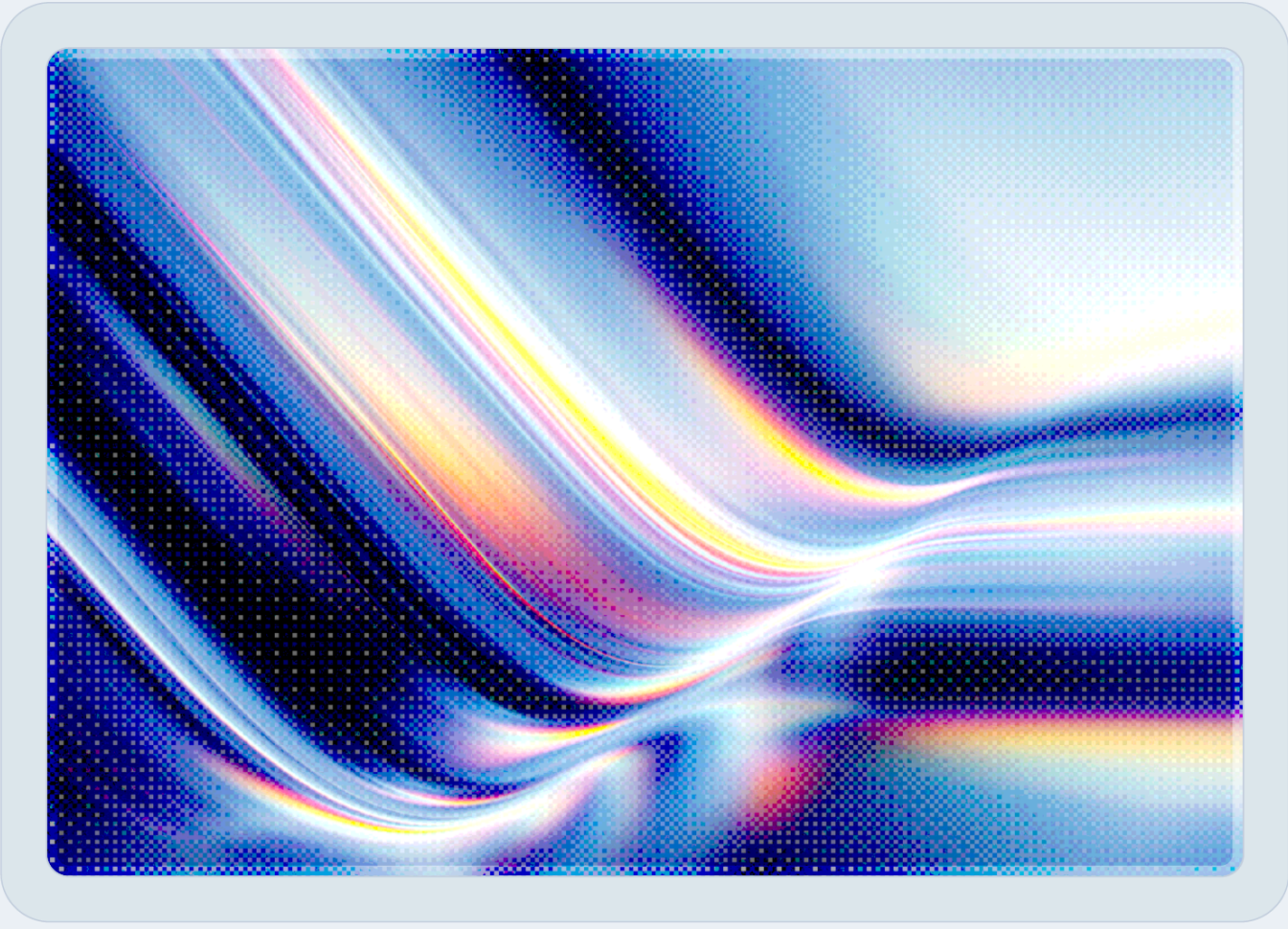
CSAI Foundation | Cloud Security Alliance

Hidden Nodes: AI Scraping SDKs as Enterprise Attack Vectors

How the AI Data Collection Industry Turns Consumer Devices into Corporate Security Risks

2026-06-08

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Table of Contents

Executive Summary

The infrastructure powering AI model training has a hidden layer that few enterprise security teams have considered: millions of consumer devices — smart televisions sitting in living rooms, mobile phones in employees' pockets, and connected appliances in corporate break rooms — are enrolled as exit nodes in commercial residential proxy networks. These networks serve AI companies, data brokers, and other commercial customers seeking to scrape the web at scale by routing requests through residential IP addresses that bypass anti-bot detection systems. The devices are enrolled through software development kits (SDKs) bundled into popular free applications, typically with consent mechanisms that Include Security's analysis found materially understated the scope of relay activity.

In June 2026, researchers from Include Security published a detailed technical analysis of one such SDK, developed by Bright Data, a Tel Aviv-based data-collection company that markets what it describes as the world's largest residential proxy network [1]. The research revealed that the SDK, embedded in applications distributed on Samsung and LG smart television platforms and in major mobile apps, includes two independent mechanisms for evading virtual private networks and standard security monitoring tools [1]. It continuously transmits device telemetry — battery level, CPU load, network type, screen state, and geographic location indicators — to Bright Data's infrastructure [1]. It relays third-party web scraping traffic even when the device screen is active and during phone calls [1].

