**State of the Internet/Security** Insights Series | V01 Issue 01

Akamai Security

SOTI

## Protecting Publishing:

# Navigating the AI Bot Era



# Contents


03 Key insights and findings

04 Introduction

05 AI's negative impacts on publishing's bottom line

06 AI bots target publishing

10 AI bot categories: Why it matters

13 How AI is undermining the traditional digital publishing economy

15 A recommended approach to managing operations in publishing

18 Security spotlight: Emerging ways to protect publishing

23 Conclusion

24 Methodology

25 Credits

# Key insights and findings

In the second half of 2025, Akamai observed that the media industry, which includes video, social media, and publishing, ranked second in Akamai-categorized AI bot activity.

Despite the dominance of training bots (which harvest vast amounts of data from websites to train large language models [LLMs]) in the media industry, AI fetchers pose a greater risk to publishing because online content loses value quickly. A total of 25.28% of Akamai-categorized AI bot types in media/publishing are AI fetchers, which collect real-time web content for user queries on AI assistants.

AI bots are undermining revenue mechanisms like paywalls, subscription models, and advertising-based income streams, which the publishing industry relies on.

Instead of just blocking all AI bots, various types of applied action are available for mitigating AI bot traffic: deny, tarpit, and delay are currently the most popular.

# Introduction

For the purposes of this report, the publishing sub-vertical data (see Methodology for more details) is grouped under the broader media industry, which also includes video and social media, as these segments share similar core business drivers: awareness, authority, and monetization. Whether it's a news site, a video platform, or a social network, success depends on brand visibility — measured through reach, impressions, unique visitors, and search rankings — to attract audiences.

Visibility alone, however, no longer guarantees revenue growth. Sustained success depends on credibility and trust, and a brand's position as a reliable source of information. Authority fosters engagement, repeat visits, and audience loyalty. Ultimately, awareness and authority form the foundation for monetization; awareness attracts advertisers and sponsors while authority sustains retention and steady revenue.

The rise of AI now threatens these business fundamentals, especially in publishing. This special Security Insights edition of the State of the Internet report explores how AI bots affect revenue, visibility, and brand awareness across the publishing industry, and what strategies can help protect its economic future in the AI era.

# AI's negative impacts on publishing's bottom line

Artificial intelligence (AI) is transforming the digital publishing ecosystem, but not always in ways that benefit publishing. The rapid rise of AI bots, especially content scrapers, has started to undermine primary income sources, tarnish brand reputation, and drive up operational costs. These bots are not just a security nuisance, they represent a profound business challenge that threatens the sustainability of quality journalism in an age dominated by zero-click searches and AI-generated content.

The publishing industry today faces an existential crisis triggered by financial concerns and the threat of being replaced by AI bots. The way audiences access information has transformed drastically and is directly related to the fact that the publishing companies no longer earn their living via subscriptions, ad engagement, and brand loyalty. Many readers and visitors still value trustworthy reporting and original content. Yet, instead of clicking through search results, users now turn to AI-driven platforms like ChatGPT and Gemini for instant answers and summaries.

This shift has real financial consequences. According to research from TollBit, AI chatbots drive significantly less referral traffic than traditional Google search — approximately 96% less in Q4 2024.

Yet, the monetary ramifications extend beyond traffic loss. AI scrapers copy and repurpose material created by the publishing industry, eroding hard-earned brand reputation and hijacking the data that drives digital strategy. Meanwhile, blocking and filtering these sophisticated bots require more infrastructure investment, which increases operational overhead even as revenue declines. The outcome is a multilayered challenge to combat reduced readership, revenue shortfalls, escalating costs, and weakened control over intellectual property.

# AI bots target publishing

AI has fundamentally changed how we find and consume information online. Instead of scrolling through search results or clicking between articles, we now get instant answers from AI platforms. These responses rely on information collected by AI bots from proprietary content, such as online articles, videos, and other exclusive digital assets, often without crediting or compensating the original creators. As a result, the media industry (including publishing) sees less web traffic, a reduction in subscription income, and a decline in overall brand visibility. The lack of proper attribution blurs the boundaries of content ownership.

In 2025, Ziff Davis, the parent company of the IT publications PCMag and Mashable, among others, filed a copyright infringement lawsuit in U.S. federal court against OpenAI. The company claimed that OpenAI used Ziff Davis's troves of articles to train LLMs without authorization or a licensing agreement. To support its case, Ziff Davis conducted research to analyze datasets like Common Crawl, C4, OpenWebText, and OpenWebText2 and examine how LLMs are trained. The findings revealed that AI companies tend to favor content from premium publishing companies like Ziff Davis, showing a strong preference for high-quality, trusted sources.

## Akamai-categorized AI bot activity

Akamai handles more than one-third of all global internet traffic. In 2025, we saw a 300% surge in Akamai-categorized AI bot activity, which accounts for only approximately 1% of overall bot activity. Most of these AI bot activities target the commerce industry (48%), but in second place is media (13%; Figure 1). This underscores the scale of automated content extraction and its direct impact on publishing organizations.

