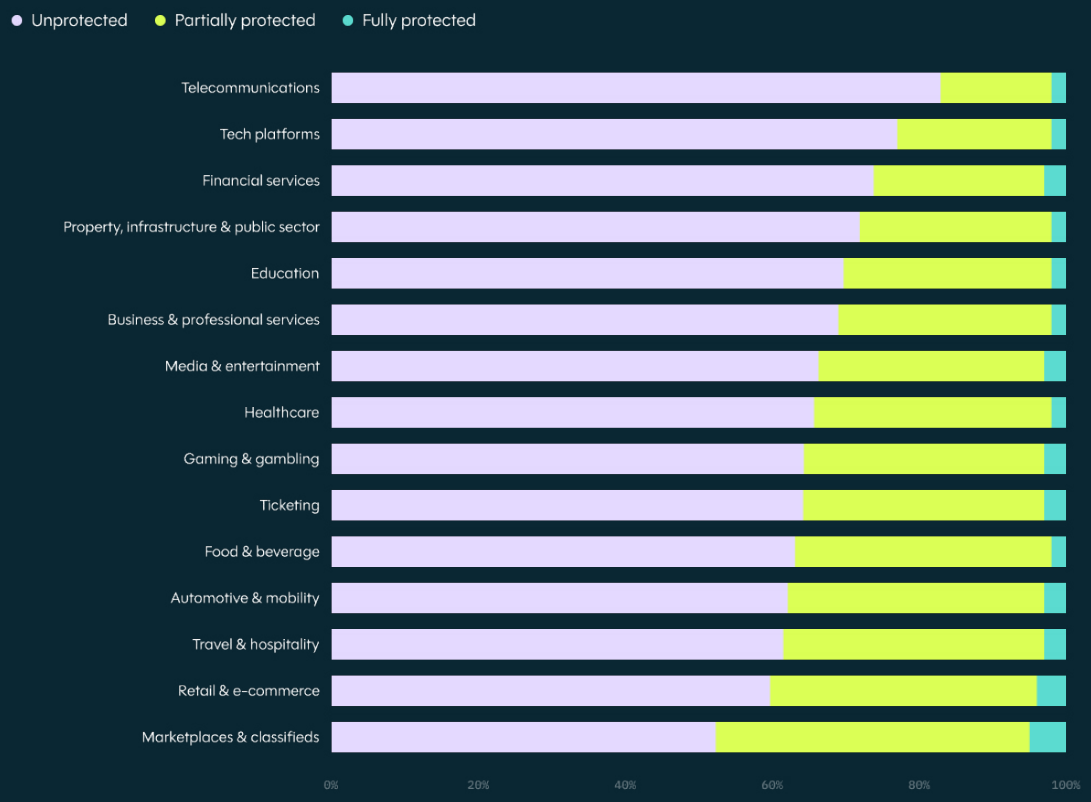
No region is improving when it comes to overall protection. While there were some shifts between full and partial protection, every region saw a decline when the two categories are combined. In other words, fewer websites are stopping at least one of the bots tested, and more are allowing every test bot through. The most mature digital markets, North America and Europe, still have nearly two-thirds of websites fully unprotected.

3.4

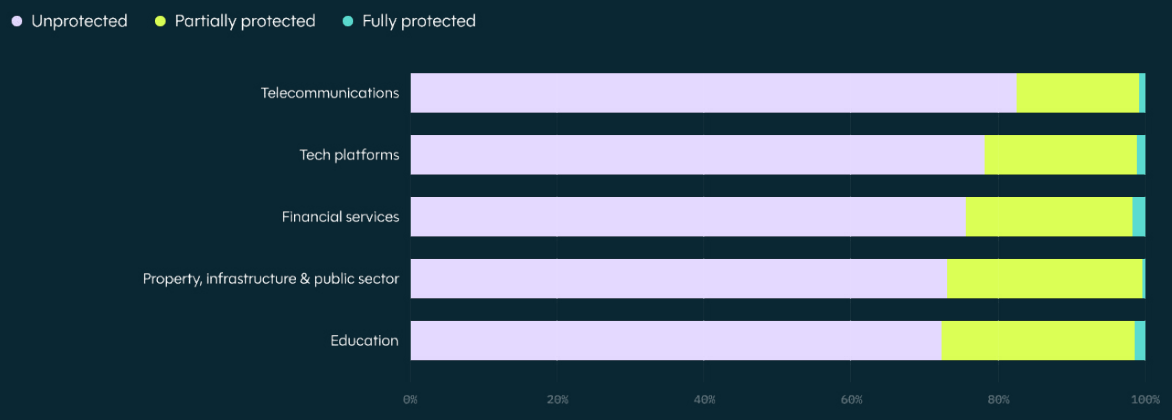
Which industry is most vulnerable to bad bots & AI agents?