For enterprise security teams, the implications extend well beyond consumer privacy. Smart televisions in corporate conference rooms, BYOD mobile devices carried by employees, and consumer-facing applications distributed through enterprise app catalogs may all participate in this proxy infrastructure without the knowledge of IT or security teams. The VPN bypass capability documented by Include Security means that standard network monitoring controls fail silently against this traffic class [1]. Meanwhile, the broader residential proxy ecosystem — of which AI-focused networks are one component — has been established by multiple research organizations and the FBI as a primary obfuscation layer used by criminal threat actors and state-sponsored attackers [6][7][9].

This paper examines the technical architecture of commercial AI scraping SDKs, their implications for enterprise security, their convergence with the criminal proxy ecosystem, the emerging regulatory landscape, and practical guidance for detection and mitigation. The findings suggest that enterprise security policy must evolve to treat embedded proxyware as a first-class threat category, comparable to spyware or unauthorized remote access tools, regardless of whether the SDK is delivered through a nominally legitimate commercial channel.

1. Introduction: The Hidden Infrastructure Tax of AI Training Data

The hunger of large AI models for training data is, by now, well understood in broad strokes: systems are trained on vast corpora scraped from the public web, and the quality, recency, and breadth of that data materially shapes model capability. What is less understood is the infrastructure that makes that scraping operationally viable at scale, and the downstream consequences that infrastructure imposes on organizations that never contracted for it.

Modern websites employ increasingly sophisticated anti-bot defenses. Services from Cloudflare, DataDome, HUMAN Security, and others analyze traffic patterns, TLS fingerprints, browser behavior, and — critically — the reputation and origin of the requesting IP address. Requests originating from cloud provider IP ranges are treated with heightened suspicion or blocked outright, because legitimate human users rarely browse the web from an AWS subnet [2]. The practical result is that AI companies and data brokers seeking to scrape at industrial scale cannot rely on datacenter infrastructure alone.

The solution the industry converged on is residential proxies: a network of IP addresses belonging to consumer broadband subscribers, whose traffic appears to anti-bot systems as ordinary home internet usage. A scraping job routed through a residential proxy arrives at its target from the IP address of a household subscriber in, say, suburban Ohio or central Manchester, indistinguishable in terms of network origin from a legitimate human visitor. The economics are straightforward: any bottleneck created by bot detection can be dissolved if you have enough residential IPs to distribute requests across.

Building and maintaining a residential proxy network requires enrolling consumer devices as exit nodes — machines that agree to relay third-party traffic through the subscriber's home internet connection. Two broad mechanisms accomplish this enrollment. The first is covert: malware, fake application downloads, and compromised app stores install proxy agents without any disclosure. The second is nominally consensual: commercial SDK providers partner with application developers who embed proxy relay functionality into otherwise legitimate free applications, in exchange for revenue sharing. The embedded SDK presents some form of disclosure during app installation or first launch, and the application developer monetizes their user base by selling those users' bandwidth to data customers.

This paper concerns the second category — not because it is less harmful, but because it occupies a legal gray zone that causes enterprise security teams to underestimate the risk. The commercial framing of the consent-based model creates an appearance of legitimacy that may discourage security response. In

practice, as the technical evidence demonstrates, the distinction between consensual proxyware and covert malware is narrower than the commercial framing suggests, and the enterprise exposure from both categories is substantially the same.

2. The Commercial Proxy SDK Ecosystem

2.1 Market Structure and Demand Drivers

Commercial residential proxy providers occupy a specific position in the AI data supply chain. They aggregate bandwidth from consumer devices and resell it to buyers who need residential IP addresses to conduct web scraping, price monitoring, ad verification, and AI training data collection at scale. The provider charges per gigabyte of traffic routed; the application developer embedding the SDK earns a share of that revenue; the consumer receives the application for free.

This model is not new — residential proxy networks predate the current AI wave — but AI data demand has expanded the market and altered its composition. Industry observers and market researchers suggest the commercial proxy services market is growing, with AI training data collection emerging as a significant demand driver alongside longstanding use cases such as price monitoring and ad verification [2][20]. Bright Data, the market's most prominent operator, advertises access to a network exceeding 400 million residential IP addresses globally [1][19]. The company claims approximately 150 million of those addresses derive from consent-based SDK partnerships, with the remainder sourced through other means [1].

The AI data collection use case is explicit in Bright Data's marketing. The company openly positions its residential proxy network as infrastructure for gathering training corpora, and its customer base includes AI developers who need web-scale data that would otherwise be blocked by anti-bot systems [4]. This is not a case of proxy infrastructure being misused for AI purposes — it is proxy infrastructure built and marketed for that purpose.

2.2 How SDK Partnerships Work

The commercial SDK model operates through a tiered partnership structure. A proxy provider such as Bright Data develops an SDK — a software library that, when embedded in a mobile or connected-TV application, transforms the device into a proxy relay node. The SDK provider recruits application developers as partners, offering revenue sharing in exchange for embedding the SDK in their applications. Those developers distribute their applications through platform app stores — Samsung's Tizen store, LG's Content Store, Apple's App Store, Google Play, and others.