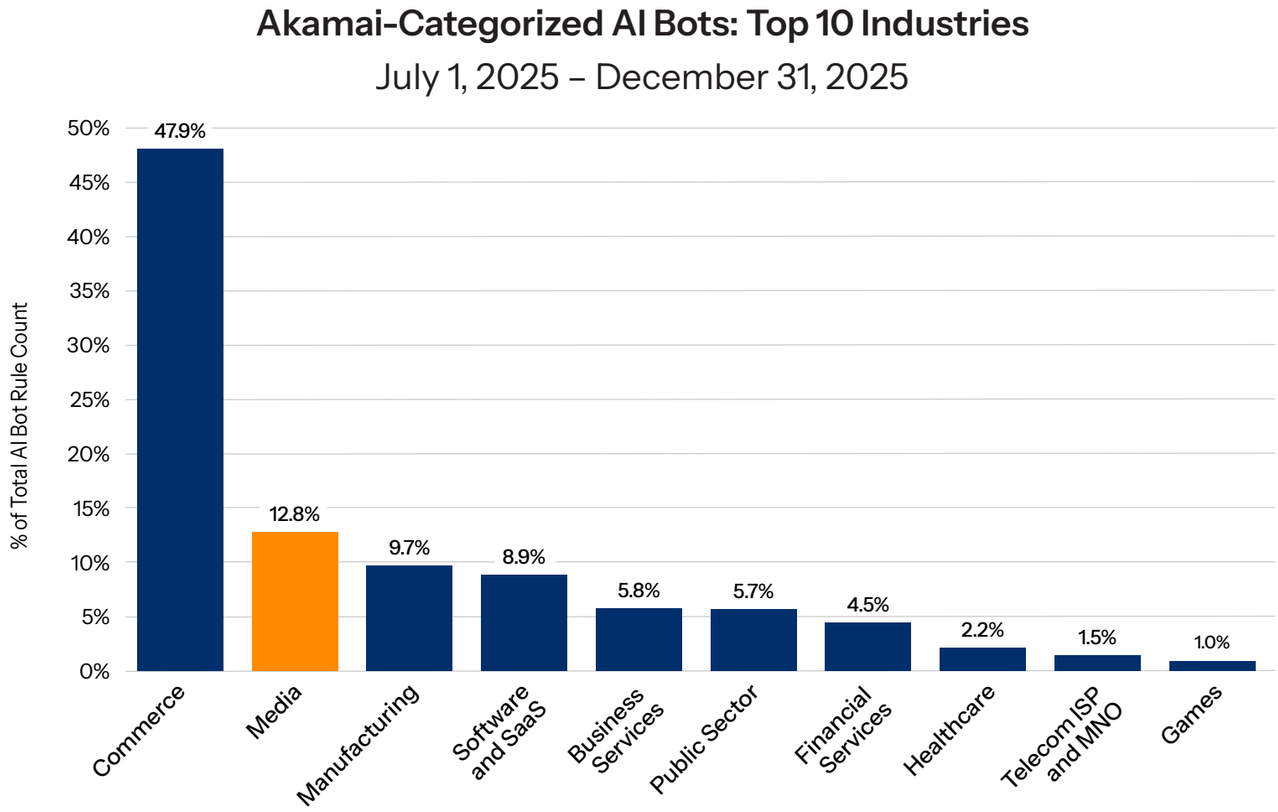



*Fig. 1: The media industry (which includes publishing) recorded the second-highest AI bot activity globally in 2H 2025*

A closer look at the media industry and its sub-verticals reveals that publishing organizations accounted for the majority of Akamai-categorized AI bots (40.1%), followed by broadcasting and TV companies (Figure 2).