We grouped the tested websites into 15 industries. Protection levels vary significantly by sector, reflecting differences in business models, the value of assets at risk, and security maturity.

Protection levels by industry



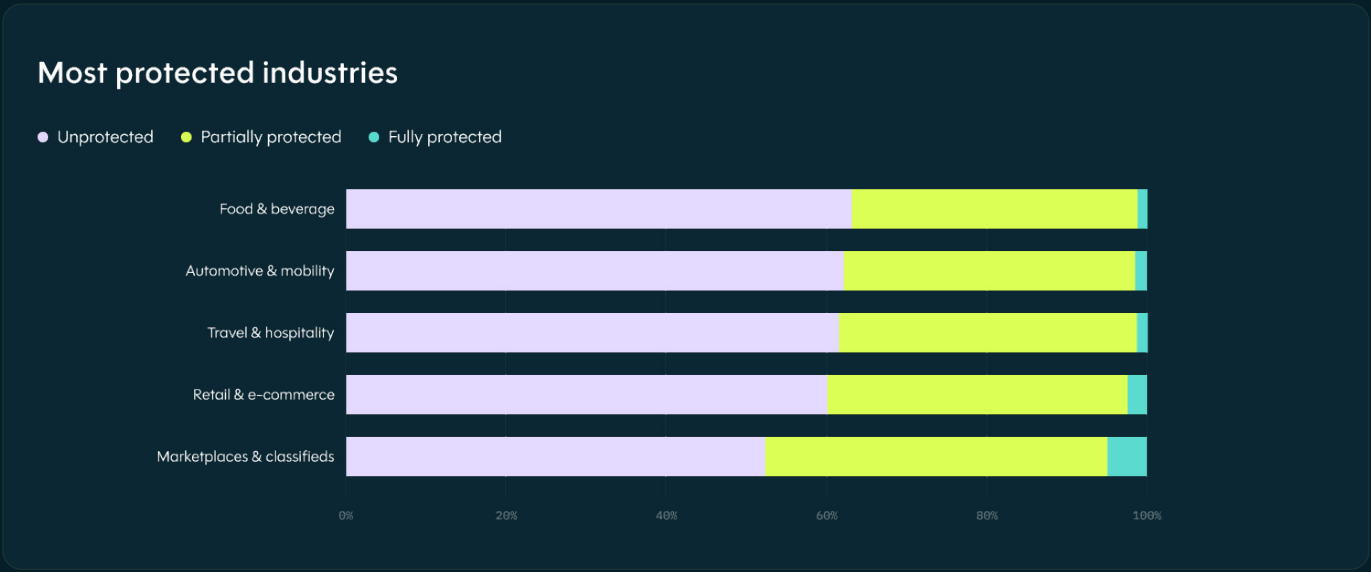
Least protected industries



Telecommunications tops the least-protected list at 82.9% unprotected. Tech platforms follow at 77%, showing that broad digital reach does not automatically translate into strong bot protection.

Financial services rank third at 73.8% unprotected. Property, infrastructure & public sector follows at 71.9%, while Education rounds out the bottom five at 69.7% unprotected.

These results show that protection gaps are not limited to one type of business. Telecommunications and technology platforms face broad exposure, while financial services continue to carry significant risk despite the value of the data and accounts they protect.



For ranking purposes, protection was measured as the combined percentage of fully protected and partially protected for each site.

Marketplaces & classifieds are the most protected industry, with 47.7% of sites showing at least partial protection. Retail & e-commerce follow at 40.4%, making these the two strongest-performing categories.

Travel & hospitality ranks third at 38.5% combined protection, followed by Automotive & mobility at 37.9% and Food & beverage at 36.8%.