The Include Security research identified the scope of Bright Data's SDK partnerships across the connected television ecosystem in particular [1][3][14]. PlayWorks Digital, a distributor of casual game titles for connected TVs, has embedded the SDK across more than 400 titles reaching an estimated 250 million television households [1]. CloudTV, which provides software to television manufacturers, has deployed the

SDK across 125 or more TV brands spanning more than 15 original equipment manufacturers [1]. Viber, a messaging application owned by Rakuten, has similarly integrated the SDK and distributes it to a user base measured in the hundreds of millions [1]. Moonfrog Labs, a mobile gaming company, reaches approximately 10 million monthly active users through SDK-embedded titles [1][21].

These figures illustrate the scale of device enrollment that a single SDK provider can achieve through the partner application model. The enrolled devices are not uniformly distributed across all platforms: Google, Amazon, and Roku restricted background proxy SDK use on their platforms following earlier reporting, leading Bright Data to drop support for those ecosystems [2][13]. As of this writing, Samsung's Tizen operating system and LG's webOS remain listed as supported platforms [1][2][3].

2.3 Consent Disclosure Limitations

The commercial SDK model's legitimacy rests on a claim of informed consumer consent. In practice, the consent mechanism documented by Include Security and other researchers falls far short of meaningful disclosure. The opt-in is presented as a EULA or privacy dialog during application installation, navigated on a television using a remote control's directional arrows — a user interface modality that is poorly suited to reading and comprehending a multi-page legal agreement [1]. The research characterizes the disclosure as materially mismatched to what the SDK actually enables: the consent dialog does not accurately describe the scope of traffic the device will relay [1].

On mobile platforms, the situation is comparably problematic. Consumer privacy expectations for mobile applications do not typically include the understanding that the application is monetizing the device's internet connection by routing third-party commercial web scraping traffic through it. According to the Include Security analysis, the consent mechanisms examined do not disclose the VPN bypass and telemetry capabilities in language a typical user could be expected to understand and evaluate [1].

This matters for enterprise security not merely as an ethical concern but as a practical risk factor. An employee installing a free game on their personal television or smartphone has no meaningful ability to evaluate whether the application embeds proxy infrastructure — and no enterprise application policy, short of a comprehensive software inventory and binary scanning program, will catch it.

3. Technical Anatomy of the SDK

3.1 Command and Control Infrastructure

The Include Security researchers conducted a detailed reverse engineering of the Bright Data SDK for iOS devices [1]. Their findings reveal an architecture that is technically sophisticated in ways that are directly relevant to enterprise security operations.

The SDK establishes a persistent WebSocket connection to `proxyjs.brdtnet.com` on port 443, routed through AWS Global Accelerator for reliability and geographic distribution [1]. This endpoint serves as the command and control channel through which the device receives instructions about when to relay traffic, what destinations to connect to, and how much bandwidth to allocate. The SDK also contacts an unauthenticated configuration endpoint at `clientsdk.bright-sdk.com/sdk_config_ios.json`, from which it retrieves operational parameters including resource limits and geographic bandwidth allocations [1]. The legacy TLS certificate presented by Bright Data's infrastructure still uses a common name of `*.luminatinet.com`, reflecting the company's prior corporate identity as Luminati Networks — a detail that may cause the traffic to be misidentified or overlooked in network inspection systems unfamiliar with the name transition [1].

The Include Security researchers described the security posture of the job dispatch channel — the mechanism by which scraping tasks are assigned to enrolled devices — as offering fewer integrity protections than are commonly implemented in malware command and control infrastructure [1]. This observation is significant: it implies that the channel through which the device receives instructions to relay third-party web traffic is itself susceptible to abuse by parties other than Bright Data.

3.2 VPN Bypass Mechanisms

Among the most significant technical findings from the Include Security research are the SDK's two independent mechanisms for evading virtual private network inspection and standard mobile security tooling [1].

The first mechanism operates at the control plane. Rather than using Apple's standard URLSession API — which is instrumented by mobile threat defense solutions and enterprise MDM systems — the SDK employs lower-level CFNetwork primitives directly [1]. Because most security monitoring hooks target URLSession, this approach causes the SDK's control-plane traffic to appear invisible to those monitoring systems.

The second mechanism operates at the data plane, which carries the actual web scraping traffic being relayed through the device. The SDK uses Apple's Network framework with an `NWConnection` object configured with a `requiredInterface` parameter that binds the connection to the device's physical network interface — either the WiFi adapter (`en0`) or the cellular modem (`pdp_ip0`) [1]. This binding causes the traffic to bypass the virtual TUN interface through which VPN connections route all traffic, meaning the scraping relay traffic exits the device directly to the internet, unencrypted by the corporate VPN tunnel and invisible to VPN-based monitoring [1].

The combination of these two mechanisms means that an enterprise deploying a VPN-based mobile threat defense solution — currently among the most common approaches to BYOD security — will not observe the SDK's traffic in their monitoring infrastructure. The proxy relay activity is, from the perspective of the enterprise security team's tooling, simply invisible.

3.3 Device Telemetry and Operational Parameters

Beyond its proxy relay function, the SDK continuously transmits a detailed telemetry stream to Bright Data's infrastructure [1]. The reported telemetry fields include: connection state indicators for WiFi and cellular interfaces; mobile network type (LTE or 5G); roaming status; battery level; screen state (on or off); active call status; current CPU utilization; memory utilization; available bandwidth metrics; IPv6 capability; and the unique application identifier and SDK version [1].