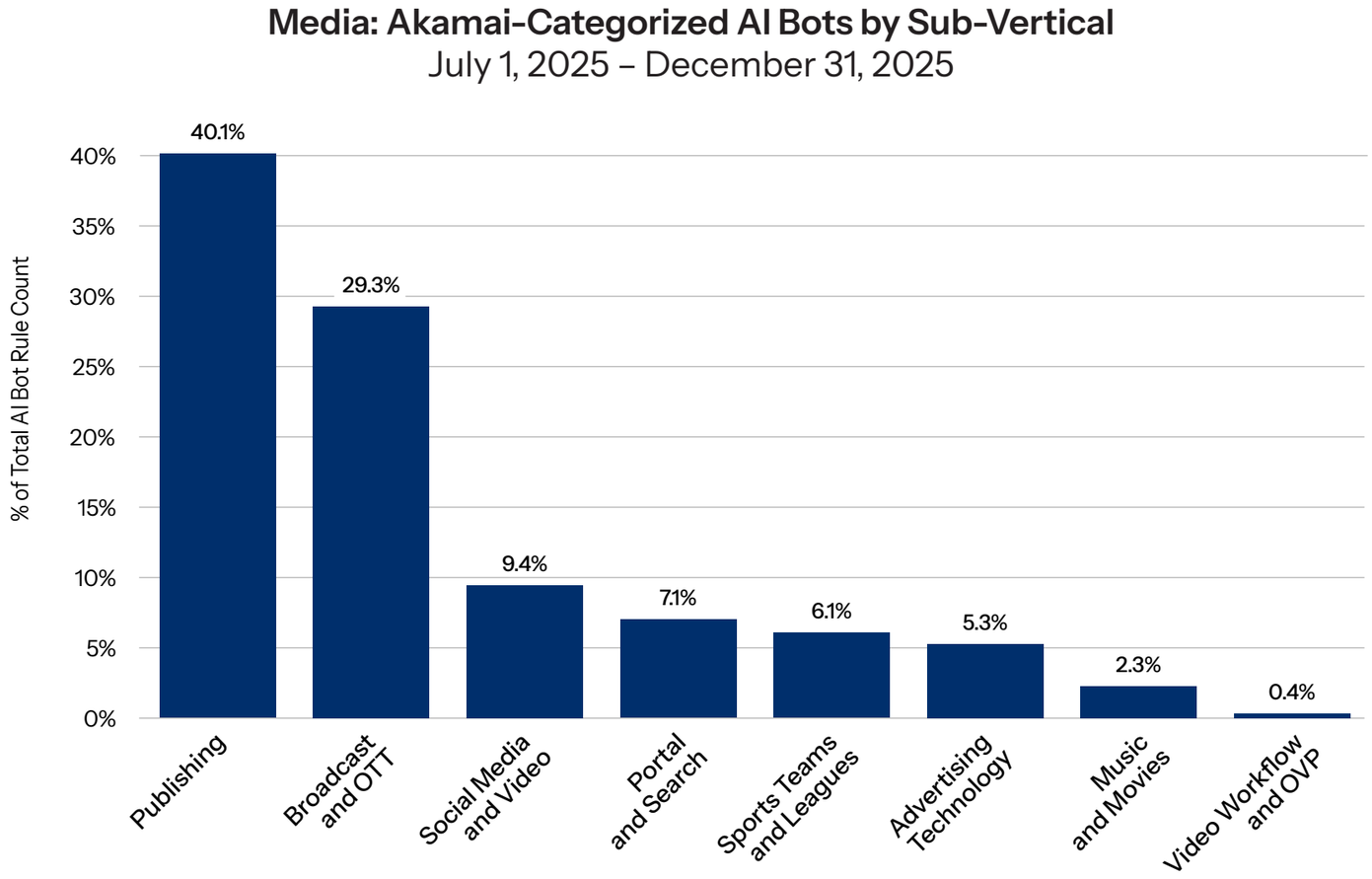



*Fig. 2: Within the media industry, publishing companies have the highest share of Akamai-categorized AI bots at 40.1%*

Between July and December 2025, OpenAI, Meta, and ByteDance were the top three vendors/owners of AI agents that impacted media companies (Figure 3).

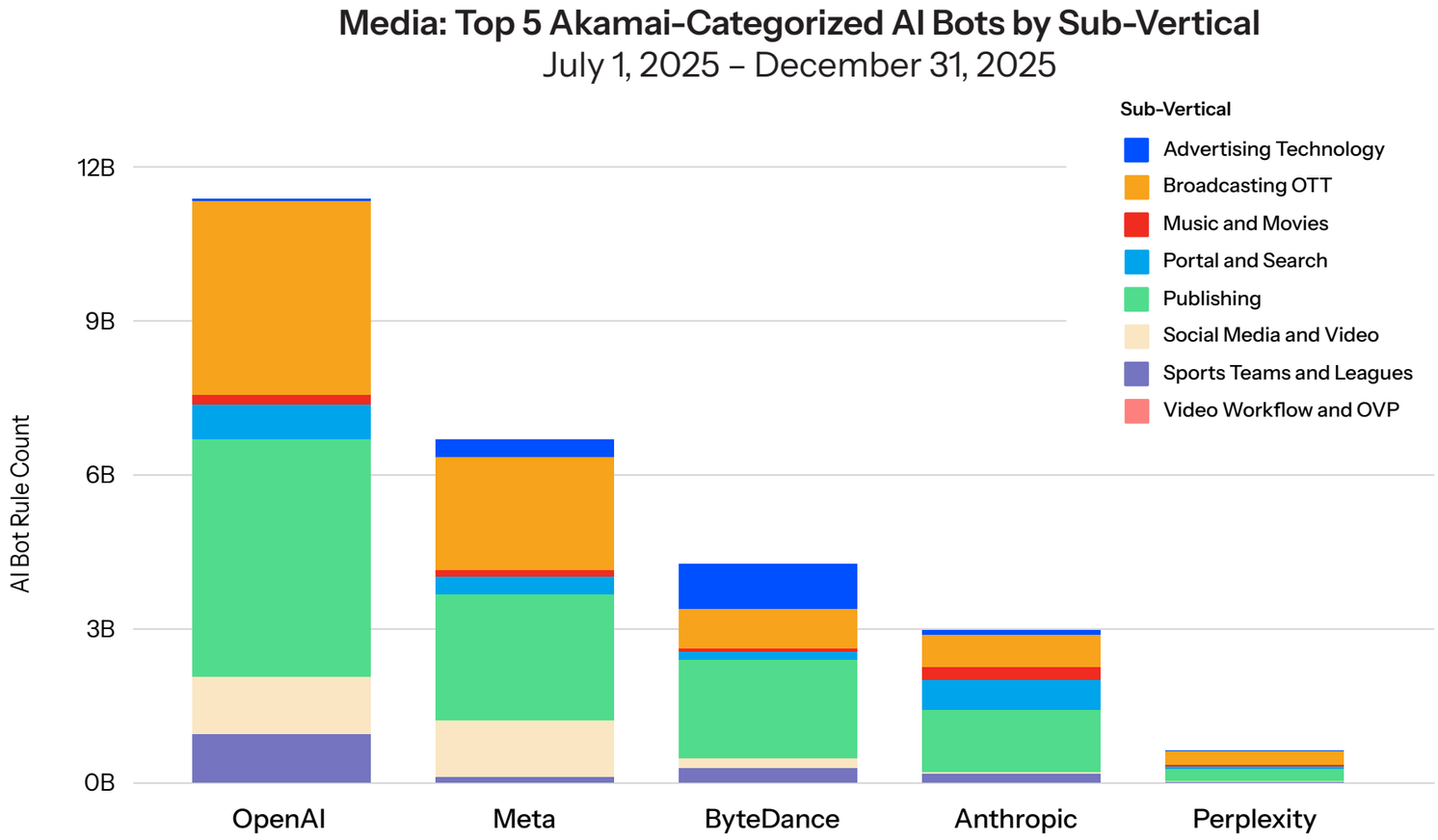



*Fig. 3: The majority of Akamai-categorized AI bot traffic in publishing comes from OpenAI*

Specifically, 40% of OpenAI requests within media are attributed to publishing companies, driven by increasing customer use, and resulting in a higher volume of requests than other vendors. This is no surprise since under the OpenAI umbrella are ChatGPT-User, GPTBot, and OAI-SearchBot, each one belonging to a different AI bot classification.

