



Quarterly Threat Report

Q2 2025

Table of Contents

Introduction	3
Intelligence requirements	3
How to read this report	3
Executive Summary	4
Threat Landscape	4
Scraping	4
Threat Landscape Update	6
Analyst Insights	6
Account Takeover (ATO)	10
ACCOUNT SALES BY INDUSTRY	10
ACCOMMODATION	11
AIRLINES	11
ENTERTAINMENT & TICKETING	11
GAMING	12
QUICK SERVICE RESTAURANTS (QSR)	12
RETAIL	12
SOCIAL NETWORKS	13
STREAMING	13
WEBMAIL	13
Spotlight On: Scraping	14
Why Scraping Matters to You	14
Profiling Scraping	15
Adversary Insights	18
Looking Ahead	29
Threat Intelligence at Kasada	29

Introduction

Kasada's Q2 CY2025 (Q2) Threat Intelligence Summary arms you with the contemporary observations and assessments you need to meet the evolving challenges set by automated threats. With millions of unique events collected by KasadaIQ each quarter, this Threat Intelligence Summary cuts through the noise and pulls out the key areas your organization needs to focus on to stay ahead of the curve. All insights are based on analysis and investigations via our in-house KasadaIQ intelligence capability, supported by deep access to open and closed sources and decades of technical and intelligence expertise combined.

Intelligence requirements

This report provides insights against the following intelligence requirements:



What were the top automated threats observed by KasadaIQ in Q2 CY2025?



How did the top automated threat observed by KasadaIQ in Q2 CY2025 impact key industries?



What movements in the automated threat landscape did KasadaIQ observe in Q2 CY2025?



Which automated threat required the most attention in Q2 CY2025?

How to read this report

Executive Summary →

A short summary of the key themes and findings from the report, with insights on where to focus and how Kasada can help.

Threat Landscape Update →

An overview of the top threats tracked by KasadaIQ over the past quarter, key industry trends and notable events across the landscape.

Spotlight on: Scraping →

A deep dive into Kasada's focus area for the past quarter, which for Q2 was scraping, using a "what", "who", "why" and "how" model. Includes [Adversary Insights](#).

How we can help →

An overview of how KasadaIQ can empower you with intelligence and the kinds of questions we can help you answer to meet the threats discussed in this report.

Executive Summary

This report provides an overview of the automated threat landscape in Q2, with a particular focus on the growing threat of unwanted scraping.

Threat Landscape

In Q2, **account takeover (ATO)** remained the most frequent and impactful threat. From government to superannuation to major retailers, credentials have been abused to breach the confidentiality, integrity and availability of critical goods and services. The top three most targeted industries for ATO and credential stuffing were **retail, QSRs and webmail**, constituting approximately **72%** of all account sales tracked by KasadaIQ in Q2.

ATO is likely to stay top threat for businesses as it sits at the intersection of low adversary cost, high business impact and automation, making investment in a strategic and enduring response a business need.

AN INTELLIGENT APPROACH

KasadaIQ monitors criminal marketplaces and online forums for fraud and abuse indicators, including credential listings, as well as open source for breaches. This data can be used to proactively identify credentials linked to user bases and assist CISOs and security teams with risk-based decision making.

Kasada's defenses are informed by dynamic client-side challenges, fingerprinting and real time telemetry, using behavioral biometrics and session analysis to identify bot-like behavior and recommend actions to security, IT & Infrastructure teams.

Scraping

The scale, speed and frequency of scraping requires all businesses to have a position on scraping, identifying the line between permitted scraping and unwanted scraping. Here's why:

-  **Unwanted scraping is costing you, and those costs are rising.** As the scale and frequency of scraping continues to increase, businesses are paying more in infrastructure costs for adversaries to profit off their hard work and data. Particularly for businesses with [cloud-native environments](#), you may be paying more to serve the adversaries exploiting and profiting off your data and infrastructure.
-  **Unwanted scraping can directly interfere with [dynamic pricing](#) mechanisms** relied on to optimize revenue, undermining your business strategy.
-  **Unwanted scraping treats data as a tradeable asset.** This means that scraping is not a background risk. Scraping is a strategic issue touching on competitive advantage, data ownership, infrastructure costs and brand integrity.

Whether scraping should be blocked or not is a choice for each business, but one that can no longer be avoided. AI has significantly amplified the scale, speed and impact of scraping, forcing businesses to reach an inflection point when it comes to scraping.

- **In Q2, Kasada observed over 120 million requests by AI scraper user agents to our customers.** Of these, the majority originated from the OpenAI user agent, which ranged from 322,000 to nearly 950,000 requests per day.
- **Your data could be powering AI tools without attribution or revenue, while also losing traffic, ad revenue and customer acquisition pipelines.** AI scrapers don't just copy your data, they learn from and repurpose scraped content, making it hard to trace and monetize.
- **Ethical data sourcing is becoming a brand issue, not just a tech one.** AI developers are increasingly using aggressive scraping tools to collect high volumes of data, often despite clear disallow rules in robots.txt. With the [indiscriminate nature of AI scraping](#), a failure to have a position on unwanted scraping, particularly where ethically problematic data is involved, could lead to regulatory action, brand erosion, social backlash or investor concern around Environmental, Social, and Governance (ESG) misalignment.
- **The threat is scaling but there are things your business can do to get on the front foot.** The first step for businesses is landing on a position for whether scraping is unwanted or permitted, and understanding why. Kasada has provided some [factors to consider](#) in internal discussions and decision-making.

THE IMPORTANCE OF NUANCE AND CUSTOMIZATION

Kasada understands that scraping can be a grey area for businesses. We understand that for some businesses, there are positive impacts, like SEO boosts and exposure; that's Kasada classifies bots as "good" and "bad". However, there is a cost to pay when you cannot turn these things off. When it comes to scraping, the playing field is uneven and the target is often moving. Kasada provides the tools to enforce a business' position on scraping effectively. Our defenses include:

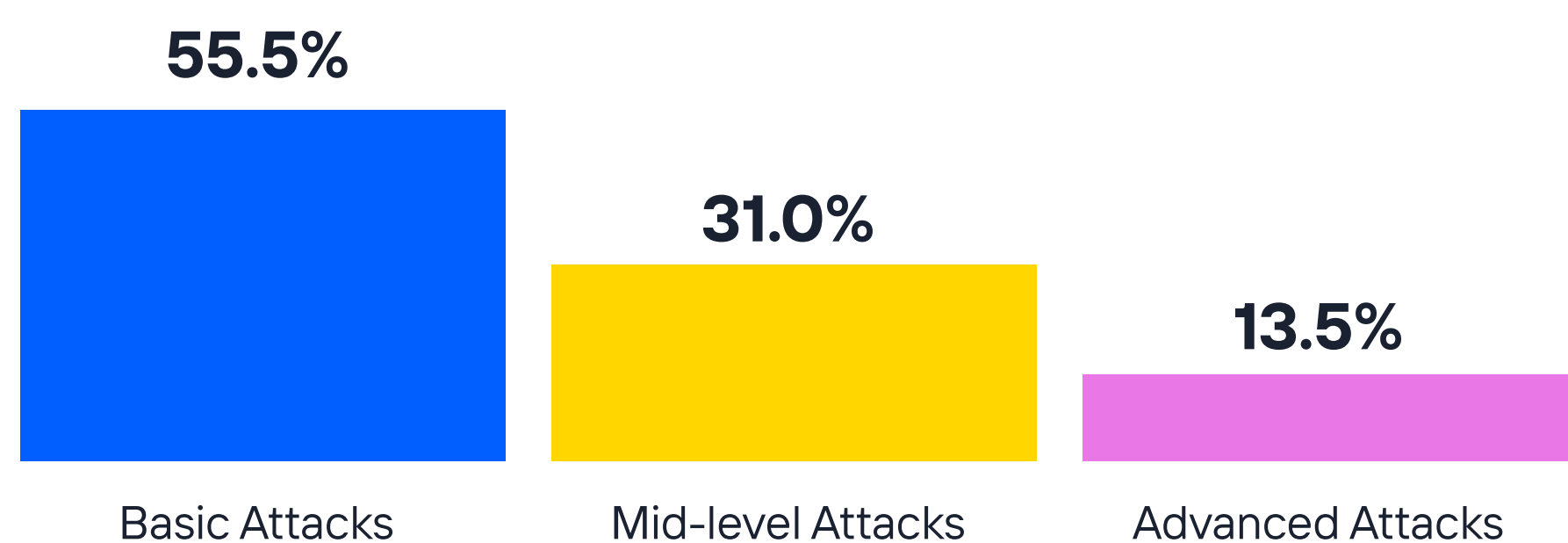
- ✓ **Client-side telemetry to detect automation from the first request.**
- ✓ **A "proof of execution" system that forces bots into a controlled environment.**
- ✓ **Built-in protections against known AI scrapers like OpenAI, Bytespider, and Anthropic.**
- ✓ **Adversary intelligence and access to a wide range of online communities and tooling.**
- ✓ **Layers of security, including client-side validation, anomaly detection and invisible computation challenges.**
- ✓ **No CAPTCHAs: Kasada avoids user friction by never using CAPTCHAs.**

Threat Landscape Update

- ▶ ATO remained the **most frequent and impactful threat**.
- ▶ The number of stores selling stolen accounts **increased across all industries** tracked by KasadaIQ.
- ▶ **Retail was the top target** for ATO based on account sales and revenue generated for adversaries.
- ▶ There was a further resurgence of **reselling communities and hype items**.
- ▶ Kasada observed over **120 million requests by AI scraper** user agents to our customers.

BOT SOPHISTICATION

Similar to Q1, sophistication of bot activity observed by Kasada varied in Q2 but the majority of attacks were assessed as basic. We note that basic does not equate to low impact, with basic bot activity often scalable and overlooked.



Analyst Insights

Based on observations from our KasadaIQ analysts in Q2 2025, below are some key insights to keep in mind for Q3 2025.

- In June 2025, the United States District Court for the Northern District of California (US District Court) ruled that Anthropic did not infringe on copyright materials by training its Large Language Model (LLM), as the work it produced was considered original. The US District Court did find that Anthropic downloaded over 7 million pirated books which could be illegal and requires a separate trial.
- In a separate case, the US District Court similarly ruled that Meta AI's use of materials produced by thirteen authors was highly transformative, citing insufficient evidence about how outputs of Meta's models would dilute the market.²



We expect to see increased focus on how data is scraped and repurposed in legal cases involving AI scrapers, particularly with regard to market implications.