Bright Data describes this telemetry as serving the operational purpose of intelligently scheduling proxy relay tasks — it would be counterproductive to relay high-bandwidth scraping traffic through a device with a dying battery or saturated CPU. Whether the data collected is limited to that purpose is not verifiable from the SDK's behavior alone. From an enterprise security perspective, the continuous reporting of device state to a third-party commercial infrastructure provider represents an unauthorized telemetry channel that organizations may not have approved and that may capture behavioral patterns of enterprise users.

The SDK's operational parameters reveal additional detail about how the relay function is designed to maximize device availability as a proxy node. Configuration retrieved from Bright Data's servers documented relay permission even when the device screen is active — a setting labeled `ignore_screen_on: true` — and during active telephone calls [1]. CPU and memory thresholds are set to 70% and 90% respectively, meaning the device may relay substantial commercial scraping traffic while the user is actively working [1]. Battery minimums are set at 20% globally, with variation by geography: devices in certain Middle Eastern markets operate under stricter battery constraints, while devices in other regions relay under a global default [1].

3.4 Smart Television as an Optimal Proxy Node

The Include Security research specifically highlights connected television devices as ideal proxy nodes from the perspective of proxy network operators, and this framing is important for enterprise risk assessment [1]. Unlike mobile phones, smart televisions maintain continuous power and internet connectivity. They are rarely monitored by their owners for unusual network activity. They do not have battery constraints that interrupt relay service. They are almost never enrolled in enterprise MDM systems, even when located in corporate conference rooms or common areas. They operate continuously, providing high uptime relative to mobile devices, which may be locked, powered off, or traveling between networks [1].

Enterprise deployments commonly include smart televisions in conference rooms for video conferencing, in break rooms for news and entertainment, and at building entrances for digital signage — all connected to the corporate network. If those devices run applications embedding proxy SDKs, they become permanently available relay nodes directly connected to the enterprise network, outside the visibility of most endpoint security tooling. The enterprise network's IP address range, not a consumer's home address, becomes the exit point for third-party scraping traffic.

4. The Enterprise Exposure

4.1 The BYOD Attack Surface

Bring-your-own-device policies, widely adopted across enterprise environments, create the primary vector through which SDK-embedded proxyware reaches corporate networks. An employee using a personal smartphone or tablet for work-related purposes — accessing email, connecting to VPN, authenticating to corporate services — may simultaneously be running applications that embed proxy SDKs on that same device. Enterprise MDM solutions provide some visibility into installed applications on enrolled devices, but their effectiveness is constrained in several ways.

First, many organizations implement limited enrollment MDM policies, sometimes called BYOD or user enrollment, rather than full device management, specifically to respect employee privacy. These limited-enrollment profiles do not provide the application inventory visibility that would be necessary to detect SDK-embedded applications. Second, even where full enrollment is in place, detecting an SDK embedded within an otherwise legitimate application requires binary analysis capabilities that go beyond a simple application catalog audit. The Swift symbols identified by Include Security — `BrdWebSocketFacade` and `BrdNetwork.DNSResolver` — are reliable indicators of the Bright Data SDK's presence, but detecting them requires app binary scanning rather than application name matching [1]. Third, as documented above, the SDK's VPN bypass mechanisms cause its traffic to evade the network monitoring capabilities that enterprise security teams often rely upon for BYOD devices.

Given the scale of the partner application ecosystem — hundreds of millions of Viber users, 250 million television households covered by PlayWorks SDK integrations — it is reasonable to assume that enterprise-adjacent devices participate in this infrastructure. The security question is what policies, if any, govern their network behavior and whether those policies would catch this specific class of activity.

4.2 Corporate Network Infrastructure as Proxy Exit Points

When an SDK-embedded application is running on a device connected to a corporate network — whether over WiFi at the office or over a split-tunnel VPN from a home office — the device's corporate IP address becomes the exit point for third-party scraping traffic. This has several consequences.

Corporate IP address reputation may be degraded if the scraping traffic directed through those addresses is flagged by target websites or threat intelligence feeds. Outbound bandwidth consumption from the corporate network increases in ways that may be difficult to attribute. Corporate network addresses appear

in the access logs of whatever websites are being scraped, potentially creating legal or compliance complications if the scraping activity violates those sites' terms of service.

More fundamentally, the corporate network's perimeter controls are bypassed. Traffic that enters the corporate network from the Bright Data infrastructure, is processed by the enrolled device, and exits directly to the scraping target does not traverse the enterprise's outbound proxy, web filter, or data loss prevention systems. The traffic is invisible to those controls because the SDK binds it to the physical network interface rather than routing it through any monitored layer.

4.3 Telemetry as an Intelligence Channel

The continuous device telemetry transmitted by the SDK to Bright Data's infrastructure warrants specific attention from enterprise security and privacy teams. The reported fields include indicators that could inform adversarial analysis: network connectivity type, roaming status, CPU and memory load patterns, and screen state together constitute a behavioral fingerprint of the device user's activity patterns [1].

While there is no evidence that Bright Data uses this telemetry for purposes beyond proxy scheduling, the transmission itself represents an unapproved data channel from enrolled devices to a third-party commercial infrastructure provider. Organizations subject to data protection regulations — particularly those operating in healthcare, finance, government, or other regulated sectors — may have compliance obligations around the transmission of employee device data to unauthorized parties, regardless of whether that transmission occurs on corporate or personal devices used for work.