# AI bot categories: Why it matters

At Akamai, we classify our AI bot types into four core categories based on their behavior: AI training crawlers, AI search crawlers, AI fetchers, and AI agents (Table 1).

| Category name | Description | Popular examples |
|---|---|---|
| **AI training crawlers** | AI training crawlers automatically scan and collect large amounts of data from websites to be used as training data for LLMs. | GPTBot (OpenAI)<br>ClaudeBot (Anthropic)<br>Meta-ExternalAgent (Meta)<br>ByteSpider (TikTok) |
| **AI search crawlers** | AI search crawlers automatically scan and index websites for AI-powered search experiences. | OAI-SearchBot (OpenAI)<br>Claude-SearchBot (Anthropic)<br>PerplexityBot (Perplexity) |
| **AI fetchers** | AI fetchers grab specific web pages in real time to fulfill specific user requests on AI assistants, such as answering a question or summarizing a webpage via an AI chatbot. | ChatGPT-User (OpenAI)<br>Claude-User (Anthropic)<br>Perplexity-User (Perplexity)<br>DuckAssistBot (DuckDuckGo)<br>MistralAI-User (MistralAI) |
| **AI agents** | AI agents are autonomous software systems that use AI to achieve goals and complete tasks for users. | ChatGPT-Agent (OpenAI)<br>GoogleAgent-Mariner (Google) |

*Table 1: The categories of AI bots based on behavior*

Across all industries, we see a significant percentage of AI bot traffic that can be attributed to **AI training crawlers**, with a main function of gathering vast amounts of data for purposes of training LLMs. **AI fetchers**, which collect information from certain web pages in real time in response to a specific user request made on AI assistants (i.e., answering a question or summarizing a specific web page via AI chatbots), are the next most prevalent AI bot type. We also observe **AI search crawlers** that scrape various industries, with these bots indexing websites for AI-powered search results. Additionally, we see a small percentage of **AI agents** that use AI to perform tasks and objectives for users.

The media industry, which includes publishing, follows a similar pattern — with AI training crawlers and AI fetchers being the most common bot types that scrape their content. In H2 2025, AI crawlers made up 63% of all Akamai-categorized AI bots that targeted media, with 37% of those in publishing. And AI fetchers made up 24% of all Akamai-categorized AI bot types in the media industry, with publishing representing 43% of that total (Figure 4).



**Media: Akamai-Categorized AI Bot Types by Sub-Vertical**
July 1, 2025 – December 31, 2025

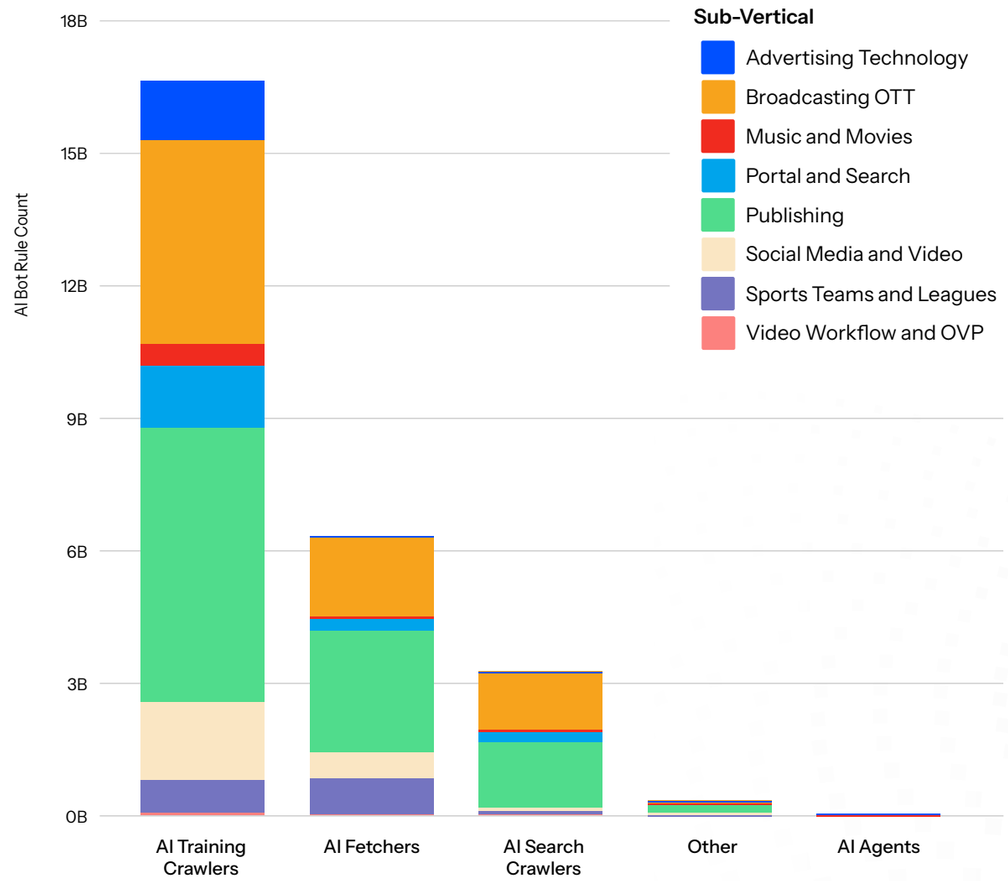



*Fig. 4: AI fetchers accounted for 24% of Akamai-categorized AI bot types in the media industry, 43% of those affected publishing organizations*

Although AI training crawlers are the most prevalent AI bots in publishing, AI fetchers are causing more immediate revenue impact for publishing organizations, partially due to the fact that some content (such as breaking news) loses value immediately. For instance, as users increasingly get their queries resolved through AI chatbots rather than by clicking through to publishing websites, page views and referral traffic decline significantly.