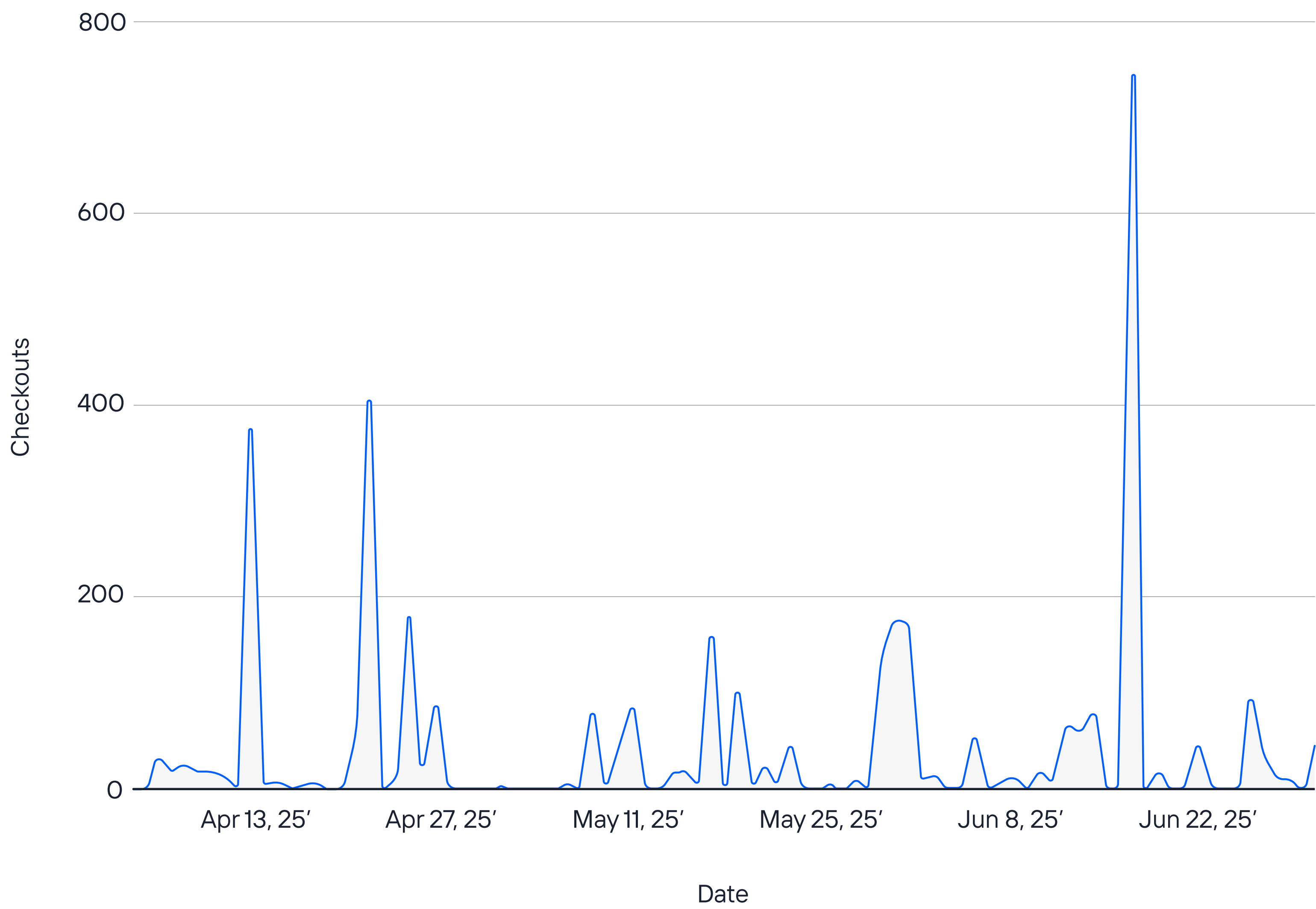
¹ <https://www.documentcloud.org/documents/25982181-authors-v-anthropic-ruling/>

² <https://fingfx.thomsonreuters.com/gfx/legaldocs/zgvozmrynpd/META%20AI%20COPYRIGHT%20LAWSUIT%20ruling.pdf>

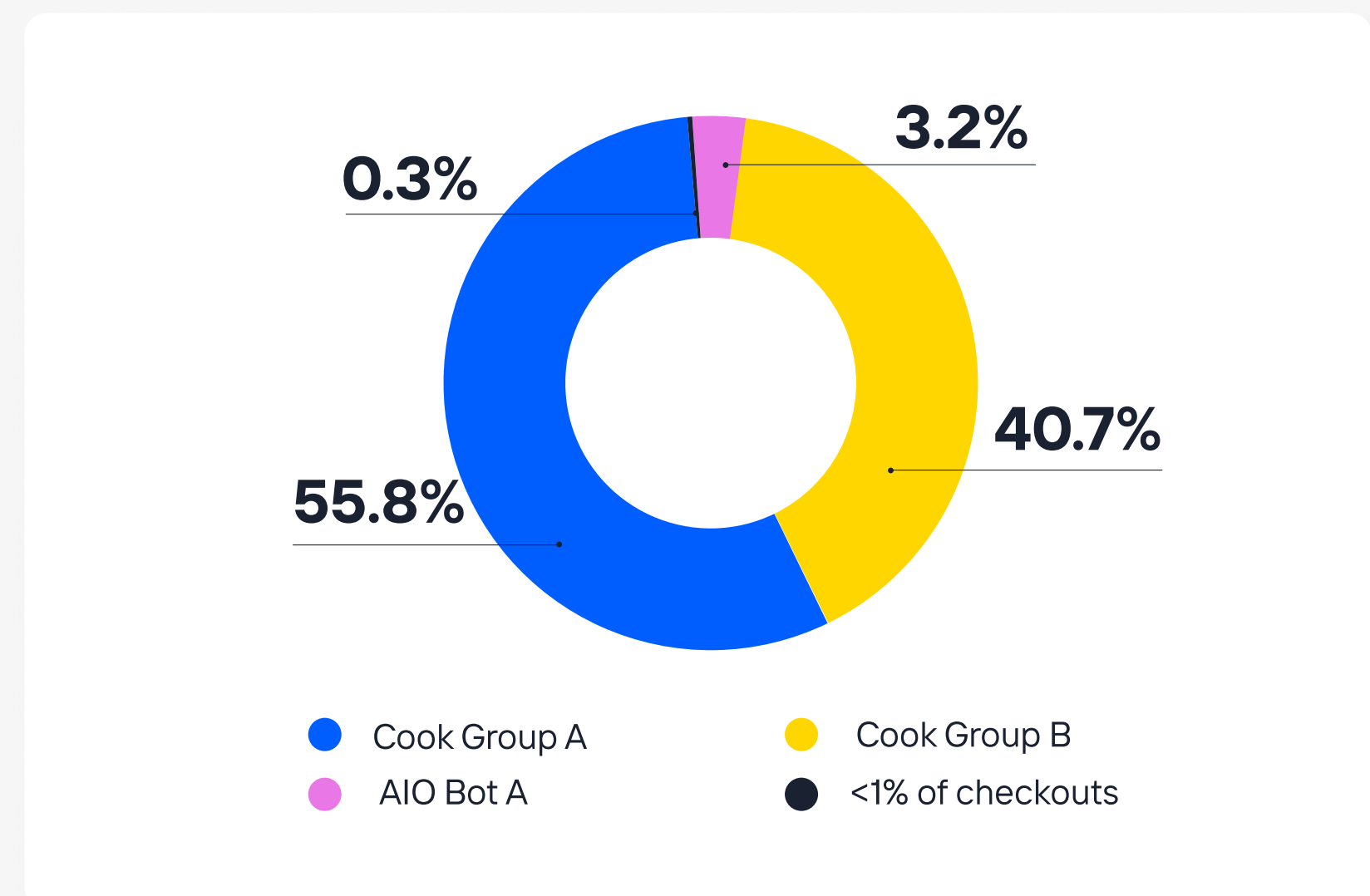
- The rulings were in response to complaints made by writers and journalists alleging that Anthropic and Meta illegally used their works to train the companies’ respective chatbots.
- These rulings are among the first to address the intersection between copyright law and AI training, meaning it could be used as precedent for several pending lawsuits.
- The rulings characterise AI’s use of the material as "transformative", an argument that could be transferred to other kinds of proprietary data held by businesses. They further reiterated the importance that how data is scraped and used matters, meaning that businesses could argue that their data wasn't just scraped, but repurposed in a way that undermines market value.
- Brands are returning to scarcity marketing, triggering a reset of reselling communities and a new wave of cultural hype.
- Collectible designer toy universe, The Monsters, has become the subject of major hype in 2025, particularly the Labubu doll. In Q2, KasadaIQ observed **3,160** bot checkouts of Labubu dolls.



As predicted in Q1, KasadaIQ has observed a further resurgence of reselling communities and refocusing on hype items. We anticipate this trend will continue for the remainder of the year.



- Labubu dolls are difficult to buy at retail, with blind-box series and limited edition figures selling out in seconds of a drop. The combination of limited-runs and blind-box series, targeting by resellers and viral social media trends are driving scarcity of Labubu dolls and the broader The Monsters universe.
- What has emerged is an aggressive resale market. Botters are not collecting these items, they are **flipping them**; selling them in resale communities and marketplaces within minutes of a drop.
- Some Labubu dolls are reselling for hundreds or thousands of dollars. From a sample of listings on online botting communities, Labubu dolls are sold for between **25-127%** more than retail price.
- Approximately **70%** of checkouts related to a single retailer.
- Two cook groups³ known to Kasada were responsible for the majority of checkouts, noting Labubu dolls are not their only focus.



A lack of basic security mitigations will continue to be subject to government and public criticism of businesses who have faced successful credential stuffing attacks. Password reuse is likely to be an ongoing challenge, with credentials from previous breaches recycled by adversaries.

- Businesses experiencing successful credential stuffing attacks are facing significant **public and regulatory scrutiny**.

CASE SNAPSHOT | AUSTRALIAN SUPER FUNDS HIT

In April 2025, a large-scale credential stuffing campaign hit several large Australian super funds, compromising tens of thousands of member's accounts. Subsequent analysis by security vendor, Proofpoint, found that 58 per cent of Australian super funds failed to implement basic email security protocols. The Australian Prudential Regulation Authority warned Australian superannuation trustees that they must urgently close gaps in cyber security systems, requiring all funds to complete a self-assessment on information security controls by 31 August 2025.

- The **financial losses** for businesses experiencing successful credential stuffing attacks are substantial, and, in some instances, can present an existential threat.

³ Cook groups are online communities, typically hosted on platforms like Discord or Telegram, that provide their members with tools, information, and support to purchase limited-edition or high-demand products—often for the purpose of reselling.

CASE SNAPSHOT | 23andMe

In June 2025, a joint investigation by the UK and Canadian Privacy Commissioners found that the 2023 breach of genetic testing provider, 23andMe, was the result of a lengthy credential stuffing attack. Despite the attack being facilitated by previous breaches outside of 23andMe's control, its failure to have controls to prevent, detect and mitigate credential stuffing was ultimately found to be a key cause of the attack. The attack allowed the adversary to access and download personal data from thousands of customer accounts, impacting almost 7 million customers. The adversary used stolen login details from other websites impacted by **previous data breaches** to find matches with 23andMe accounts.

23andMe filed for bankruptcy in March 2025 and announced plans to sell assets following years of financial struggles. On top of this, the UK Information Commissioner's Office fined 23andMe USD \$3.12 million.