The configuration endpoint is also unauthenticated [1]. The configuration endpoint's lack of authentication on the response channel means the SDK accepts configuration updates without verifying they originate from Bright Data, creating a theoretical attack surface for adversaries who could intercept or substitute the configuration payload. This attack path would require defeating transport-layer protections, but the absence of response authentication represents a weaker security posture than a properly integrity-protected control channel.

4.4 The Silent VPN Failure

Perhaps the most significant operational security implication of the Include Security findings is the documented failure of VPN-based monitoring against SDK traffic. Enterprise security architecture has increasingly relied on mobile VPN connectivity as both a security control and a monitoring channel for BYOD devices. The premise is that traffic from enrolled BYOD devices, routed through the corporate VPN tunnel, becomes visible to enterprise security tooling and subject to the same filtering and logging as traffic from managed endpoints.

The Bright Data SDK's `NWConnection` binding technique directly undermines this assumption [1]. Traffic the SDK relays exits the device through the physical network interface, bypassing the VPN tunnel interface entirely. An enterprise security team relying on VPN-enforced traffic inspection will observe no evidence of this traffic in their logs — not because the traffic does not exist, but because it never enters the monitored channel. The security control that the team believes covers these devices is silently non-functional for this traffic class.

This failure mode is not unique to Bright Data's SDK. Any application that uses the `requiredInterface` API to bind connections to a specific network interface achieves the same VPN bypass effect on iOS. The use of this technique in a commercial proxyware context makes visible a class of monitoring gap that security teams should evaluate independently of any single vendor.

5. Criminal Infrastructure Convergence

5.1 A Shared Ecosystem

The residential proxy ecosystem does not maintain clean separation between commercial AI-scraping infrastructure and criminal abuse infrastructure. The same category of proxy network — residential IP addresses used to route web requests through consumer devices — serves fundamentally different purposes for different customers, and the infrastructure is often intermingled.

In January 2026, Google's Threat Intelligence Group disrupted IPIDEA, which it characterized as one of the world's largest residential proxy networks, with between 9 and 11 million daily active proxy nodes [6]. The disruption revealed that IPIDEA's network was used by more than 550 distinct threat groups, including state-sponsored threat actors from China, North Korea, Iran, and Russia [6]. These groups routed cyberattack traffic through residential proxy nodes to obfuscate the true origin of their operations, making attribution more difficult and bypassing IP-based blocking controls.

A large-scale study of the residential proxy ecosystem published in 2026 examined 53 million globally unique proxy exit nodes across commercial proxy services [5]. The findings revealed that 15 percent of distinct egress IP addresses in those networks — approximately 8.2 million addresses — were simultaneously flagged for active malware infections [5]. An additional 13 percent, approximately 6.8 million addresses, exhibited riskware activity [5]. The infected node subset showed persistent co-infections with malware families including Vo1d, Badbox, RootSTV/Pandoraspear, and others associated with connected television device compromises [5]. GreyNoise Intelligence has similarly documented a significant proportion of commercial residential proxy IP addresses as simultaneously compromised by malware, corroborating the scale of this ecosystem overlap [10]. These figures describe the broader residential proxy ecosystem rather than any single provider, but they establish that the pool of residential proxy nodes mixes consensually enrolled devices with covertly compromised ones, and that the underlying infrastructure is shared.

5.2 Botnet Convergence on the Same Device Categories

The device categories that commercial proxy SDKs target — smart televisions and connected streaming devices — are the same categories that criminal botnet operators have simultaneously targeted. In December 2025, a botnet designated Kimwolf infected more than two million Android TV streaming devices, leveraging residential proxy software to enroll compromised devices as relay nodes [9]. The attack vector and the commercial SDK vector produce functionally identical outcomes: a device routes third-party traffic through the household's internet connection, the device owner is unaware, and the device's IP address appears in traffic logs as a residential address indistinguishable from a human user.

In February 2026, Malwarebytes documented a campaign distributing trojanized versions of the 7-Zip archiving utility that enrolled Windows PCs into a residential proxy network [8]. The campaign demonstrates that proxyware distribution through deceptive channels continues in parallel with the commercial consent-based model, targeting the same infrastructure category and producing the same endpoint behavior.

The operational convergence is significant for enterprise defenders: an endpoint enrolled as a proxy node — regardless of whether enrollment occurred through a commercial SDK or covert malware — creates the same class of monitoring gap, and detection and network segmentation controls are substantially transferable across both categories. While threat actor motivations and incident response procedures differ between commercial SDK and botnet-enrolled devices, treating them as entirely separate threat categories creates unnecessary analytical complexity at the network layer.

5.3 Criminal Abuse and Legal Liability

The FBI's Internet Crime Complaint Center issued a public service announcement in March 2026 specifically addressing residential proxy network criminal abuse, warning consumers that their devices may be enrolled as proxy nodes and used to facilitate criminal activity [7]. The announcement notes that traffic routed through a residential proxy node originates from the enrolled device's IP address, meaning that criminal activity conducted through the network appears, in access logs and law enforcement records, to originate from the residential subscriber's address.

For enterprises, this creates a concrete legal exposure: if corporate network addresses appear in connection logs associated with criminal activity facilitated through a proxy node running on a corporate-connected device, those logs may trigger law enforcement inquiries or discovery obligations. While the enterprise would presumably not be found liable for traffic it did not knowingly enable, the investigation burden and reputational risk are real. Organizations in sectors with heightened regulatory scrutiny — financial services, healthcare, government contracting — face a particularly acute version of this exposure.