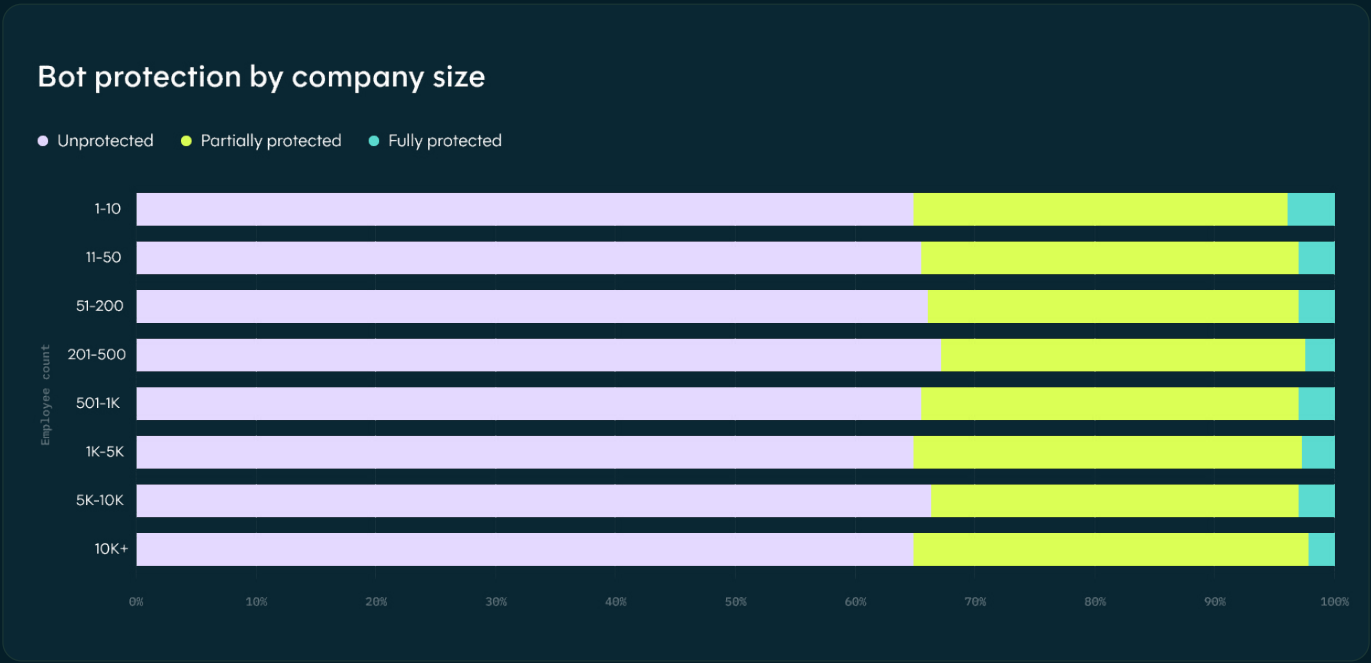
Marketplaces & classifieds are the strongest-performing category, but 52.3% of sites in the category remain unprotected. Even the best-performing industries therefore have a clear opportunity to improve detection across more advanced bot types.

KEY TAKEAWAY

Protection is uneven across industries, but no category has reached a strong baseline. Unprotected rates range from 52.3% in Marketplaces & classifieds to 82.9% in Telecommunications. Consistent, behavior-based detection is needed across every sector, especially as attackers reuse the same automation tools across different business models.

3.5 Do bigger companies have better bot protection?

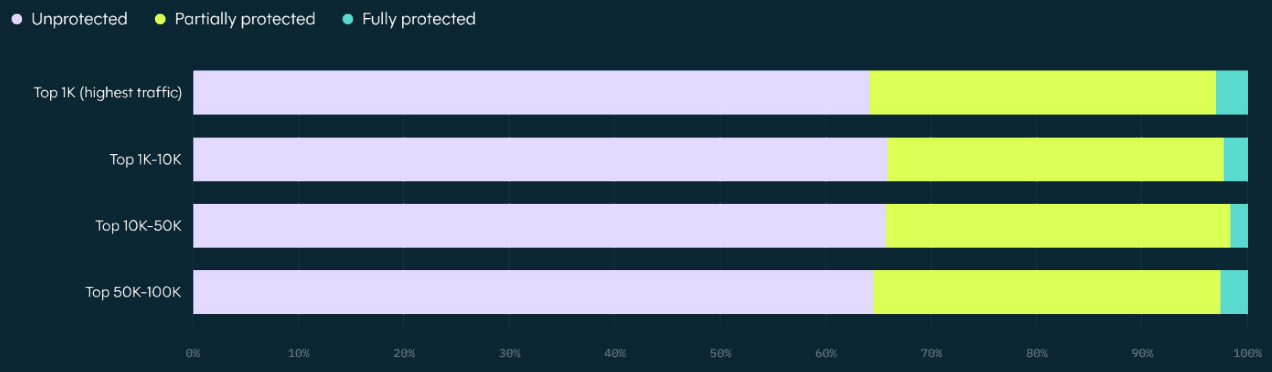
Protection levels show very little variation across company size tiers. One might expect large enterprises with substantial security budgets to significantly outperform small businesses. In 2026, the gap is narrow to the point of being negligible.