- In June, there was significant media attention on an alleged data breach impacting 16 billion passwords.⁴ The credentials were likely circulating online for some time and collected and repackaged into a database that was exposed to the internet. There is a lack of credible sourcing and evidence to demonstrate this was a new breach, or a breach at all.
- While Kasada did not identify any evidence indicating this was a breach, we note the reporting was a useful reminder of the importance of preventing, detecting and mitigating credential theft.



As predicted in Q1, adversaries have continued to exploit emergent AI agents and capabilities to increase the scale and efficiency of their operations. In Q3, we expect to see ongoing focus on agentic AI, particularly how it can be manipulated to facilitate malicious activities.

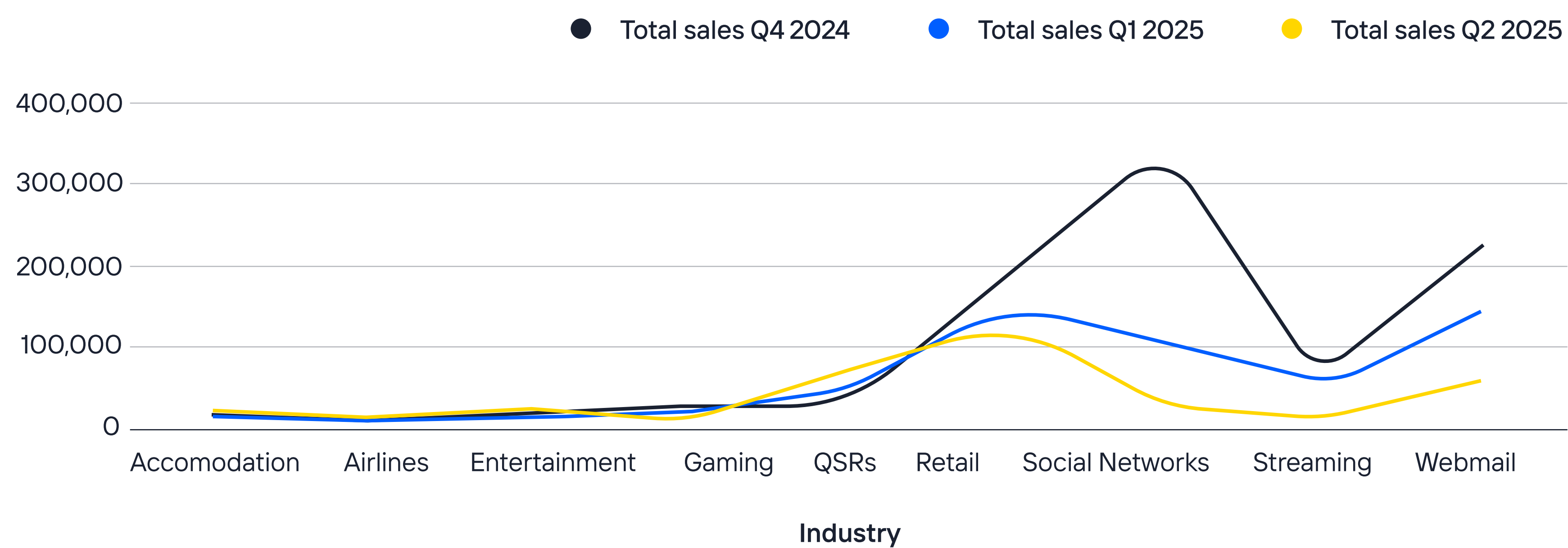
- In Q2, security researchers continued to uncover new techniques to evade security mechanisms and manipulate AI agents. In response, a large focus has been on secure and structured prompting, demonstrating the need to ensure humans have the knowledge to engage with AI effectively and safely.
- Developments included:
 - ✓ The novel Echo Chamber Attack, which uses context poisoning and multi-turn reasoning to guide models into generating harmful content without ever providing a dangerous prompt.⁵
 - ✓ Vulnerabilities introduced by vibe coding, which involves the use of natural language to create software. Researchers emphasise that LLMs are trained to complete, not protect. If security is not explicitly in the prompt, it is likely it won't be a factor in the code.
 - ✓ Adversaries hijacking LLM APIs like Grok and Mixtral with prompts to relaunch WormGPT. WormGPT gained traction in 2023 as a private AI service marketed as a tool for writing malicious software. Its author ceased sales in 2023 in response to unwanted public attention and scrutiny.

⁴ <https://cybernews.com/security/billions-credentials-exposed-infostealers-data-leak/>

⁵ <https://neuraltrust.ai/blog/echo-chamber-context-poisoning-jailbreak>

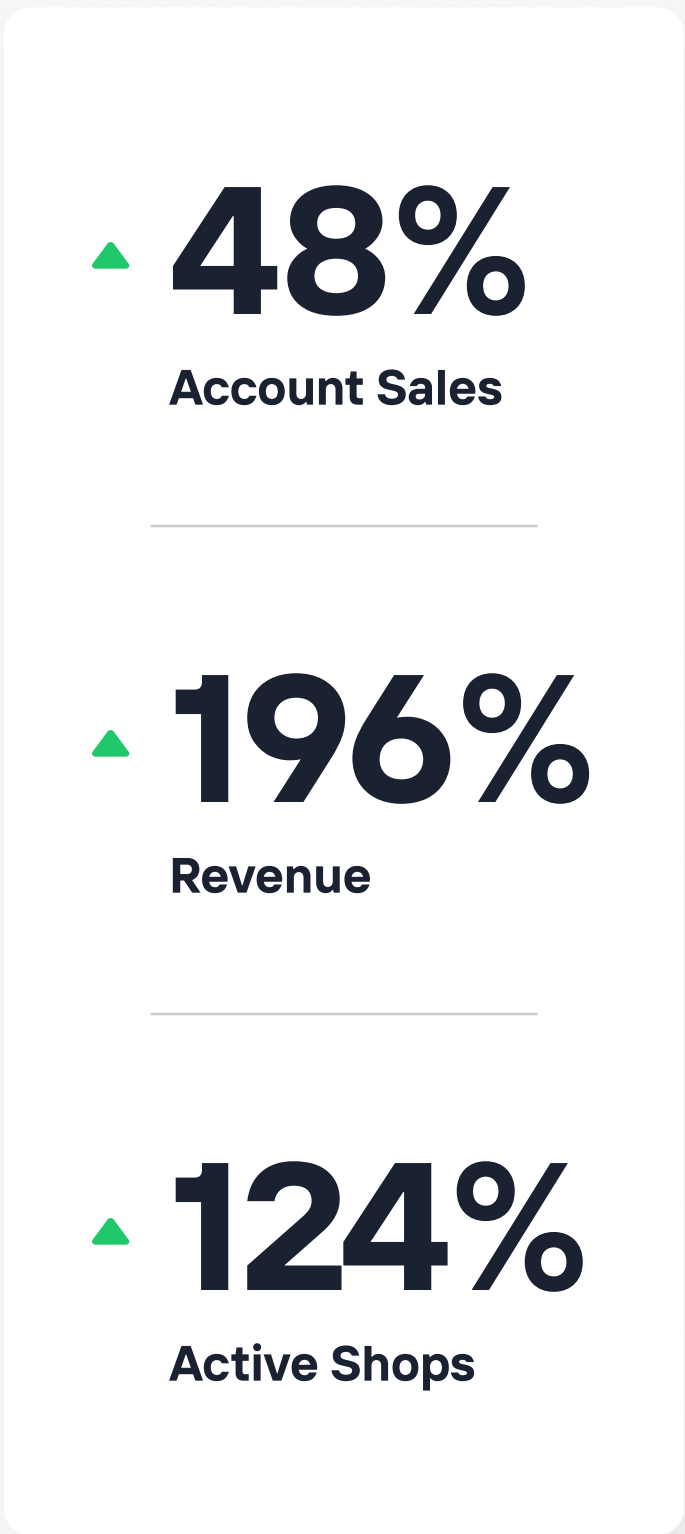
Account Takeover (ATO)

ATO remained the top threat tracked by KasadalQ in Q2. The top three most targeted industries for ATO and credential stuffing were **retail, QSRs and webmail**. Collectively, these industries constituted approximately **72%** of all account sales tracked by KasadalQ in Q2.



ACCOUNT SALES BY INDUSTRY

	Total Sales	Stores Selling Accounts	Maximum Available Stock	Revenue from Sales	Average Account Price
Accommodation	▲ 19,859	▲ 65	▲ 106,187	▲ \$164,170	▲ \$8.27
Airlines	▲ 12,252	▲ 81	▲ 117,420	▲ \$273,938	▼ \$22.36
Entertainment & Ticketing	▲ 20,310	▲ 112	▼ 99,704	▼ \$63,898	▼ \$3.15
Gaming	▼ 6,413	▲ 338	▲ 408,247	▼ \$7,869	▲ \$1.23
Quick Service Restaurants	▲ 68,980	▲ 232	▲ 487,392	▲ \$151,377	▼ \$2.19
Retail	▼ 115,993	▲ 564	▲ 985,942	▲ \$402,818	▲ \$3.47
Social Networks	▼ 26,292	▲ 376	▼ 291,482	▼ \$15,319	▲ \$0.58
Streaming	▼ 10,033	▲ 777	▲ 608,785	▼ \$111,455	▼ \$1.59
Webmail	▼ 57,176	▲ 171	▲ 302,322	▲ \$73,889	▲ \$1.29

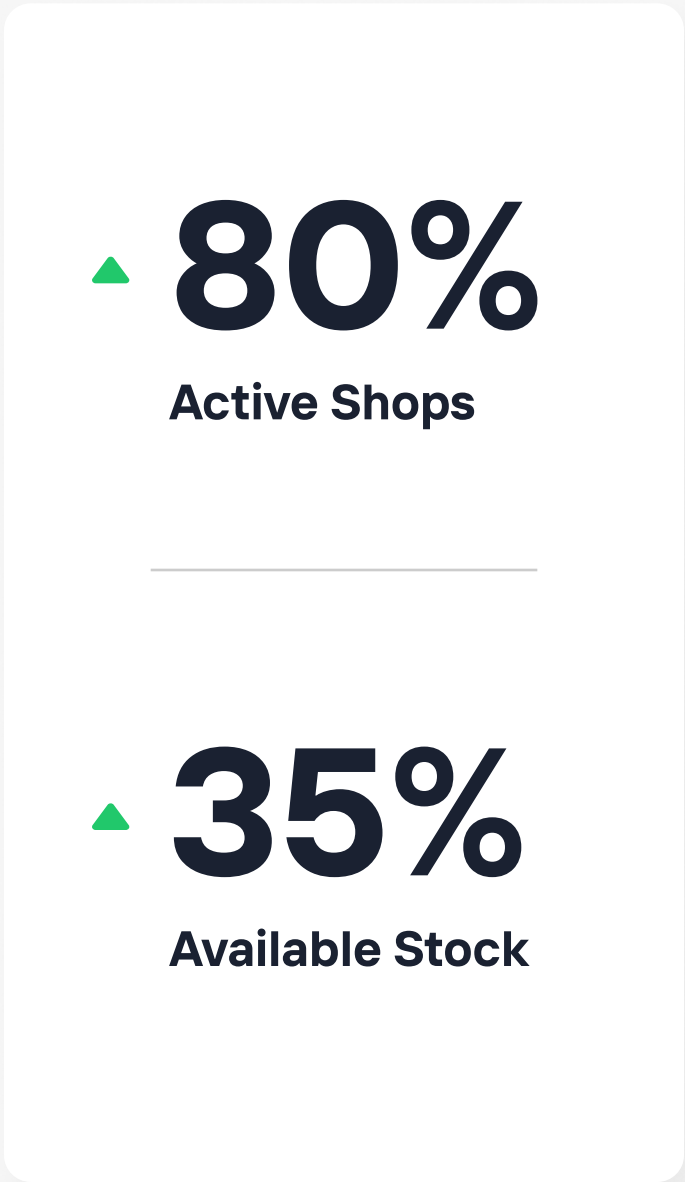


ACCOMMODATION

The accommodation sector experienced remarkable growth in Q2 2025, with account sales surging **48.2%** to 19,859 units compared to 13,403 in Q1. This increase coincided with the spring travel season, as consumers began planning summer vacations and business travel resumed.