6. The Regulatory and Legal Dimension

6.1 Consent Frameworks Under Scrutiny

The commercial proxy SDK model's legal foundation is consumer consent. Operators such as Bright Data have invested in compliance infrastructure: the company publishes detailed documentation of its SDK consent requirements for partner application developers and maintains a trust center describing its ethical sourcing policies [4]. The regulatory question is whether the consent mechanisms deployed in practice satisfy the requirements of applicable privacy law.

Under the General Data Protection Regulation, consent must be freely given, specific, informed, and unambiguous [12]. The disclosure described by the Include Security researchers — a EULA navigated via a television remote control, whose text does not accurately describe the scope of the relay activity being consented to — raises substantial questions about compliance with this standard [1]. While no Data Protection Authority has ruled specifically on this consent model, the characteristics identified by Include Security are inconsistent with the informed consent standard as articulated in Article 7 and related Working Party guidance. The California Consumer Privacy Act imposes similar requirements on businesses collecting personal data from California residents. The relay activity, which involves transmitting behavioral data about a consumer's network usage patterns to a third party, arguably falls within the scope of personal data under both frameworks, though this characterization would require definitive regulatory or judicial determination.

In March 2026, the Federal Trade Commission published a policy statement applying Section 5 of the FTC Act to AI-related practices, including data collection for AI training [11]. The statement creates enforcement risk for companies whose consent mechanisms are materially misleading about the scope or purpose of data collection — a characterization that may apply to proxy SDK deployments where the disclosure understates the relay activity. The FTC has indicated that models trained on improperly collected data may be subject to remedial orders, and that civil penalties may be assessed against companies that engage in unfair or deceptive practices in AI data collection. Organizations should consult the FTC's published guidance directly for current enforcement parameters, as these are subject to revision as rulemaking and enforcement activity evolve.

6.2 Third-Party Liability and Due Diligence

Organizations that distribute applications embedding proxy SDKs — whether as enterprise-approved applications in a managed app catalog or as platform operators offering third-party applications to consumers — may face third-party liability questions. Regulatory attention to application store gatekeeping responsibilities — illustrated by the EU's Digital Markets Act and evolving FTC guidance on mobile

applications — creates a framework within which SDK-level disclosure requirements may emerge. Enterprise procurement teams and platform operators should monitor this regulatory trajectory and evaluate their current SDK governance practices against the direction of travel [17].

For enterprise IT and procurement teams, the third-party SDK risk is an instance of a broader software supply chain governance challenge. Applications approved for enterprise use undergo varying levels of security vetting, and that vetting rarely includes binary analysis for embedded SDKs that perform functions not described in the application's stated functionality. The NSA and seven allied national cybersecurity agencies specifically identified third-party services as one of six AI/ML supply chain components requiring explicit risk management in their March 2026 joint guidance [16]. Embedded SDKs fall squarely within that risk category, a threat pattern that security researchers have characterized as an escalating challenge throughout 2026 [17].

6.3 Data Residency and Cross-Border Considerations

The geographic bandwidth tier configuration documented in the Include Security research reveals that the SDK's operational parameters vary by country [1]. Devices in different jurisdictions relay different volumes of traffic under different conditions. This variability reflects Bright Data's management of its proxy supply relative to customer demand by geography — but it also means that the cross-border data flow implications of SDK enrollment are jurisdiction-specific.

Organizations operating in the European Union, where cross-border data transfers to third countries require legal basis under the GDPR's Chapter V, should evaluate whether employee device enrollment in commercial proxy networks constitutes a cross-border data transfer subject to those requirements. Given that the relay traffic originates from the enrolled device's network connection, and that device telemetry is transmitted to Bright Data's infrastructure, there are grounds for arguing that the SDK creates data flows requiring legal basis analysis under Article 46.

7. Detection, Monitoring, and Mitigation

7.1 Network-Level Controls

The most accessible detection and blocking mechanism for organizations is DNS-based blocking of the domains associated with Bright Data's SDK infrastructure. The Include Security researchers identified the following domains as control and relay endpoints:

- `proxyjs.brdtnet.com` – primary WebSocket command and control endpoint
- `proxyjs.luminatinet.com` – legacy control endpoint under prior company identity
- `proxyjs.bright-sdk.com` – additional control endpoint
- `clientsdk.bright-sdk.com` – unauthenticated configuration endpoint
- `clientsdk.brdtnet.com` – configuration endpoint under current branding [1]

Blocking these domains at the corporate DNS resolver or outbound firewall will prevent SDK-embedded applications on corporate-connected devices from establishing control channels. This control is effective for devices connected to the corporate network, including smart televisions on corporate WiFi, and for devices that route DNS through the corporate resolver via VPN. It is not effective for the VPN bypass traffic class identified in the research, which exits the device through the physical network interface and bypasses the corporate DNS resolver.

Organizations should also consider blocking these domains in their secure DNS services and threat intelligence platform enrichment pipelines, so that any resolution of these domains by devices on corporate networks generates a security alert. The appearance of these domains in outbound traffic logs is a reliable indicator of SDK-enrolled device presence.

