ARMATURE
SYSTEMS

JUL — SEP 2026

Global Cyber Threat Landscape Report

© 2026 Armature Systems Inc.

Table of Contents

1 EXECUTIVE SUMMARY 2

2 THREAT ACTOR TRENDS 5

3 ARMATURE THREAT RESEARCH 7

4 ATTACK VECTORS, TECHNIQUES & EXPLOITATION TRENDS 8

5 EMERGING THREATS & TECHNOLOGY RISKS 13

6 RECOMMENDATIONS 18

APPENDIX 20

© 2026 Armature Systems Inc.

1 Executive Summary

Reporting period: July 1 – September 30, 2026.

Last quarter we asked why attackers would bother with a backdoor when the front door was open. In Q3 they kept walking through it, and started eyeing the new wing we're all building:

- **Edge devices keep falling.** According to Recorded Future, 142 of the 215 vulnerabilities actively exploited in the first half of 2026 could be triggered remotely with no authentication. In Q3 alone, attackers exploited internet-facing security devices from several vendors, including firewalls, VPNs and management consoles.
- **Identity attacks surged.** Vishing intrusions doubled and device-code phishing attempts rose fifteen-fold (CrowdStrike). Authentication abuse appeared in 65% of Cisco Talos incident-response engagements in Q2, up from 35% in Q1. We've reported on this trend in previous assessments: ShinyHunters, Scattered Spider and similar threat actors keep picking up the phone and calling their targets' help desks.
- **The vulnerability backlog keeps growing.** Published vulnerabilities rose more than 40% in the first half of 2026 (VulnCheck), but only about 1.4% are known to have been exploited. Engineering teams are buried in thousands of critical and high findings while trying to meet SLAs and juggle other priorities. The goal is to fix what attackers can actually reach, but that analysis takes time and effort. Monitoring CISA's Known Exploited Vulnerabilities (KEV) catalog and reviewing exploit chains, where attackers combine several medium or low severity flaws into a critical attack path, are a must.

And, of course, all of this is happening against the backdrop of a powerful innovation cycle in AI. Almost every company we've worked with this year is pushing forward with AI engineering initiatives, to enable business activities, develop new or updated products, or both. AI takes people, infrastructure, tooling, training, and policies and procedures, at the cost of increased systemic complexity and risk.

If this quarter has a headline, that's it. The complexity, and the relatively new governance controls around it, have become a target. Attackers aren't just using AI to attack; they are actively targeting AI infrastructure, including AI gateways, MCP servers, identities, APIs, and even the instruction files AI agents read, like CLAUDE.md and skills. When you think about it, the approach makes sense. Agents are often given significant access to data, systems and tooling, so compromising the agent means inheriting that access.

© 2026 Armature Systems Inc.

It's no surprise, then, that many in our community have shifted strategy. "Exploitability," "AI governance," and "agentic remediation" are now top of mind, driven by AI enabled threat actors adding to an already overwhelming backlog and by continued business pressure to adopt AI.

The organizations that will succeed in protecting themselves this year won't be the ones that patch the most or enable AI services the fastest. They'll be the ones that know what is actually exposed, get in front of systemic risk in AI, govern their AI infrastructure like other complex systems in their ecosystem, and treat their identity infrastructure as the new front line that it is.

Where to start

- **Prioritize by exploitability, not volume.** Focus vulnerability remediation on what is reachable, on the KEV catalog, or part of a known exploit chain. Look for workarounds where patching isn't practical, such as configuration changes or network / permission level compensating controls.
- **Treat edge devices as crown jewels.** Take management interfaces off the internet, and check exposed devices for compromises. In at least a few cases this year, the device didn't log the malicious authentication event at all, meaning your first indication of compromise is traffic already on your network.
- **Harden the help desk.** Require call back verification for MFA and password resets, use phishing-resistant MFA, and block device-code sign-in where it isn't needed. ID verification technology can also help here, especially at scale.
- **Put AI under the same controls as any privileged system.** Inventory AI gateways, MCP servers and agents, give them least privilege, and review AI instruction files like code. Many new and existing vendors are actively developing new AI governance technologies as well and are worth a look.