Revenue growth was even more dramatic, jumping **196.3%** to \$164,170, driven by a near-doubling of average account prices from \$4.15 to \$8.27. This price increase reflects the growing value of loyalty program points and rewards as travel demand increased throughout the quarter. Premium hotel chain accounts commanded particularly high prices due to their transferable points and elite status benefits.

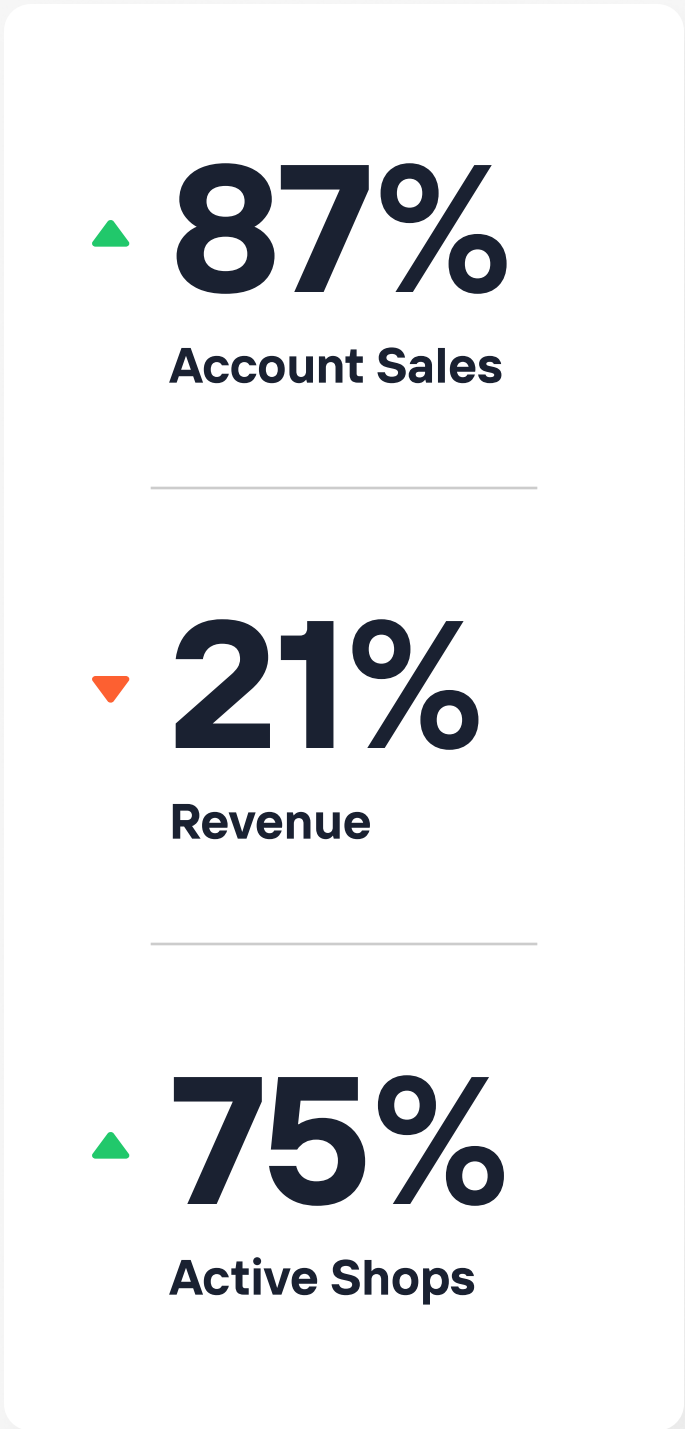
The number of shops selling accommodation accounts **more than doubled** from 29 to 65, indicating increased criminal interest in this sector. Available stock rose **31.6%** to 106,187 accounts, suggesting improved supply chains and potentially new data breaches in the hospitality industry.



AIRLINES

Airline account sales maintained strong momentum from Q1, increasing **33.1%** to 12,252 accounts sold. However, average prices declined **24.9%** from \$29.79 to \$22.36, resulting in minimal revenue growth of just **0.6%** to \$273,938. This price reduction may indicate market saturation or increased competition among sellers.

Despite the price decline, airlines remained the second-highest revenue generator after retail, highlighting the continued value of frequent flyer programs and accumulated miles. The **80%** increase in active shops from 45 to 81 suggests expanded criminal interest, while stock levels grew **35.3%** to 117,420 accounts.



ENTERTAINMENT & TICKETING

The entertainment & ticketing sector saw explosive growth in account sales, jumping **87.2%** to 20,310 units. This surge aligns with the spring concert season and major sporting events, as criminals targeted accounts with valuable pre-purchased tickets.

However, average prices plummeted **57.6%** from \$7.43 to \$3.15, causing overall revenue to decline **20.8%** to \$63,898. This price reduction suggests a shift toward lower-value entertainment accounts, possibly cinema memberships and regular event tickets, rather than the premium concert and sporting event tickets that commanded higher prices in Q1.

The number of active shops increased **75%** to 112, while stock levels decreased **9.3%** to 99,704, indicating strong demand that outpaced supply.

▼ **67%**
Account Sales

▼ **54%**
Revenue

▲ **69%**
Available Stock

GAMING

Gaming accounts experienced the most significant decline in sales, dropping **66.5%** to just 6,413 units sold. This dramatic decrease occurred despite a **20.7%** increase in active shops and a **68.5%** rise in available stock to 408,247 accounts.

The substantial stock increase combined with reduced sales suggests either decreased demand or potential quality issues with available accounts. Average prices did increase **35.2%** to \$1.23, but this was insufficient to prevent a **54.7%** revenue decline to \$7,869.

The gaming sector's poor performance may reflect seasonal patterns, with fewer major game releases during Q2 compared to the holiday season that typically drives gaming account demand.

▲ **42%**
Account Sales

▲ **93%**
Active Shops

▲ **38%**
Active Shops

QUICK SERVICE RESTAURANTS (QSR)

QSR accounts continued their strong performance from Q1, with sales increasing **42.8%** to 68,980 units. The sector benefited from expanded loyalty programs and increased consumer adoption of mobile ordering platforms.

Despite a **28.4%** decrease in average prices from \$3.06 to \$2.19, revenue growth remained modest at **2.3%** to \$151,377.

The price decline may indicate market maturation or increased competition among sellers.

The number of shops nearly doubled from 120 to 232, while stock levels increased **32.7%** to 487,392 accounts. This expansion suggests continued criminal interest in QSR loyalty programs, which are perceived as low-risk targets with consistent redemption opportunities.

▼ **16%**
Account Sales

▲ **18%**
Revenue

▲ **57%**
Active Shops

RETAIL

Retail maintained its position as the **highest-revenue sector** despite a **16%** decline in sales volume to 115,993 accounts.

The sector's resilience was demonstrated by an **18.3%** revenue increase to \$402,818, driven by a **40.5%** rise in average prices from \$2.47 to \$3.47.

This price increase suggests criminals are focusing on higher-value retail accounts, potentially those with stored payment methods, loyalty points, or premium membership benefits. The **56.7%** increase in active shops from 360 to 564 indicates continued strong criminal interest in retail targets.

Stock levels grew modestly by **7%** to 985,942 accounts, suggesting steady supply despite reduced sales volumes.

▼ **75%**
Account Sales

▼ **63%**
Revenue

SOCIAL NETWORKS

Social network accounts suffered severe declines in Q2, with sales dropping **74.8%** to 26,292 units. This dramatic decrease occurred alongside a **57%** reduction in available stock to 291,482 accounts, suggesting continued impact from platform security improvements and law enforcement actions.

Revenue declined **62.9%** to \$15,319, though average prices increased **28.9%** to \$0.58. The price increase may reflect scarcity value as supply constraints drove up remaining account prices. The number of shops increased modestly by **13.6%** to 376, indicating some sellers remained active despite challenging market conditions.

▼ **86%**
Account Sales

▲ **64%**
Available Stock

STREAMING

Streaming accounts experienced the most dramatic decline in sales, plummeting **86.2%** to just 10,033 units. This massive decrease occurred despite a **64.1%** increase in available stock to 608,785 accounts and a **37.5%** rise in active shops to 777.

The combination of reduced sales and increased supply suggests fundamental changes in the streaming account market, possibly due to improved security measures by streaming platforms or changes in user behavior. Revenue fell **38.3%** to \$111,455, while average prices dropped **38.1%** to \$1.59.

▼ **61%**
Account Sales

▲ **37%**
Available Stock

WEBMAIL

Webmail account sales declined **61.1%** to 57,176 units, though this sector showed resilience in other metrics. Revenue increased **33.3%** to \$73,889, driven by a remarkable **239.5%** price increase from \$0.38 to \$1.29 per account.

This dramatic price increase suggests growing scarcity of quality webmail accounts or increased demand for accounts with specific characteristics. Stock levels increased **37.4%** to 302,322 accounts, while active shops grew **31.5%** to 171. The webmail sector's price appreciation may indicate its evolving role as a gateway to more valuable accounts, justifying higher prices for quality compromised email accounts.

Spotlight On: Scraping

Why Scraping Matters to You

Unwanted scraping is costing you, and those costs are rising.

Unwanted scraping profits off your infrastructure, while you foot the bill. Particularly for businesses with dynamic pricing models and/or cloud-native environments, **you may be paying more to serve the adversaries exploiting and profiting off your data and infrastructure.**

WHAT ONCE WAS NOISE IS NOW A THREAT

While once considered a nuisance, unwanted scraping can now present **strategic and existential threats** to businesses thanks to the likes of [generative AI](#). Advancements in AI have evolved the threat into something far more tangible and pressing, and this threat is set to grow. Businesses can no longer afford to not care about scraping. The scale, speed and frequency of modern scraping requires all businesses to have a position on scraping, identifying the line between permitted scraping and unwanted scraping.

SCRAPED AND BILLED

Unwanted scraping is driving up your infrastructure costs. Scrapers simulate thousands of users accessing pages, images and API data. Cloud providers charge for data egress, so more scraping means higher bandwidth bills. APIs often incur per-request charges. Where scraping slows down site performance, businesses may need to over-provision infrastructure to meet business requirements.