The largest companies (10,000+ employees) have an unprotected rate of 64.6%. This is worse than the very smallest businesses (62.7%). Large enterprises contend with the complexity of protecting vast and diverse digital infrastructures, often with fragmented tooling and legacy systems that create gaps.

Counterintuitively, the highest-traffic websites are not better protected. Sites in the top 1,000 by monthly traffic had a 64.4% unprotected rate, virtually the same protection rate as sites in the lowest-traffic tier tested.

Bot protection by website traffic tier



High-traffic sites are more attractive targets. They host more users, more data, and more monetizable assets. Not having stronger defenses makes them disproportionately valuable to attackers.

KEY TAKEAWAY

Neither company size nor traffic volume predicts protection quality in 2026. The most powerful websites on the internet are failing the same basic bot tests as the smallest.

Takeaways & recommendations

The 2026 data demands a new approach to detection

The 2026 data tells a consistent story: the gap between what most websites deploy for defense and what attackers actually use is widening. **Only 5.5% of sites stopped every basic bot** — the lowest-sophistication attack in the test set. For real-browser bots, that figure drops to 3.1%. The tools enabling these attacks — from fingerprint-spoofing frameworks to bot-as-a-service platforms — are increasingly accessible and affordable. Defenses that cannot stop simple bots will not stop the fraud and abuse that follows.

The challenge in 2026 is not only about stopping more bots, but making better decisions about all automated traffic. Traffic type alone is no longer a reliable signal. An AI agent hitting a login page could be an authorized customer tool or an attacker running a credential stuffing campaign. The security question is no longer "is this a bot or a human?" It is: "What is this visitor trying to do, and does it serve my business?"

Intent-based detection answers that question by evaluating not just who or what is making a request, but what it is trying to accomplish — using signals like behavioral sequencing, endpoint sensitivity, rate and volume patterns, and session context. The goal is not to block automation. As AI agents become active participants in commerce, productivity, and research workflows, the right automation should operate freely. The task is identifying and stopping the wrong automation in real time.

Key recommendations

- **Implement intent-based detection, not just bot detection.** Traffic type alone does not determine risk. AI agents, good bots, and human users all generate web requests. The security decision is not whether a visitor is automated — it is whether the automation is authorized, behaving consistently with its declared purpose, and targeting appropriate endpoints. Detection systems that classify based on identity signals alone (user-agent, IP, header patterns) will continue to miss the majority of modern threats.

- **Extend protection to high-risk endpoints, especially APIs and login flows.** Login endpoint targeting by AI agents grew more than 8x (735.8%) across H1 2026. Forms, APIs, and cart flows are increasingly touched by automated traffic — both legitimate and malicious. Protection that covers the homepage and misses the login page, checkout flow, mobile apps, or a Model Context Protocol (MCP) endpoint is incomplete. Access control policies for AI agents need to be applied with the same rigor as access control for human users.

- **Test your bot and AI agent defenses.** Tool presence does not equal effective protection. should test their own sites to understand what actually gets through. DataDome makes this easy to do anytime for free. Test your site.

- **Define your AI agent access policy.** AI agents will interact with your website whether or not you have a policy for them. The question is whether those interactions are governed. A published AI agent access policy — specifying which agents are permitted, what they can access, and under what conditions — gives your security team the criteria to classify traffic as authorized or unauthorized, and gives legitimate AI systems the information they need to operate appropriately.

Named a Leader in The Forrester Wave™: Bot And Agent Trust Management

Trusted by over 400 leading brands worldwide

Etsy

The New York Times

SOUNDCLOUD

ASUS

Tripadvisor

petco

PayPal

Why DataDome?

Traffic you can trust

DataDome delivers real-time bot and agent trust management, giving businesses complete visibility and control over all traffic — human, bot, or AI.

Acting as a traffic control plane, DataDome's multi-layered AI engine leverages thousands of models and 5 trillion signals daily to analyze intent and stop fraud in under 2 milliseconds — letting legitimate users through seamlessly.