Understanding which AI bots are crawling a site, what content they're collecting, and what operators are behind them can give an organization an advantage in the high-speed cat-and-mouse game of publishing vs. AI platforms. In this digital AI era, staying ahead means having visibility into AI traffic.

# How AI is undermining the traditional digital publishing economy

As AI reshapes long-standing digital publishing practices, the industry is experiencing a profound period of transformation. This shift comes with serious implications for the publishing industry overall. Publishing is experiencing sharp declines in both traffic and revenue as users are increasingly turning to AI-generated answers instead of visiting websites directly.

As we discussed in the previous section, one of the ways that AI tools have been significantly impacting publishing is through the scraping of their websites. Scraping refers to the automated collection of website content, commonly conducted by AI systems without consent. These systems direct large-scale content harvesting and use the collected data to build and enhance their models to generate responses based on information that the publishing companies created and financed.

Ultimately, the AI system serves up this repurposed original content, leading users to bypass publishing sites entirely. It's a trend that has been undermining long-standing revenue mechanisms that the digital publishing industry depends on. This includes paywalls, subscription models, and advertising-based revenue streams. These traditional monetization methods, which have sustained the digital publishing industry for nearly 30 years, rely on consistent user engagement, and revenue plummets when audiences no longer click through website content.

In addition to lost revenue, web scraping leads to increased infrastructure expenses. Web scrapers contribute directly to higher operational spending caused by rising server, content delivery network (CDN), and cloud costs. The extra strain on infrastructure degrades site performance by slowing content delivery (increasing server and delivery costs) and diminishes overall user experience, as websites and applications struggle to serve unwanted automated traffic efficiently.

Beyond these immediate technical and financial burdens, web scraping can also inflict lasting reputational damage on publishing companies. Web scrapers extract full articles, headlines, and feature stories, which are then republished on impersonation, aggregation, or counterfeit sites that appear legitimate to readers and advertisers. When attackers, competitors, or counterfeiters misuse original content, it compromises trust in the brand, misleads consumers, and erodes the credibility companies work hard to establish.

AI-generated responses do not often give clear credit or provide linkage to the original creators, limiting recognition and attribution, so brand visibility is also weakened for both publishing organizations and authors. This has consequently damaged the trust, authority, and overall relationship that established media brands and publishing companies have cultivated with their audiences. A 2025 study from the Pew Research Center, which analyzed more than 68,000 queries, revealed that users only click on the sources cited in AI summaries approximately 1% of the time.

As AI increasingly intermediates information sharing and scrapes content without authorization, publishing is pushing harder to safeguard intellectual property and maintain fair compensation in digital spaces. It's crucial for publishing businesses to stay up-to-date on the evolving AI ecosystem that's shaping how content is discovered and monetized. These dynamics are redefining the relationship between publishing and audiences and making it increasingly vital to understand how AI tools interact with online content. Additionally, the publishing industry needs to determine which revenue, legal, technical, or collaborative strategies best align with their goals and long-term interests.

# A recommended approach to managing operations in publishing

To stay successful, the publishing industry needs to understand how AI tools and bots are changing the internet — and learn how to protect their original work. That means being able to view, understand, and manage AI bot activity on their sites.

Many companies in the publishing industry don't realize how many various AI bots are out there. Blocking all of them by default might seem like a quick fix, but that can backfire. For example, some AI companies are open to partnerships or licensing content, which could lead to new revenue. If publishing companies block all AI bots without knowing which ones are which, they risk losing those opportunities, and their content could become less visible online. By blocking all bots, publishing companies may turn away from the very innovations that will shape future monetization in their industry.

These trade-offs are influencing publishing today. One of our customers, who prefers to remain undisclosed, evaluated whether to block AI bots by default but ultimately chose to tarpit all AI bot requests to slow responses rather than deny access outright. In doing so, they were able to control 97% of AI bot requests and assess strategies to monetize this traffic. This example highlights the importance of managing how AI bots interact with a business to ensure protection without compromising legitimate engagement opportunities.

It's important to work with business partners to establish operational goals. Publishing companies may have different interests than other industries when it comes to handling AI bots. A publishing organization might decide to allow search bots that help drive readers to their content, but block AI bots that scrape data without permission. A retailer, on the other hand, might allow both kinds of AI bots if they improve product discovery or recommendations.

At a more technical level, when choosing which AI bots to handle, businesses can adopt a range of strategies to manage their interactions more effectively. Some popular options include deny, tarpit, and delay.

- **Deny** blocks and rejects bot requests completely and immediately. It conserves server resources entirely and prevents any data extraction, since there is no served content.
- **Tarpit** prolongs bot connections indefinitely, delivering minimal or no response data at a very slow rate. It's an aggressive approach aimed at causing bots to exhaust themselves of resources in their attempts to connect until they time out.
- **Delay** introduces a brief, intentional pause before fulfilling requests, which can help slow down bots during peak traffic times.

Of these options, deny seems to be the most popular strategy among Akamai customers, followed by tarpit, and then delay (Figure 5). However, before deciding to block AI bots outright, most customers first allow and monitor bot traffic to gain visibility into which ones are harmful (as opposed to using a blanket block), since not all AI bots are being used maliciously. Nonetheless, the visibility often leads customers to selectively deny unwanted bots.