7.2 Endpoint and Application Controls

For managed device fleets, mobile device management platforms should be evaluated for their ability to detect applications embedding proxy SDKs. The binary-level indicators identified by Include Security – the Swift symbols `BrdWebSocketFacade` and `BrdNetwork.DNSResolver` within iOS application binaries – provide reliable detection signatures [1]. Organizations with mobile threat defense solutions should evaluate whether those solutions have added detection for these indicators following the June 2026 disclosure.

For devices enrolled in full MDM, application allowlist policies that restrict installation to approved applications provide a more comprehensive control than reactive detection. Any application not in the approved catalog should be blocked, regardless of whether it is known to embed proxy infrastructure. This approach requires ongoing catalog maintenance but eliminates the detection gap created by novel SDK-embedded applications for which binary signatures do not yet exist.

For corporate smart television deployments, organizations should evaluate whether those devices require app store access at all, and consider network segmentation that isolates televisions from the corporate network. Smart televisions used for conference room video conferencing can often be configured to connect only to the video conferencing service's endpoints, with all other outbound connections blocked at the network level. This architectural approach eliminates the television as a proxy ingress point regardless of which applications are installed.

7.3 Software Composition Analysis

The proxy SDK risk is an instance of the broader third-party software component risk that software composition analysis tools are designed to address. Organizations that internally develop or review applications for enterprise distribution should incorporate SDK-level analysis into their security review process, specifically examining applications for embedded network relay components alongside the more commonly evaluated dependencies for known vulnerabilities.

For enterprise application stores and internal application distribution platforms, the same scrutiny should apply to third-party applications. Application vetting processes should include SDK composition review, not merely application-level functionality review. Platform operators — mobile operating system vendors and connected TV platform providers — are positioned to impose SDK-level disclosure requirements on application developers, and several have done so in specific cases. Enterprise procurement teams should favor platforms with active SDK governance programs when selecting collaboration tools, streaming services, and entertainment applications for corporate distribution.

7.4 Security Awareness and Policy

Technical controls are most effective when complemented by policy that defines acceptable use and supports employee reporting. Enterprise acceptable use policies should explicitly address the installation of applications that commoditize device bandwidth without organizational consent. Employees who are aware that such applications exist and understand the risks they create are more likely to report suspicious network behavior and to make informed choices about application installation on devices that connect to corporate systems.

Security awareness programs covering AI data supply chain risks should address the specific scenario of consumer applications embedding proxy functionality, rather than confining instruction to more familiar threat categories. The Include Security research illustrates that the risk manifests in application categories — casual games, messaging applications, entertainment apps — that employees are unlikely to associate with enterprise security risk.

8. Conclusions and Strategic Recommendations

The AI data supply chain has constructed residential proxy infrastructure at a scale and technical sophistication that now presents a measurable enterprise security challenge. The combination of widespread SDK deployment, VPN bypass capability, and convergence with criminal proxy networks creates a threat category that existing enterprise security architectures were not designed to address and that most current security programs have not incorporated into their threat models.

Several strategic conclusions follow from the analysis in this paper.

Enterprise network perimeter assumptions need updating. The VPN bypass technique documented in the Include Security research is not a vulnerability in Apple's platform — it is a documented, supported API feature that any application can use. Security architectures that rely on VPN-based monitoring for BYOD device traffic should account for application-level VPN bypass techniques, which the Bright Data SDK demonstrates are practical and commercially deployed at scale. The SDK's traffic class is invisible to VPN-based inspection — a monitoring gap that may be larger than security teams have previously assessed.

Smart television deployments warrant a security category revision. Connected televisions in corporate environments should be treated as unmanaged endpoints equivalent to IoT devices, subject to network segmentation, outbound traffic filtering, and application restriction policies. The characteristics that make smart TVs appealing to proxy network operators — constant power, persistent connectivity, absence of endpoint security agents — are also characteristics that make them poorly suited to corporate network trust zones without compensating controls.

Software composition analysis should extend to SDK-level review. Application approval processes should examine embedded SDKs, not merely application-declared functionality. The Bright Data example is illustrative but not unique: the commercial proxy SDK market includes multiple providers with similar technical architectures, and new entrants may use different binary signatures than those currently documented.

The consent model for proxy SDKs is a regulatory inflection point. The June 2026 disclosures occurred against a backdrop of active FTC enforcement posture, evolving GDPR interpretation, and renewed regulatory attention to AI data collection practices. Organizations that distribute consumer applications — either directly or through enterprise application catalogs — should evaluate their SDK governance against current regulatory expectations rather than the more permissive standards of prior years.

Criminal and commercial proxy infrastructure convergence warrants a coordinated detection approach. From a network security standpoint, an endpoint enrolled as a proxy node — regardless of whether enrollment occurred through a commercial SDK or covert malware — creates the same class of monitoring

gap: outbound traffic that bypasses enterprise controls. The Bitsight research establishes that commercial and criminal proxy nodes share infrastructure and that significant portions of commercial residential proxy networks are simultaneously compromised by malware [5]. Detection and network segmentation controls are therefore substantially transferable across both categories, even though threat actor motivations and incident response procedures differ.

9. CSA Resource Alignment

The threats documented in this paper engage several dimensions of existing Cloud Security Alliance frameworks and guidance.