Threat Landscape at a Glance

The Q3 2026 landscape can be summarized through six major developments:

AREA	Q3 2026 OBSERVATION	ASSESSMENT
AI enabled attacks	AI increasingly used as an operational layer rather than simply a productivity tool	Rapidly developing
Identity attacks	Credential theft, MFA abuse, vishing and cloud account compromise remain central	Very high concern
Edge device exploitation	VPNs, firewalls and security management appliances repeatedly targeted	Very high concern

© 2026 Armature Systems Inc.

AREA	Q3 2026 OBSERVATION	ASSESSMENT
Ransomware	Ransomware groups continue combining stolen credentials, vulnerability exploitation and data theft	Persistent
Supply chain	Open source AI tooling and developer environments increasingly targeted	Growing
AI infrastructure	AI gateways, RAG systems, orchestration platforms and model environments becoming attack targets	Actively targeted; this quarter's headline

For example, an attacker may obtain credentials through phishing, use those credentials to enter Microsoft 365, compromise a developer account, access a GitHub repository, steal cloud credentials, compromise an AI workload and then use AI automation to process stolen data. The traditional distinction between phishing, cloud compromise, supply chain attacks and AI attacks is therefore becoming less useful.

Last Quarter's Calls

How our Q2 2026 calls held up in Q3.

Q2 CALL	Q3 EVIDENCE	SCORE
ShinyHunters will continue to be a threat in Q3	Baxter confirmed unauthorized access to third-party apps (Aug 13; the 7.1M-record figure is claimed); McKesson disclosed a breach on Aug 28 that ShinyHunters claims it carried out through help-desk vishing; the FBI said it was investigating ShinyHunters' claims involving FBIJobs.gov (late Sep); Mandiant reported a WAF bypass for the patched PeopleSoft flaw (Sep 25)	Held
OT targeting by Iran-nexus actors is the most strategically concerning trend	More than 30 Minnesota water systems were hit Jul 26–28, within attacks on water systems in at least seven states; no public US attribution	Likely (no hard evidence but follows attack pattern)

© 2026 Armature Systems Inc.

2 Threat Actor Trends

The groups below were the most operationally significant in Q3 2026. Each entry ends with who should pay closest attention. Leak-site victim counts are actor claims.

The clearest pattern of the quarter: criminal and state actors went after the same appliances. A Qilin affiliate and a cluster with Sandworm overlaps both exploited Cisco FMC flaws to get inside firewall management.

1. Qilin (Agenda) | Highest-Volume Ransomware Brand

Qilin remained the highest-volume ransomware brand: 119 leak-site victims in July (tied with The Gentlemen, per Breachsense) and 145–165 in August depending on the tracker (Cyble, ZeroFox). Its affiliates used edge appliances as a way in, including Cisco FMC CVE-2026-20316 (Talos cluster UAT-11988), followed by SOCKS5 proxies, reverse SSH tunnels, Impacket and EDR killers.

Relevant to you if: you run Cisco FMC, GlobalProtect or other VPN and firewall management interfaces reachable from the internet.

2. ShinyHunters | SaaS Data Theft and Extortion at Scale

ShinyHunters stayed active across SaaS data theft and extortion. Baxter confirmed unauthorized activity in third-party apps in August; the group claimed about 7.1 million Salesforce records. McKesson disclosed on August 28 that attackers had taken data through third-party applications; ShinyHunters claims it used help-desk vishing to compromise Okta accounts and reach Salesforce and Snowflake, and claims 284 million patient records, which McKesson has not confirmed. In late September the FBI said it was aware of and investigating the group's claims involving the FBIJobs.gov portal. Mandiant (tracking the actor as UNC6240) reported on September 25 a WAF bypass that reopened the patched Oracle PeopleSoft flaw CVE-2026-35273. Dutch police made an arrest on September 28 in a related investigation; ShinyHunters denies the suspect is a member.

Relevant to you if: you hold customer data in Salesforce or other SaaS with third-party app integrations, or run Oracle PeopleSoft.

3. Sandworm-linked activity (UAT-11823) | State Actor Inside Firewall Management

Cisco Talos reported on September 9 a state-sponsored cluster exploiting both Cisco FMC flaws to gain a root reverse shell, collect configuration data from managed devices and deploy a Cyclops Blink variant. Talos assessed overlaps with Sandworm, a Russian state-sponsored group.

© 2026 Armature Systems Inc.

Relevant to you if: you manage firewalls centrally through Cisco FMC, especially in critical infrastructure.