Cloud-native environments in particular face significant impacts from scraping. In cloud-native environments, scrapers trigger auto-scaling. Cloud infrastructure automatically scales to handle the load, meaning more CPU, memory and resource usage. These are generally billed by the second or per request. Bandwidth overages result in higher data egress charges. This means that as the scale and frequency of scraping continues to increase, businesses are **paying more in infrastructure costs for adversaries to profit off their hard work and data.**

DYNAMIC PRICING AT RISK

Businesses with dynamic pricing models (like airlines, eCommerce and ridesharing) **face significant losses from scraping**, particularly with the scale and speed of [AI scraping](#). For these businesses, scraping can **directly interfere** with the mechanisms relied on to optimize revenue and undermine business strategy, as demonstrated in the matrix below.

Impacts/ Scraping Methods	Price prediction Continuous price monitoring to detect patterns and changes	Price changes Simulating interest by flooding product or services page	Signal theft Analyzing pricing shifts to adjust own	Copying Scraping prices to instantly match or undercut	False demand signals Repeatedly querying items or regions
Loss of competitive edge	✗		✗	✗	
Infrastructure costs	✗	✗	✗	✗	✗
Price surges or instability from skewed analytics		✗			✗
Site performance issues		✗			✗
Business strategy disruption	✗	✗	✗	✗	✗

Profiling Scraping

WHAT

Scraping is the use of automation tools to extract valuable data or content from a website, such as pricing information, product details or proprietary content. Scraping is a **grey area**. Businesses must consider if scraping is allowed or unwanted based on a range of factors.

Scraping can be periodic or continual, depending on the intended use case for the acquired information. Continual scraping may monitor for a change on a website, like the release of a limited edition item, and places a sustained load on servers. Periodic scraping may be used to gather information such as daily listings on a real estate website, with short bursts of load on servers.

Scraping alone is not illegal and has both positive, neutral and negative use cases. How scraping is characterised and whether it is blocked is a choice for each business with regard to how it is conducted, the intent behind it and potential harms.

The below provides a non-exhaustive guide for the kinds of factors which can help identify unwanted scraping.



Ethical and transparent

Legally and with respect to terms of service

With website user's permission

Respectful request rate

Minimal server load, non-disruptive

Targets publicly accessible, non-sensitive data

Fully respects robots.txt directives

Gives credit or links back to original content

Neutral or positive net impact on business



Unethical, exploitative or deceptive

Violates legal agreements or terms of service

Without website user's permission

High volume requests which overload servers

Causes downtime, lags or exhausts resources

Targets private, copyrighted data

Ignores or bypasses robots.txt

Presents scraped content as original

Negative net impact on business

WHY

The primary objectives of unwanted scraping include:

- ✓ Data monetization by selling scraped data to competitors or aggregators.
- ✓ Price optimization or price/stock arbitrage by leveraging proprietary data.
- ✓ Obtaining competitive intelligence by using data to undercut pricing, improve their own offerings or gain another market advantage.

WHO

Scraping is typically conducted by specialised scraping providers, AI firms, freelancers, data brokers, researchers and enthusiasts. Scraping is typically consumed by commercial entities and B2B competitors. The adversaries and tooling used by them for scraping activities are diverse. This presents challenges for the detection and mitigation of this type of traffic. The ability to closely mimic genuine human behaviour exacerbates these challenges.



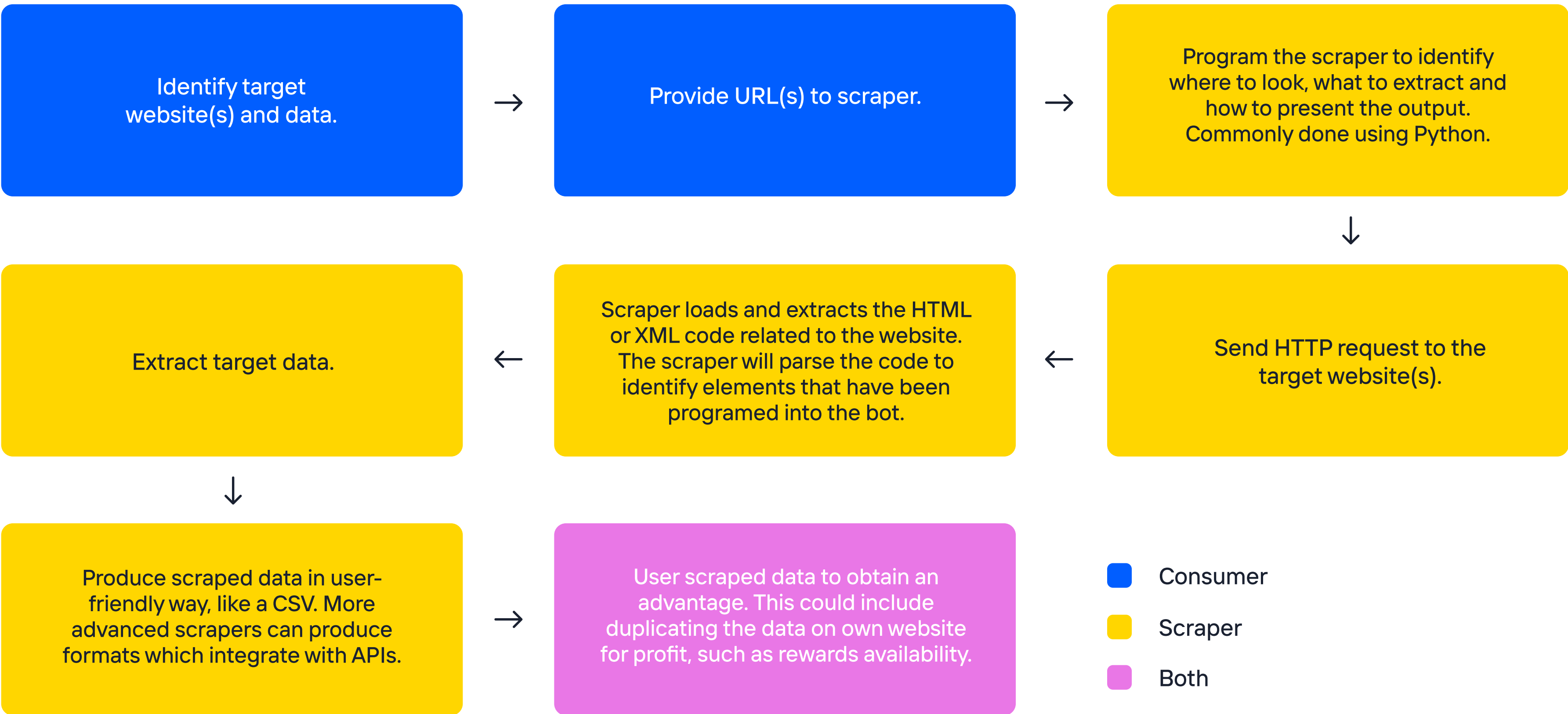
There is an important distinction to make between scrapers and users. **Scrapers** provide the infrastructure and tools that actually scrape the data and often operate in a commercial model. Users are individuals who outsource scraping and data harvesting to a third party. See [Adversary Insights](#) for different use cases.

Overall, we assess adversaries behind scraping have a moderate capability level based on the considerations below.

Technical capability	Access to resources	Scale	Persistence
Moderate	Moderate	High	High

HOW

Scraping gathers a website's HTML code and the visible data it presents, as demonstrated in the attack chain below.



Common methods used by scrapers to extract data include:

- ✓ **Automated bots** which simulate human behavior.
- ✓ **Open-source tools**, like GitHub, host repositories with pre-built scraping scripts which compare product prices across retailers.
- ✓ **Solver services**, like CAPTCHA-solvers and anti-bot solvers, are used to bypass anti-bot defenses.
- ✓ **Residential proxies** rotate IP addresses to evade detection and mimic legitimate users.
- ✓ **Commercial scrapers** take the complexity out for a user by implementing defense bypasses and solver services to obtain data from the website. See [USE CASE: SCRAPING-AS-A-SERVICE](#).
- ✓ **Human emulation** follows established human patterns to blend into legitimate, human traffic.

Adversary Insights

Scraping sits at the centre of how data is collected, repackaged and repurposed at scale. While each use case might appear distinct, they often overlap and feed into one another. What emerges is not a set of isolated tools, but a dynamic and interconnected scraping ecosystem where **data is a tradable asset**. This means that scraping is not a background risk.

Scraping is a strategic issue touching on competitive advantage, data ownership, infrastructure costs and brand integrity. Understanding how these use cases overlap is essential for managing exposure and deciding when to block, license or capitalize on your own data.

USE CASE | AWARDS AGGREGATION

Awards aggregation platforms provide visibility in award availability and comparisons, particularly for travel-related programs. These platforms use automated tools to scrape data on award availability, such as award flight seats or hotel bookings, from the websites of the originating business. This involves the awards aggregator sending a **high volume of requests** to the originating business' endpoints to extract information and present in a user-friendly way on their own platform.

HOW THIS COULD IMPACT YOU:

- ▶ The high volume of automated requests can strain systems, potentially leading to **performance issues or downtime**. Kasada has seen systems taken offline due to activity correlated with scraping by aggregators.
- ▶ Businesses may face challenges in **maintaining accurate and secure data** due to the sophisticated methods used by some aggregators to bypass anti-bot measures.
- ▶ Scraping activities can expose disclosure discrepancies, creating complications for businesses.
- ▶ These issues can create downstream consequences for **business reputation** and **customer experience**.

CASE SNAPSHOT | AWARDS AGGREGATION PLATFORM

LOYALTYPRO⁶ is an aggregator platform that provides visibility into awards availability. LOYALTYPRO operates as a legitimate business, managed by a single person. It is widely used by frequent flyers and travel enthusiasts to find and book award flights using points or miles.

The platform aggregates data from various legitimate websites and presents it in a structured, user-friendly interface. Through its website and API, it offers customized searching across several websites, as well as bulk availability of awards for one website.

⁶ Pseudonym applied.

CASE SNAPSHOT | AWARDS AGGREGATION PLATFORM (CONT.)

What differentiates LOYALTYPRO from other awards aggregating services is that it has kept up with shifts to [dynamic pricing](#) models. It automatically displays dynamically-priced awards only if they have reasonable pricing and hides those which are undesirable or unreasonable. These factors make it appealing to users looking for a streamlined experience. These types of business models have faced legal challenges in the past, with owners of scraped websites claiming that the harvesting and use of their data is unauthorised and unlawful (see [USE CASE: MARKET COMPETITION](#)).

Industries targeted	Airline, hotel
Scraper motivation	Financial gain
User motivation	Financial savings
Cost	Free, < USD \$10 for subscribers
Userbase	500,000+
Revenue	> USD \$6 million

USE CASE | SCRAPING-AS-A-SERVICE

Both commercial entities and individuals provide scraping services “as-a-service”, enabling users to collect data without needing to build and maintain their own scraping infrastructure. Unlike casual or academic scrapers, scraping-as-a-service providers are **strategically motivated** to extract and **monetize** data belonging to other organizations.