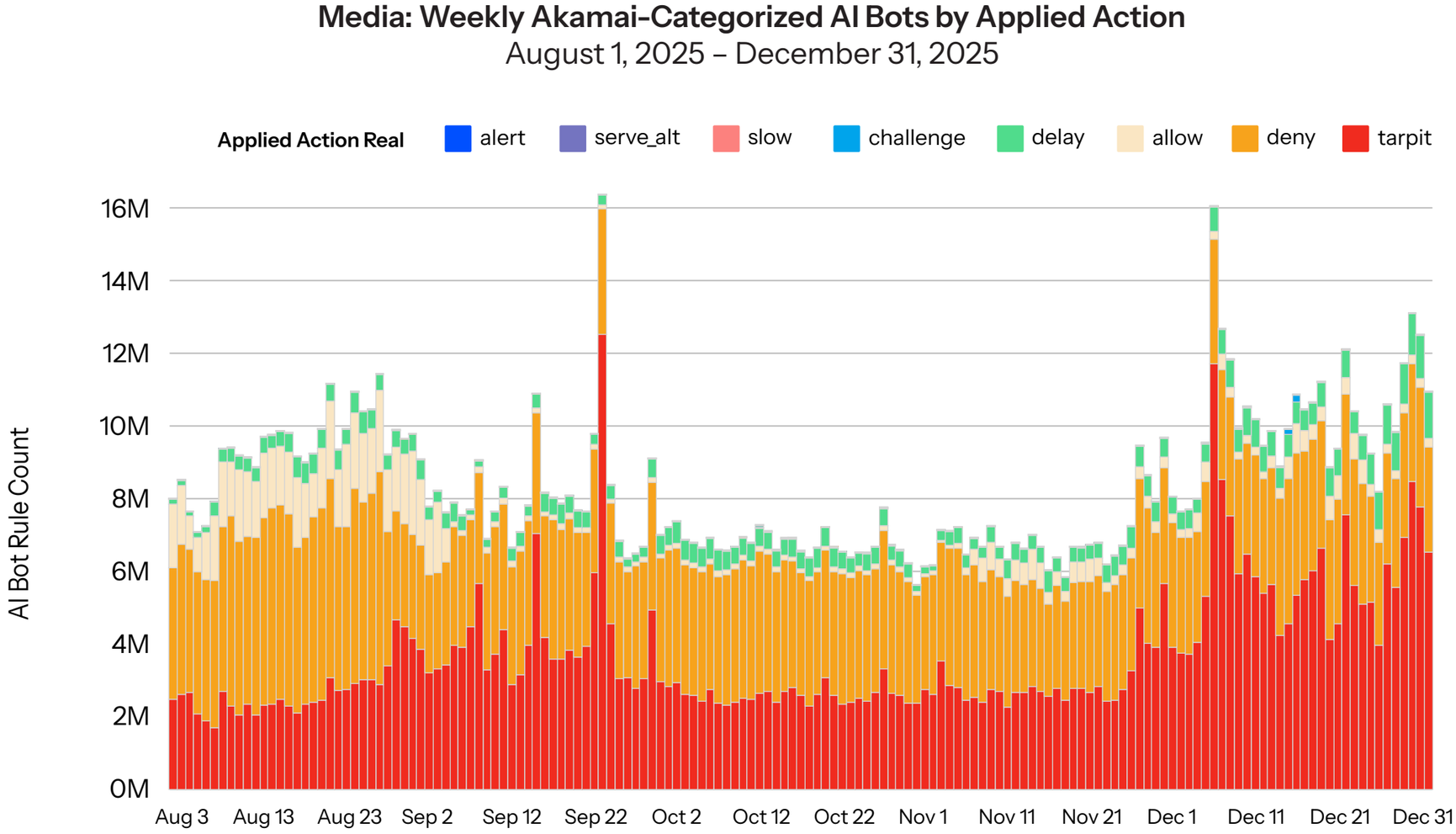



*Fig. 5: Total AI bot traffic mitigation by mitigation action across all Akamai customers*

### AI bot management checklist for publishing

Use this checklist to assist in the management of AI bots.

- ☐ **View.** Don't just block all AI bots.
- ☐ **Understand.** Identify the AI agents that are willing to pay.
- ☐ **Manage.** Control AI bot interaction. Options may include deny, tarpit, or delay.
- ☐ **Verify.** Validate the identity of the AI bots that are accessing the website through web protocols.
- ☐ **Enforce.** Establish a payment system for bots that are scraping content.

Security spotlight

# Emerging ways to protect publishing

As automation dominates web traffic, the publishing industry needs ways to distinguish legitimate, user-driven agents from unauthorized agents. A practical solution is a trust-and-commerce layer that makes approved agent access attributable and transactable by allowing protection, monetization, and transparency without forcing publishing companies to choose between blocking everything or accepting unpaid use. The goal is to turn automated demand into accountable, permissioned access that can be priced, audited, and enforced without overhauling the publishing industry's existing systems.

## Partnerships with platforms: Skyfire and TollBit

Instead of denying all AI bots, advanced strategies enable content licensing and fair value exchange through partnerships with platforms such as Skyfire and TollBit.

### How does Skyfire work?

Skyfire's Know Your Agent (KYA) operates as an identity foundation that allows publishing companies to infer intent from fingerprints and behavior alone. It does so by providing a portable, verifiable identity signal for agents — an identity that comes from a verified source rather than self-asserted claims, which can be delivered through existing infrastructure that publishing companies already operate at the edge.