MAESTRO (Multi-Agent Environment, Security, Threat, Risk, and Outcome). CSA's agentic AI threat modeling framework provides a structured methodology for analyzing threats to AI systems at the supply chain layer [15]. In applying MAESTRO's framework, the proxy SDK threat aligns most closely with Layer 1 (Foundation Models) and Layer 7 (Agent Ecosystem) concerns: the data collected through proxy infrastructure shapes foundation model training quality and provenance, while the SDK ecosystem represents a supply chain trust boundary that current threat models rarely address. Organizations applying MAESTRO to their AI adoption risk assessments should incorporate data supply chain infrastructure as a threat surface [15].

AI/ML Supply Chain Security Guidance. In March 2026, CSA synthesized the joint guidance published by the NSA and seven allied national cybersecurity agencies on AI/ML supply chain risk management [16]. That guidance defines a six-component AI supply chain — training data, models, software, infrastructure, hardware, and third-party services — and maps threat classes to each component. The residential proxy SDK vector represents a training data supply chain risk (data collected through inadequate consent mechanisms) and a third-party services risk (SDK embedded in consumer applications without enterprise visibility). Organizations implementing the allied guidance's recommended controls should evaluate their data procurement practices against this risk category.

Agentic Trust Framework and Zero Trust for AI. CSA's Agentic Trust Framework applies Zero Trust principles to AI agent environments, emphasizing that trust must be continuously verified rather than assumed based on network position or prior authentication [18]. This principle applies directly to the proxy SDK context: the presence of a device on the corporate network, or its enrollment in a corporate MDM system, does not establish that the device's traffic is contained within enterprise-approved channels. Zero Trust architectures that enforce per-connection policy based on application identity — rather than network position — are more resilient against the VPN bypass techniques documented in this paper.

Cloud Controls Matrix (CCM). Several CCM control domains are directly relevant to proxy SDK risk management. The Supply Chain Management domain (STA-09 through STA-12) addresses third-party software component risk and vendor due diligence requirements. The Infrastructure and Virtualization Security domain addresses network segmentation controls applicable to IoT and unmanaged endpoint categories, including smart televisions. The Application and Interface Security domain (AIS) addresses software composition and security testing requirements for distributed applications.

STAR (Security Trust Assurance and Risk) Program. Organizations procuring applications or platforms for enterprise use should evaluate whether those vendors have completed a CSA STAR self-assessment or certification, and whether that assessment addresses third-party SDK governance. As the proxy SDK risk becomes better understood, STAR questionnaires should be updated to include explicit inquiry into SDK disclosure practices and third-party bandwidth monetization.

References

[1] Include Security. "The Smart TV in Your Living Room Is a Node in the AI-Scraping Economy." Include Security Research Blog, June 2026.

[2] The Hacker News. "Free Apps Are Quietly Turning Smart TVs Into Web-Scraping Proxies for AI." The Hacker News, June 2026.

[3] CybersecurityNews. "Free Apps on Samsung and LG Smart TVs Secretly Turning Your Devices Into AI Proxies." CybersecurityNews, June 2026.

[4] Bright Data. "Ethically Sourcing Residential Proxies." Bright Data Trust Center, 2026.

[5] Bitsight Technologies. "Residential Proxy Services and Malware Ecosystems." Bitsight Research, 2026.

[6] The Hacker News. "Google Disrupts IPIDEA — One of the World's Largest Residential Proxy Networks." The Hacker News, January 2026.

[7] Federal Bureau of Investigation Internet Crime Complaint Center. "Evading Residential Proxy Networks: Protecting Your Devices from Becoming a Tool for Criminals." IC3 Public Service Announcement PSA260312, March 12, 2026.

[8] Malwarebytes Threat Intelligence. "Fake 7-Zip Downloads Are Turning Home PCs into Proxy Nodes." Malwarebytes, February 2026.

[9] Spamhaus. "Let's Talk About the Danger of Residential Proxy Networks." Spamhaus Resource Hub, 2026.

[10] GreyNoise Intelligence. "New Report Points to a Significant Number of Compromised Residential IP Addresses." GreyNoise Intelligence Press Release, 2026.

[11] Hash Scraper. "FTC AI Policy Statement March 2026 — 5 Regulatory Areas Web Crawling Companies Should Know." Hash Scraper Technology Blog, March 2026.

[12] International Association of Privacy Professionals. "Training AI on Personal Data Scraped from the Web." IAPP, 2023.

[13] Lowpass. "Your Smart TV May Be Scraping the Web for AI." Lowpass, June 2026.

[14] TechSpot. "Smart TV Apps Are Quietly Scraping Web Data for AI Training." TechSpot, June 2026.

[15] Cloud Security Alliance. "Agentic AI Threat Modeling Framework: MAESTRO." CSA Blog, February 2025.

[16] Cloud Security Alliance. "Eight-Nation AI/ML Supply Chain Risk and Mitigation Guidance." CSA Lab Space, March 2026.

[17] eSecurity Planet. "AI Software Supply Chain Threats Escalate in 2026." eSecurity Planet, 2026.

[18] Cloud Security Alliance. "The Agentic Trust Framework: Zero Trust Governance for AI Agents." CSA Blog, February 2026.

[19] Bright Data. "Introduction to Residential Proxies." Bright Data Documentation, 2026.

[20] Orange Cyberdefense. "Uncovering Residential Proxy Providers: Risks and Market Insights." Orange Cyberdefense Research, 2026.

[21] QPulse / Quasar CyberTech. "Consumer Apps Embedding Bright Data SDK Turn Smart TVs and Mobile Devices into Residential Proxies." QPulse, June 2026.