HOW THIS COULD
IMPACT YOU:

- ▶ Scraping-as-a-service is designed to make the business or developer money, meaning there is significant investment in ensuring tooling is successful.
- ▶ In addition to imposing **infrastructure costs** from a high volume of requests, scraping-as-a-service providers are **persistent**. They prioritize defense evasion, IP rotation and mimicking real user behavior.
- ▶ Adding to these challenges, scraping is outsourced, presenting businesses with a moving target. The true client is often hidden behind layers of intermediaries, making enforcement difficult and driving up operational and legal costs.

COMMERCIAL SCRAPERS

Commercial scrapers are cloud-based, software-as-a-service providers that offer **on-demand, low effort** web scraping via an application programming interface (API). Users only need to type a few lines of code to start scraping a target, while the commercial scraping provider will attempt to bypass any detected anti-bot defenses.

We have provided a side-by-side comparison of a sample of commercial scrapers to demonstrate the variability of services available.

Scraper	Price/1000 requests (USD)	Primary use case	Request timings	Userbase	IP Pool
A	\$2.70	API-based scraping with proxy rotation and JavaScript rendering	39–43 seconds between requests	200,000+	190 million+
B	\$6.00	Full-service data collection platform with residential proxies, scraping APIs, and SDK-based traffic routing	~21–49 seconds between requests	15,000+	150 million+

Some commercial scrapers offer specific APIs for industries or specific organizations, such as for Quick Service Restaurants and food delivery services, providing highly accessible bulk data per industry. This is of particular interest to market competitors. See [USE CASE: MARKET COMPETITION](#).

CASE SNAPSHOT | MONETIZATION-AS-A-SERVICE

Developments in [AI](#) have driven the emergence of new “monetization-as-a-service” companies, advertised as open-source alternatives to advertisements. Monetization-as-a-service companies build specialized platforms, provide Software Development Kits (SDKs) or offer APIs that clients can integrate into their own products. A recent example of this is the Mellowtel library which pays extension developers in exchange for the “unused bandwidth” of users who have an extension installed.⁷

- Mellowtel is a JavaScript library and service that aims to provide a way for developers, particularly those creating browser extensions or free mobile apps, to monetize their products without relying on traditional advertising or selling user data. According to Mellowtel, its core concept is that users can voluntarily share a small portion of their unused internet bandwidth to support their favorite developers.⁸ Mellowtel indicates that verified institutions, trusted partners and companies pay Mellowtel to access the internet via this network.

⁷ <https://secureannex.com/blog/mellow-drama/>

⁸ <https://www.mellowtel.com/blog/responding-to-ars-technica-and-mellow-drama-article>.

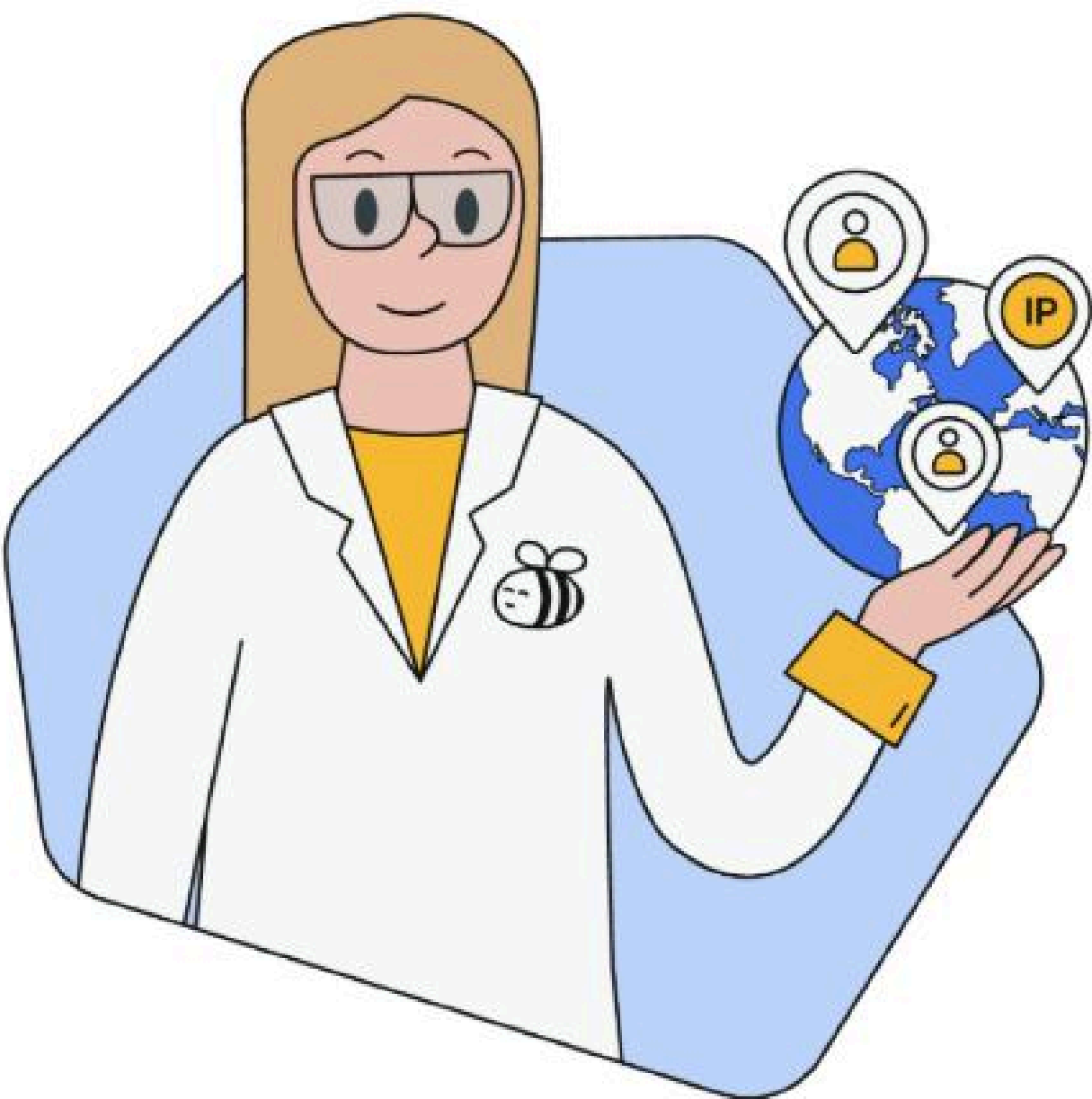
CASE SNAPSHOT | MONETIZATION-AS-A-SERVICE (CONT.)

- Security researchers have alleged that Mellowtel effectively turns users' browsers into unintentional nodes in a large-scale web scraping botnet.⁹ Over a million devices are reported to have installed extensions containing the Mellowtel library. Research has suggested a connection between Mellowtel and a company advertising a "cost-effective web scraping API, Olostep, which is designed to circumvent bot detection.

Mellowtel is just one example in a growing ecosystem of monetization providers, broken down below through the lens of scrapers and users on the following page.

Scrapers: Users opt-in or are opted-in by default to platforms or services to share bandwidth and generate passive income. The shared bandwidth is sold to businesses, primarily for web scraping or market research.

- An example is Honeygain, a bandwidth-sharing application that allows users to earn a small amount of passive income by sharing their unused internet bandwidth.¹⁰ Honeygain state they sell this unused bandwidth to scientists and companies



How does Honeygain work?

When you run Honeygain, your unused internet bandwidth makes you money. Scientists and companies use your unused bandwidth to perform tasks like ad verification and more. Honeygain pays you for helping them make the internet a better place!

[Learn more](#)

Users: Residential proxy network providers buy the bandwidth from networks like Mellowtel and Honeygain and resell access to their own clients for a range of purposes, including web scraping.

- An example is Bright Data, which offers a pool of over 150 million residential IPs, data collection tools and proxy types.¹¹ Bright Data embeds its brightSDK into freemium apps and games, transforming end-user devices into residential proxies without their explicit awareness. Bright Data states it uses these devices for at-scale web scraping activities.¹² App developers are also incentivised to integrate the brightSDK, gaining access to an alternative revenue stream outside of traditional advertising.

⁹ <https://secureannex.com/blog/mellow-drama/>

¹⁰ <https://www.honeygain.com/what-is-honeygain/>

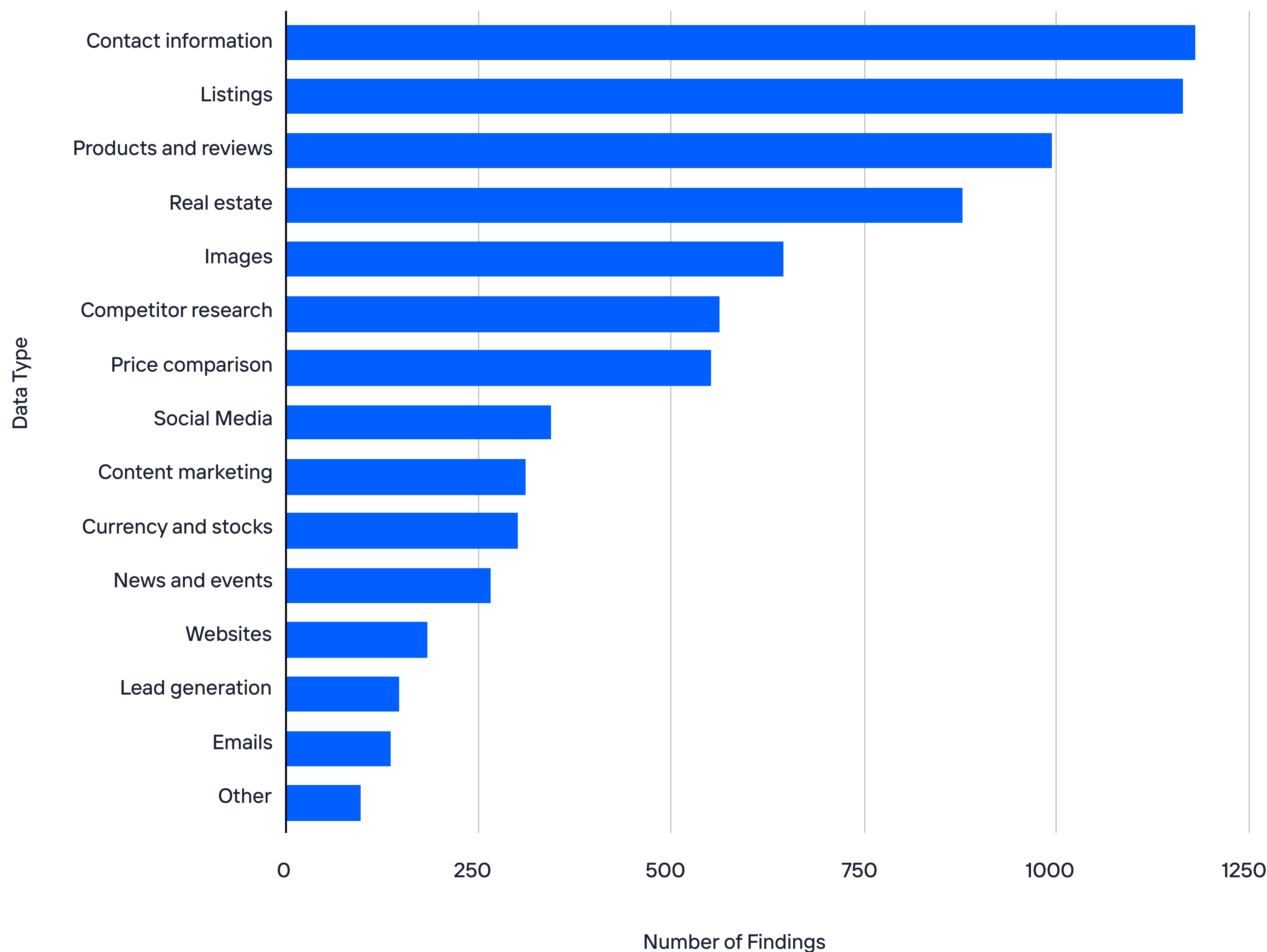
¹¹ <https://docs.brightdata.com/proxy-networks/introduction>

¹² <https://www.youtube.com/watch?v=Sq-otzS2nl8>

FREELANCING

Freelancers advertise their scraping services on platforms like Upwork and Fiverr, or respond to scraping requests from organizations.