4. Iran-nexus actors | Pressure on Water Systems

More than 30 Minnesota water systems were disrupted July 26–28, part of attacks the FBI and EPA said hit water systems in at least seven states. The US government has not publicly attributed them; later reporting says Iran-linked groups, including CyberAv3ngers, claimed responsibility (unverified). On September 17 Group-IB tied the Handala persona to HEAVYGRAM, a Telegram-based backdoor used by the MOIS-linked Void Manticore against dissidents and journalists. We found no Q3 activity from Pay2Key.I2P.

Relevant to you if: you operate water, energy or manufacturing OT with internet-exposed PLCs or remote access.

5. China-nexus espionage | Shared Zero-Day Chains

Three China-linked groups (UTA0560, JungleBamboo/APT31 and UTA0565) used the same Chrome-to-Windows zero-day chain in early September against US NGOs and Asian government targets, according to Volexity (see Shared exploit chains).

Relevant to you if: your staff work with NGOs, government or policy organizations, or your browser updates lag.

6. Other ransomware and extortion groups

INC Ransomware affiliates exploited SonicWall SMA1000 appliances (Resecurity, August 3). The Gentlemen, Akira and DragonForce continued posting victims, including claimed listings of Mercado Libre (The Gentlemen, July 8) and TUI China (DragonForce, August). Cl0p exploited PTC Windchill CVE-2026-12569 from late July and named more than 40 alleged victims from August 12. Alleged Scattered Spider member Peter Stokes was extradited from Finland and appeared in US federal court around July 1.

Relevant to you if: you run SonicWall SMA1000 or PTC Windchill / FlexPLM.

Sources: Breachsense, Cyble and ZeroFox (ransomware tracking); Cisco Talos; Mandiant; BleepingComputer; Group-IB; NBC News and Fortune; Volexity; Resecurity; SecurityWeek.

© 2026 Armature Systems Inc.

3 Armature Threat Research – Exploitability

Narrowing 30,000 Findings to the Ones That Matter

Q3 made the backlog problem worse and the answer clearer. Disclosures kept surging while the share of vulnerabilities actually exploited stayed small, so the work is not patching more; it is finding the few that matter and closing them faster. This section sets out the approach we use with clients to do that.

Many of the clients we speak with regularly report 30,000 or more raw High and Critical findings, and worry that AI driven discovery will make it worse. That list cannot be patched in full, and it does not need to be. Only about 1.4% of disclosed vulnerabilities are known to be exploited (Root Evidence, May 2026), and AI-discovered vulnerabilities are exploited at the same rate as all others (VulnCheck). The challenge is moving fast on the few that matter: SAP patched CVE-2026-58231 on August 11 and attacks began on August 14.

By exploitability we mean how likely a flaw is to be attacked where it actually sits in your environment: whether it is reachable, exposed, and has known exploit activity. This is different from the CVSS Exploitability subscore.

STEP	QUESTION	WHAT Q3 SHOWED
1. Deduplicate	How many unique fixes are behind the findings?	One CVE across many hosts, or one product upgrade, is a single fix
2. Known exploited	Is it on CISA KEV or under active exploitation? If not, rank it with EPSS.	Public exploit code is a weaker evidence than a confirmed exploitation
3. Exploitability	Is it reachable and exposed where it sits, alone or chained with other flaws?	The exploited edge instances in this report were internet-reachable; Microsoft saw two LiteLLM flaws chained into unauthenticated code execution
4. Fix by configuration	Can a configuration change close it today?	GlobalProtect CVE-2026-0257 (Q2 intrusions) needed authentication override cookies; F5 CVE-2026-94127 needed APM acting as an OAuth server; Citrix CVE-2026-88772 needed DTLS enabled
5. Patch and verify	Is the fix tested and confirmed?	1Password's Off-by-1 Labs (Aug 6) found only about one in four AI-generated patches was a clean fix

Two cautions. First, 23% of exploited vulnerabilities were attacked on or before the day they were disclosed (VulnCheck), so pair the funnel with reducing what is exposed at all; zero-days will not be on any list. Second, internal findings still matter once an attacker is inside, and every configuration workaround needs an owner and a date to cross check, because upgrades and configuration drift can re-expose. Compensating controls can fail too: in September Mandiant reported a WAF bypass that reopened a patched PeopleSoft flaw.

Sources: Armature Threat Research; VulnCheck; Root Evidence; Off-by-1 Labs (1Password).

© 2026 Armature Systems Inc.