Key platform capabilities:

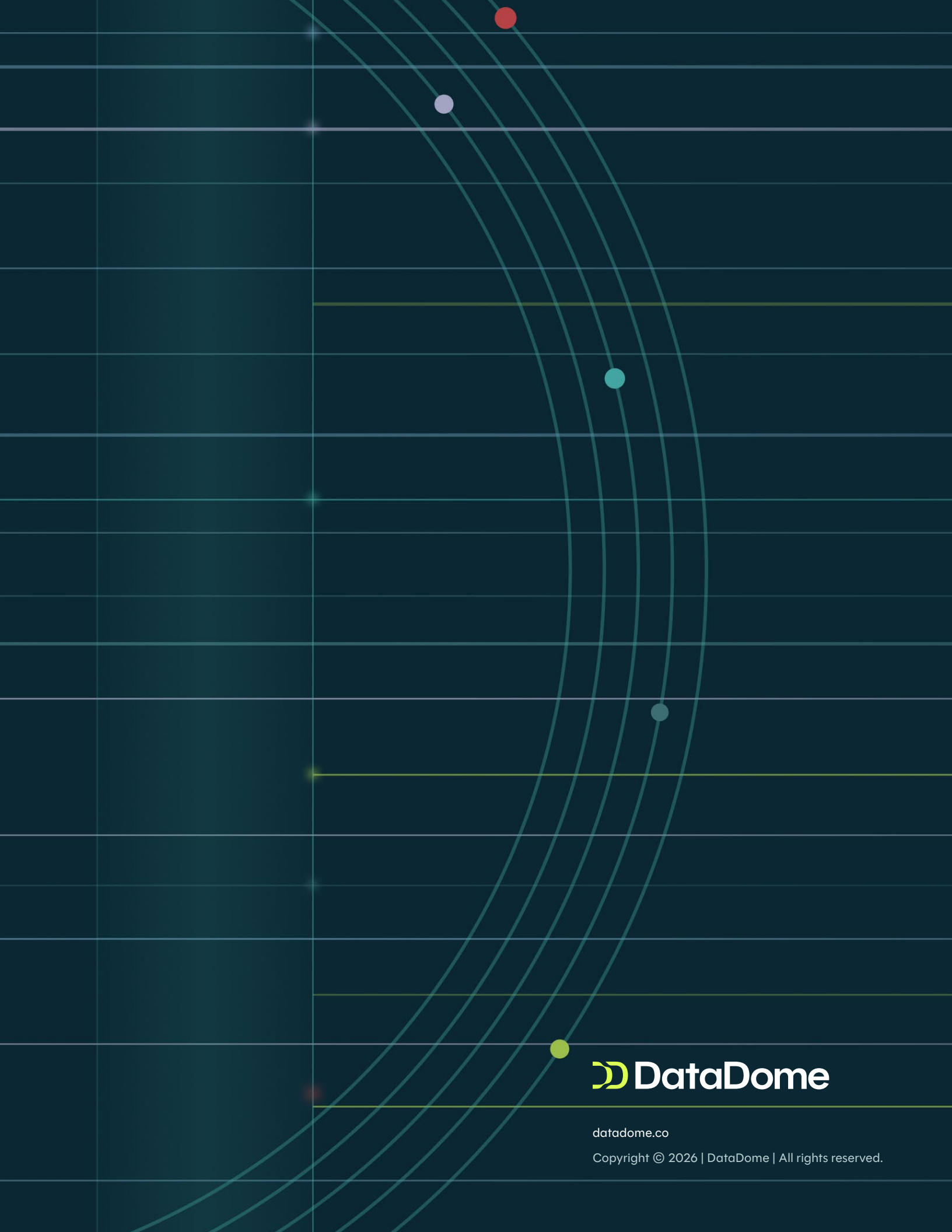
- **Intent-based detection in real time:** DataDome evaluates behavioral intent — not just identity — to distinguish authorized automation from malicious activity
- **Full-journey coverage:** protection across web, mobile, API, and MCP server endpoints
- **Industry-leading accuracy:** less than 0.01% false positive rate, ensuring legitimate users and authorized AI agents are never incorrectly blocked

See it for yourself

If this report raises questions about your own exposure, there are two ways to find out where you stand.

Test your site's defenses for free. See exactly which bot and AI types get through, and where your protection gaps are.

Talk to our team to see how DataDome works across your full traffic environment, including web, mobile, API, and MCP endpoints.



DataDome

datadome.co