From a point-in-time search, KasadaIQ was able to find **11,872** advertisements for scraping on a freelancing platform ranging from USD \$6 to USD \$1,064. The most advertised data types for scraping were **contact information, listings, products and reviews, real estate, images, competitor research** and **price comparison**.



When outsourcing to a freelancer, a user is relying on the individual to uphold the Terms of Service of the websites it is scraping and align with relevant legal requirements, like copyright laws, privacy regulations and anti-hacking laws. This emphasises the potential legal considerations and costs for organizations that are targeted in scraping by freelancers.

USE CASE | MARKET COMPETITION

Competitors may scrape pricing or inventory data in real time to inform their own business decisions, creating a **race to the bottom** or **distorting the market**.

HOW THIS COULD IMPACT YOU:

▶

Competitor-led scraping can result in the resale of your pricing, content or customer data. This can **undermine** your direct sales model, **misrepresent** your brand and force you to invest in enforcement and infrastructure.

▶

Competitors use scraping to directly profit from rival data and operate in a competing commercial ecosystem. Without proactive measures, this kind of scraping can **erode your competitive edge and consumer trust**.

▶

Legal challenge is timely, resource-intensive, costly and often has limited operational impact.

▶

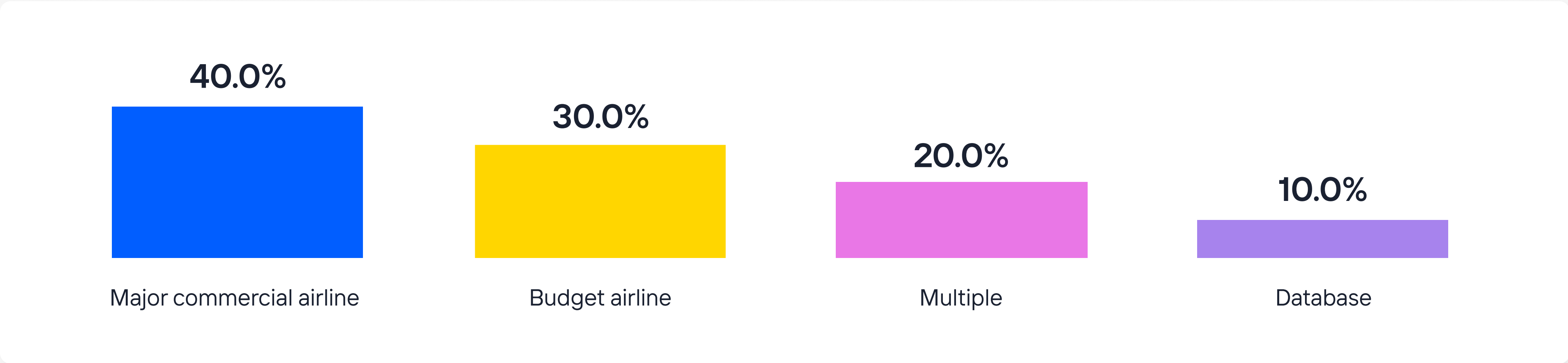
Collectively, these burdens can **divert resources** away from core business priorities.

Kasada has observed high volumes of automated requests straining systems, leading to performance issues or downtime. This included a North American company being taken offline at one point due to activity correlated with unauthorised scraping by a market competitor.

Organizations may outsource to [scraping-as-a-service providers](#) or use open source repositories, like GitHub, to extract data from websites. KasadaIQ monitors open-source code repositories for tooling that affects our customers. These tools are freely available for anyone to use or modify and are frequently updated to bypass anti-bot protections.

From a point-in-time search of GitHub, KasadaIQ was able to identify **726** real estate scrapers, **114** retail scrapers and **81** airline scrapers in public repositories. Looking at a sample of prominent **airline scrapers**, we observed:

- Many of the scrapers were designed for multiple airlines/sites and can be adapted to any country.
- **90%** were written in Python, making them relatively simple, readable and versatile for users.
- **70%** were specific to major or budget airlines in North America and Europe.



CASE SNAPSHOT | SCRAPING OF RYANAIR FLIGHT INFORMATION

Budget European airline, Ryanair, has been in ongoing legal battles with online travel agencies (OTAs) over scraping of Ryanair’s website without permission. For over a decade, Ryanair has claimed various OTAs have hired commercial scrapers or operated their own to display and sell Ryanair flight information, using Ryanair’s proprietary fare and schedule data for profit.

Ryanair’s claims	Scraper's claim
Unauthorised scraping violated terms of use.	Ryanair’s terms of use did not apply because the scraper was not a “user” in a legal sense, no explicit contract was formed between Ryanair, the OTA or scraper, and the Terms of Service were not enforceable against automated systems or third party bots .
Scraping undermined its direct-to-consumer business model.	They were facilitating access to Ryanair flights and charging a service fee.
OTAs misrepresented Ryanair’s brand and misled customers.	Aggregating data promoted price transparency and enhanced market competition. They were offering a service to consumers.

OTAs often outsourced the scraping to specialised third-party scraping firms or used complex bot networks to disguise traffic and bypass Ryanair’s bot detection systems.

Courts in multiple European jurisdictions sided with Ryanair, reinforcing that scraping and reselling flight data without permission **was unlawful**, finding that its Terms of Use and contractual terms were legally valid and enforceable. In several cases, the Court found that the OTA violated Ryanair’s intelligence property and terms of use. Ryanair has banned over 20 OTAs and data scrapers from accessing its content, launching a “Verified OTA” program in 2023 which published a list of unauthorised resellers.

While Ryanair has not disclosed an exact total cost for its long-running legal and operational battles against OTAs and commercial scrapers, the financial and operational impacts have likely been significant. A prime example of this is Ryanair’s four-year court case seeking injunctive relief against Booking[.]com. Ryanair alleged Booking[.]com and affiliated brands were unlawfully scraping its website. While a jury initially sided with Ryanair, the decision was overturned in March 2025, finding that Ryanair had failed to prove it had incurred the minimum amount of damages required under the *Computer Fraud and Abuse Act*. In addition to the financial losses from scraping claimed by Ryanair, it also was impacted by court fees, legal fees, resource diversion and operational measures over a four year period.

USE CASE | TRAINING AI MODELS

AI scrapers are automated bots that extract data from websites to train LLMs like those from OpenAI, Anthropic, ByteDance (Bytespider), and others. Once ingested, the LLMs can reproduce factual, stylistic or thematic elements from the original sources, often verbatim and without attribution.

HOW THIS COULD IMPACT YOU:

▶

AI scrapers place **heavy strain** on servers, bandwidth and APIs, without acknowledgement or compensation, leading to **higher cloud and CDN bills** and **degradation of service** for real users.

▶

AI companies gain value from scraped business data without attribution or licensing which can **undermine IP monetization strategies**.

▶

Competitive edge is at risk when unique content becomes part of a public AI model, leading to **diminished ROI in content creation**.

▶

AI trained on scraped data can misrepresent your brand, amplify outdated or inaccurate information or undermine SEO investments.

▶

The indiscriminate nature of AI scraping, particularly involving ethically problematic data or repurposed content, can introduce **additional legal and strategic concerns**. This could include regulatory action, brand erosion, social backlash or investor concern around **ESG misalignment**.

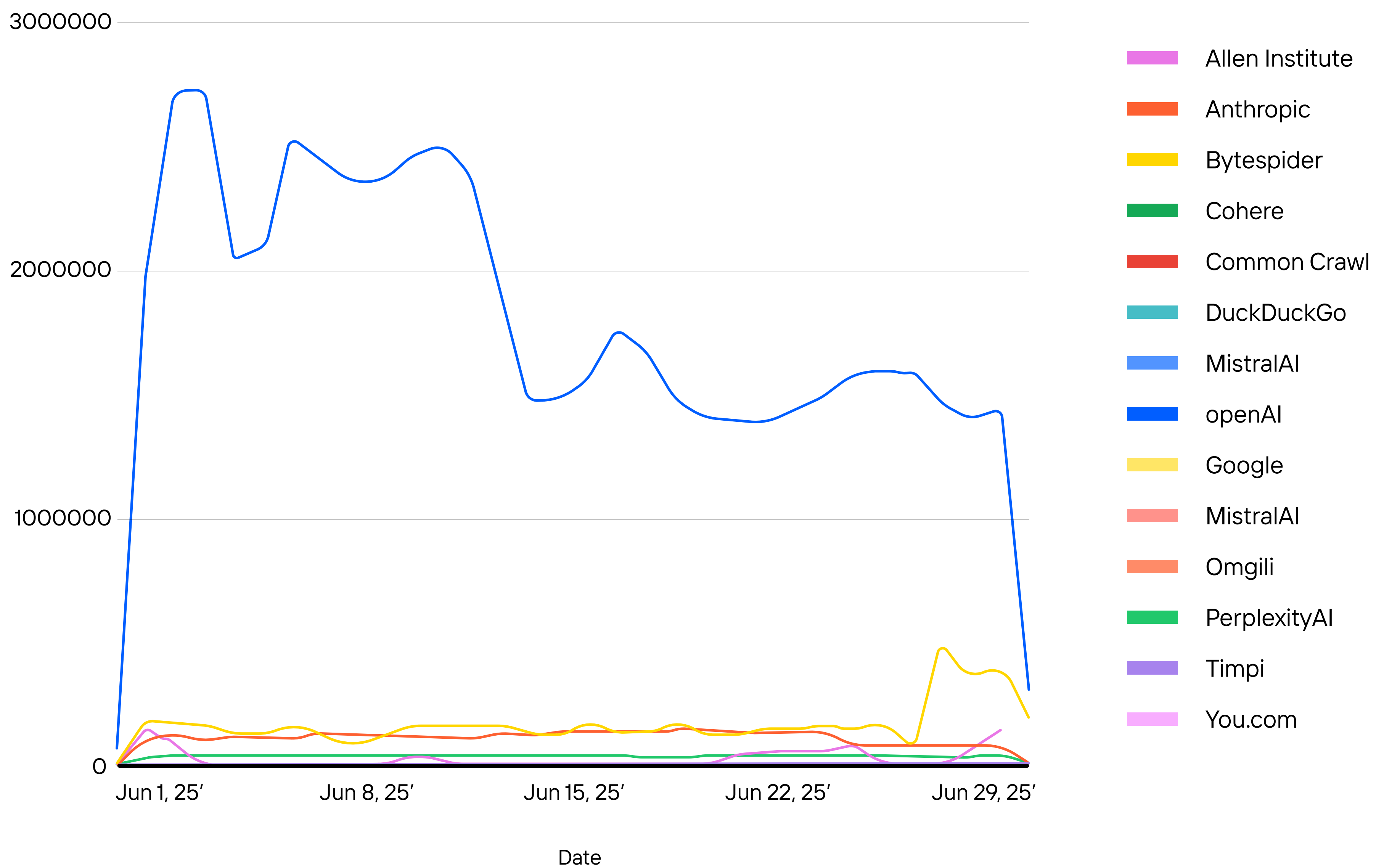
▶

Ethical data sourcing is becoming a **brand issue**, not just a tech one.