When a bot that meets the integration criteria (such as an AI bot) sends a request, Akamai responds with instructions directing the bot to register with Skyfire, generate a token, and include it in subsequent requests via the skyfire-pay-id header. When the bot submits a request containing this token, Akamai validates the token's signature and claims, then forwards it to Skyfire for payment processing. Upon successful payment, the requested content is retrieved from the origin and served to the bot.

This gives publishing a new control surface — once identity is verifiable, policy becomes selective. Legitimate, user-initiated agent traffic can be allowed or metered under clear rules, while unknown automation is still challenged or blocked with the same rigor.

KYAPay completes the picture by turning verified identity into enforceable commerce. Skyfire's KYA + Pay Tokens package identity context and payment intent in a signed, interoperable envelope (structured as JSON Web Tokens), allowing the publishing industry to couple access decisions with monetization in a way that is both automatable and auditable (Figure 6).

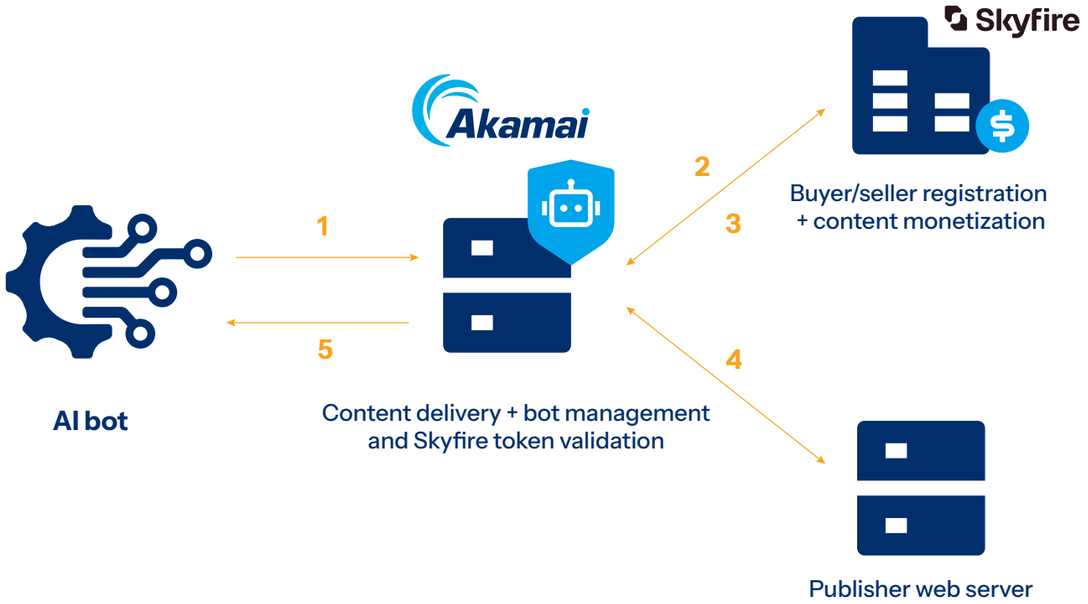



*Fig. 6: KYAPay is a new mechanism that verifies users by creating a digital token that is validated with each request and identifies who is making the request and their purpose*

## Turning access into revenue

This strategy directly impacts how publishing makes money because it maps to real business needs: Single-use access, metered use, and subscription-like entitlements initiated by agents, all with accountability about who is behind the request. As licensing standards evolve to define the rules for access, KYA and KYAPay handle the operational side: Verifying who's asking and managing the flow of payments or permissions. In short, publishing companies can protect their content, participate in the agent economy, and earn revenue when their content is being used.

## How does TollBit work?

TollBit enables a fair exchange of value between AI companies and websites by transforming scraping into sanctioned, revenue-generating access. Purpose-built for publishing companies and site owners, TollBit provides the transaction layer that governs how AI agents authenticate, access, and pay for content at scale.

Once Akamai identifies and redirects AI bots, their requests are handed over to TollBit's bot and agent paywall where they're verified, assigned a value, and managed to ensure proper payment and access (Figure 7).