AN INFLECTION POINT

AI has forced businesses to reach an inflection point when it comes to scraping. AI has significantly amplified the scale, speed and impact of scraping. While once a nuisance, AI scraping can now present **strategic and existential threats** to businesses.

Between April and June 2025, Kasada observed **over 120 million requests by AI scraper user agents to our customers**. Of these, the majority originated from the OpenAI user agent. In June alone, Kasada observed over **56 million** OpenAI requests, vastly exceeding other self-identified AI user agents. This is a change of behavior from July 2024, where we observed more than 25 Bytespider requests per OpenAI request.¹³



This shift aligns with broader trends in the market, with reported increases in ChatGPT’s web traffic share and challenges by way of regulatory scrutiny of China-based ByteDance’s practices.

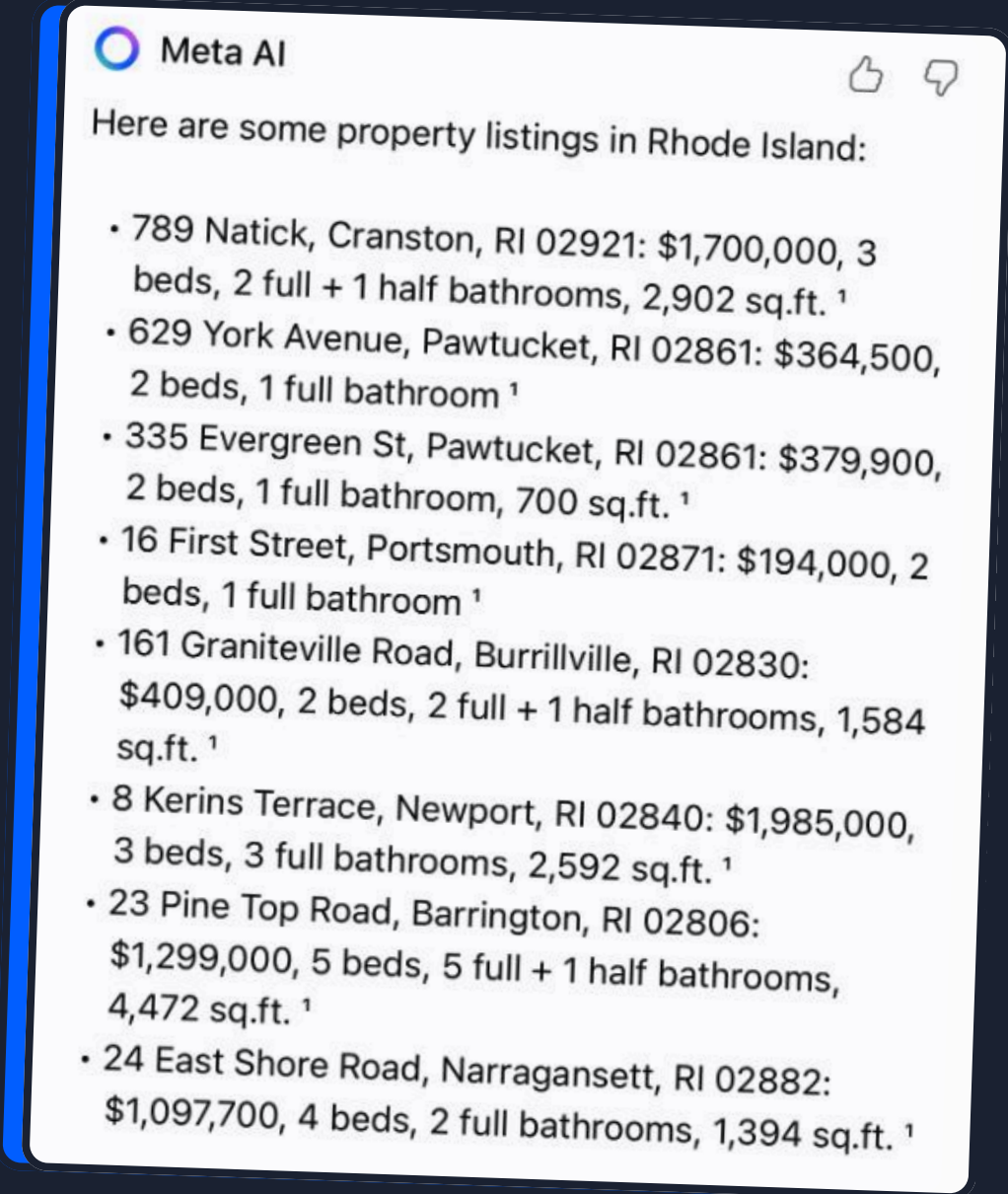
Based on Kasada’s observations, AI user agents make the highest volume of requests against the **retail and real estate** industries.

CASE SNAPSHOT | META AI SCRAPING REAL ESTATE WEBSITES

Facebook’s AI Meta platform, which powers tools like MetaAI, uses generative AI to provide users with answers to queries, often including data scraped from external websites. Kasada received the following results from searching “Show me property listings in Rhode Island” using the Meta AI search function on Facebook.

CASE SNAPSHOT | META AI SCRAPING REAL ESTATE WEBSITES

Meta AI returns results derived from real estate websites, which in this case was a large United States real estate organization. This raises questions about whether such scraping activities comply with the terms and conditions of the targeted websites. Kasada’s previous investigations into Meta AI identified that traffic is often indistinguishable from legitimate requests, as it employs user agents and Autonomous System Organization (ASO)¹⁴ identifiers attributed to Facebook. There is little to differentiate Facebook spider traffic between a Meta AI request and a request sent by other means, such as sharing a link within Facebook.



LLMs need large amounts of data, making scraping a default part of AI development. To train AI models, developers scrape large portions of the internet, including proprietary content, product data and user-generated content, without permission. AI scrapers do not just copy the data, they learn from and repurpose scraped content, making it hard to trace and monetize. This means as a business, your data is effectively powering AI tools **without attribution or revenue**, while also losing traffic, ad revenue and customer acquisition pipelines.

When thinking about whether AI scraping is unwanted scraping in a specific organizational context, key cost considerations include:

Bandwidth Overage

Bot Mitigation Tools

Engineering Time

Lost SEO Traffic

Legal/Compliance Burden

Content ROI Loss

NO HONOR AMONGST AI SCRAPERS

AI scrapers have not been playing by the same rules when it comes to digital boundaries. AI developers are increasingly using aggressive scraping tools to collect high volumes of data, often without consent and despite clear disallow rules in robots.txt.

¹⁴ An ASO is a large, independent network that manages its own routes and traffic rules within the internet.

Robots.txt is effectively a **do not enter** sign for bots on a website. It is a plain text file at the root of the website which tells bots and scrapers which part of the site they are allowed or disallowed from accessing. robots.txt was developed in the 1990s in response to the indiscriminate crawling of websites which would overload servers and expose private parts of websites. Now, it is used by ethical bots to respect digital boundaries. It is not a security mechanism, it just requests compliance.

AI scrapers typically bypass robots.txt restrictions, ignore rate limits, target APIs and endpoints not intended for public use and use distributed IPs or mobile proxies to avoid detection.

AI scrapers are indiscriminate, exposing children and cultural institutions. In June 2024, Human Rights Watch found the LAION-5B dataset contained links to identifiable photos of Australian children. The dataset compiled over 5 billion image-text pairs scraped from the web, including content from nonprofit and archival websites, and was used to train AI models like Stable Diffusion.¹⁵

More recently in June 2025, the United States university-led Galleries, Libraries, Archives and Museums (GLAMs) initiative reported that cultural institutions in North America, Australia and Europe are alarmed by the aggressive scraping of their content.¹⁶ Of 43 cultural institutions surveyed by GLAMs:

- ~91% had experienced a recent increase in traffic.
- ~63% experienced an increase in traffic from AI scrapers.
- Some institutions have observed increased traffic since 2021, while others did not observe their first spike until 2025.
- Some scrapers self-identify, but several take measures to hide their source.
- Robots.txt is not currently effective in preventing scraping.
- The institutions are reluctant to move collections behind login screens, noting it would present a barrier to the larger goal of making collections easily available online.
- The institutions are worried that AI scrapers will unsustainably escalate costs for providing online access to collections.

¹⁵ <https://www.hrw.org/news/2024/07/03/australia-childrens-personal-photos-misused-power-ai-tools>

¹⁶ <https://www.glamelab.org/products/are-ai-bots-knocking-cultural-heritage-offline/>

Looking ahead

Threat Intelligence at Kasada

Kasada uses threat intelligence to drive the identification of new and emerging threats, techniques and tactics used by adversaries. This identification, coupled with broad access and analysis of open and closed data sources, allows Kasada to rapidly implement new mitigations or identify alternate methods to reduce the impact of automated threats, bot attacks, and online fraud.

The people we're up against are dynamic, well-resourced, and technically proficient. This evolving field means that Kasada needs to be at the forefront of cybersecurity to provide subject matter expertise relating to the capabilities, development pipeline, and targets of specific actors or groups against their networks.

The goal of threat intelligence within Kasada is to reduce the efficacy of the threat ecosystem by providing complete, accurate, relevant, and timely intelligence to customers.

How KasadaIQ can help:

- ✓ Unparalleled insights into the adversary ecosystem. Our collection consists of deep access across 2000+ unique collection locations, with over 23 million messages ingested each month.
- ✓ Deep expertise on the automated threat landscape, adversary behaviour and intelligence practice. Dedicated analyst hours can help you dig deeper to investigate what's most relevant to your needs – whether that's intelligence on fraud groups, reverse-engineering, or pinpointing security weaknesses in your environment.
- ✓ Adversary engagement. We bridge the gap between you and the adversary to help better understand behaviours and patterns.
- ✓ Threat reporting. We equip you with regular cadence reporting on relevant trends at the organisation and/or industry level, tailored for all levels of audiences across your organisation.

Kasada is here to help.

Reach out through the following options

-  team-threat-intel@kasada.io
-  kasadaiq
-  @kasadaiq