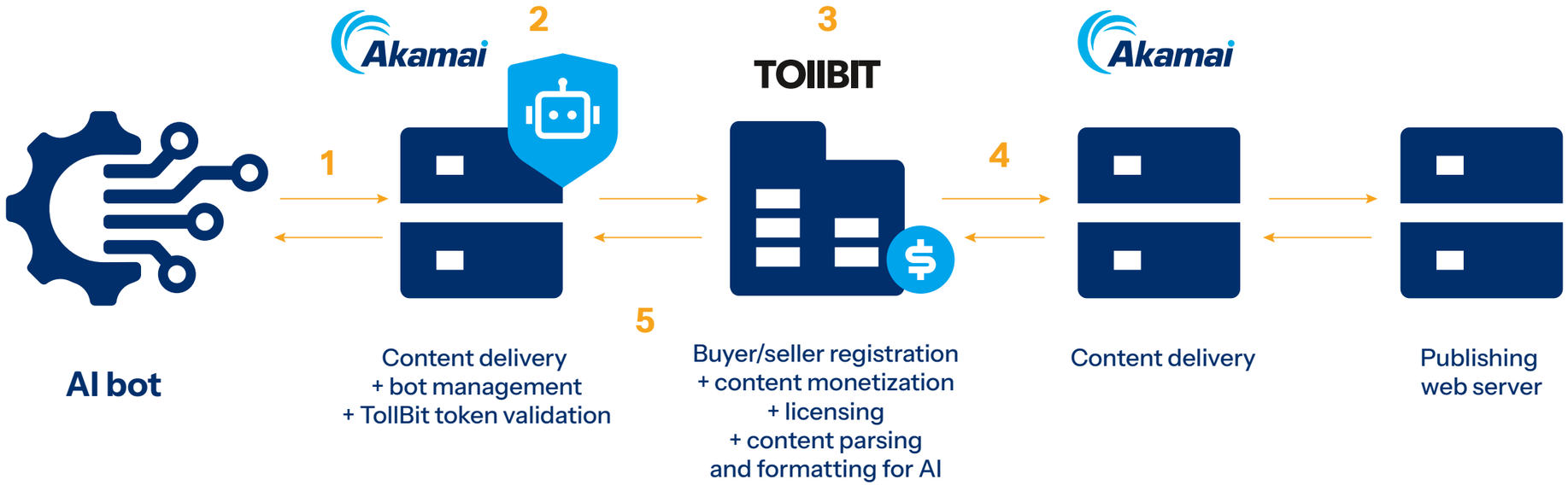



*Fig. 7: Once AI bots are identified and routed by Akamai, they are sent to TollBit's bot and agent paywall so requests can be authenticated, priced, and enforced*

The paywall allows publishing companies to actively enforce robots.txt policies, turning them from a passive signal into real access control. Additionally, unauthorized bots and agents are denied access and shown a clear message indicating that they are not permitted to use the content, while approved AI systems are required to authenticate and pay (when applicable) before accessing it.

TollBit uses short-lived, cryptographically signed tokens to verify the identity and intent of each AI request. These tokens prevent spoofing and reuse, ensuring that access is limited to approved AI organizations and agents. Each request is verified in real time and tied back to a specific AI partner, making AI access authenticated, attributable, and enforceable at scale, without adding operational complexity for the publishing company.

With authenticated access in place, TollBit enables publishing companies to monetize AI use with flexibility and control. Site owners define pricing via pay-per-use economics, supported by standard licenses for retrieval-augmented generation (RAG).

Users can also apply granular pricing rules by bot, page, directory, keyword, or time window.

Today, more than 6,500 publishing organizations and websites use the TollBit platform to monitor over 6 billion quarterly AI bot scrapes, set access terms, and receive direct compensation when AI bots and agents access their content. In the second half of 2025 alone, TollBit analyzed more than 500 billion website visits and redirected approximately 2 billion AI bots through the TollBit bot and agent paywall.

### The push for RSL

Another key development in the publishing sub-vertical is the adoption of Really Simple Licensing (RSL). RSL is a framework that ensures that AI companies collect and use digital content responsibly, under clear terms set by the content owner. Encouragingly, several major technology companies, including Akamai, have expressed support for this initiative. Work is also underway to make RSL compliance-ready and standardized, which will help more companies in the media industry protect their digital assets.

# Conclusion

This report explored the growing challenges of web scraping and the ways AI bots have intensified the problem for the publishing industry. Using Akamai's research data, it examined where publishing ranks within the broader media landscape to provide a granular view of the issues (and their impact) faced by publishing from automated scraping. Additionally, the report also shared practical insights on how publishing companies can reclaim control of their content.

LLMs are often trained without the content originators' knowledge or consent, and we believe that content creators deserve fair compensation when their work is used for this purpose. Through Akamai's partnership with TollBit and Skyfire, we make that possible. When an AI bot tries to access protected content, it will be redirected to a payment gateway before allowing access. This ensures that AI platforms compensate publishing companies fairly for their content and gives visibility into who's using it and for what purpose.

This approach reinforces the idea that in the AI era, there is no such thing as a free lunch, as access to valuable content comes at a cost, even for bots.

Effective bot solutions can help publishing organizations reduce infrastructure costs during traffic spikes by filtering out unwanted traffic from scrapers and scalpers. This avoids overprovisioning for bot-driven peaks, helps reduce spending on firefighting performance incidents, and keeps more inventory and revenue tied to real customers instead of bots.

The publishing industry holds a position of strength in this fight. By combining bot management, licensing frameworks, web protocols, and monetization strategies, publishing can push back against unauthorized scraping and reclaim control over how their content is used and distributed. This approach helps ensure that their digital and proprietary assets are protected, valued, and accessed on their own terms.

# Methodology

This data describes application-layer traffic seen through our bot management tools. Our rules are triggered when we detect bot telemetry from an Akamai-categorized AI bot within a request to a managed website, application, or API. These tools help detect, manage, and mitigate bot traffic according to customer expectations.

*This research dataset covered data from July 1, 2025, through December 31, 2025, unless otherwise noted.*

# Credits

## Research director
Kimberly Gomez

## Writing and editing
Lance Rhodes  Maria Vlasak
Badette Tribbey

## Review and subject matter contribution
Emily Lyons  Ari Weil

## Data analysis
Chelsea Tuttle

## Promotional materials
Ashley Linares  Ellen O'Brien

## Marketing and publishing
Kimberly Gomez
Georgina Morales Hampe

## State of the Internet/Security

Read back issues and watch for upcoming releases of Akamai's acclaimed State of the Internet/Security reports.

## Akamai threat research

Stay updated with the latest threat intelligence analyses, security reports, and cybersecurity research.

## Akamai security research

Read the Akamai security research blog for a rapid response perspective on today's most important research.

## Access data from this report

View high-quality versions of the graphs and charts referenced in this report. These images are free to use and reference, provided that Akamai is duly credited as a source and the Akamai logo is retained.

Akamai Security

Akamai Security protects the applications that drive your business at every point of interaction, without compromising performance or customer experience. By leveraging the scale of our global platform and its visibility to threats, we partner with you to prevent, detect, and mitigate threats, so you can build brand trust and deliver on your vision. Learn more at **akamai.com** and **akamai.com/blog**, or follow Akamai Technologies on **X** and **LinkedIn**. Published 04/26.