4 Attack Vectors, Techniques & Exploitation Trends

In attack techniques, Q3 was about continuity: attackers kept using what already works: internet-facing security appliances, the help desk, and the packages and tools developers depend on. Edge-device exploitation continued at pace, often as zero-days, while identity attacks moved further from the inbox to the phone.

Security Appliances: Still the Leading Way In

Q3 confirmed that internet-facing security appliances remain a leading way in; Verizon's 2026 DBIR (2025 data) ranks vulnerability exploitation as the top initial access vector. Citrix, Cisco, SonicWall, Check Point, F5, Fortinet and Arista all had products exploited or added to CISA's Known Exploited Vulnerabilities (KEV) catalog during the quarter. Cisco disclosed exploitation of Secure Firewall Management Center CVE-2026-20316 on July 29 and became aware of CVE-2026-20079 exploitation in August; on September 9 Cisco Talos tied the activity to three clusters, including a Qilin affiliate (UAT-11988) and a state-sponsored cluster (UAT-11823) deploying a Cyclops Blink variant with overlaps to Russia's Sandworm. SonicWall SMA1000 flaws CVE-2026-15409 and CVE-2026-15410 were patched and added to KEV on July 14, linked to an INC Ransomware affiliate on August 3, and followed by two new SMA1000 zero-days (CVE-2026-83548 and CVE-2026-83549) on September 1. Check Point disclosed on September 22 a management-server zero-day (CVE-2026-93616) exploited since July 23, two months before a patch existed. Citrix NetScaler ended the quarter with zero-days CVE-2026-88771 and CVE-2026-88772, patched and added to KEV on September 27. Arctic Wolf's July 20 report on Qilin affiliates exploiting GlobalProtect CVE-2026-0257 describes June (Q2) intrusions, but Arctic Wolf assessed exploitation was likely ongoing.

Armature's MDR team saw the same pattern in its own casework this year: in some intrusions the compromised edge device never logged the malicious authentication, so the first sign of compromise was attacker traffic already moving inside the network.

Exploited vulnerabilities, Q3 2026: edge devices, browsers and other notable flaws

CVE	PRODUCT / FLAW	Q3 DATES	ZERO-DAY?	USED BY
CVE-2026-88771, -88772	Citrix NetScaler ADC / Gateway RCE	Exploited for weeks before patch; patch and KEV Sep 27	Yes	Not public
CVE-2026-93616	Check Point Management / Log Server path traversal	Exploited from Jul 23; disclosed, patched and KEV Sep 22	Yes	Not public

© 2026 Armature Systems Inc.

CVE	PRODUCT / FLAW	Q3 DATES	ZERO-DAY?	USED BY
CVE-2026-85102	Check Point Security Gateway / Spark VPN	Disclosed Sep 9; exploited from Sep 12; KEV Sep 22	No	Not public
CVE-2026-94127	F5 BIG-IP APM (OAuth server) RCE	Exploited before patch; KEV Sep 22	Yes	Not public
CVE-2026-93952	Arista VeloCloud Orchestrator	KEV Sep 22	Not confirmed	Not public
CVE-2026-83548, -83549	SonicWall SMA1000 SSRF and command injection	Patched Sep 1; KEV Sep 2	Yes	Not public
CVE-2026-15409, -15410	SonicWall SMA1000 SSRF and command injection	Patch and KEV Jul 14; ransomware link Aug 3 (exploited since Jun 22)	Yes	UTA0533; INC affiliate
CVE-2026-20079	Cisco Secure FMC authentication bypass (CVSS 10.0)	Exploitation noticed Aug; Talos report and KEV Sep 9	No (patched Mar 2026)	UAT-11823 (Sandworm overlap), UAT-12197
CVE-2026-20316	Cisco Secure FMC static credential	Disclosed as exploited Jul 29; Talos Sep 9	Not confirmed	UAT-11988 (Qilin), UAT-11823
CVE-2026-19490	Citrix NetScaler authentication bypass	Exploit attempts from Sep 3; KEV Sep 9	Not confirmed	Not public
CVE-2025-25249	Fortinet FortiOS heap overflow RCE	Exploited since July; KEV Sep 9	No	Not public
CVE-2026-8452	Citrix NetScaler (DoS bug later shown to allow RCE)	PoC Aug 14; web shells seen; KEV Aug 26	No (patched Jun 30)	Not public
CVE-2026-59822	LiteLLM MCP authentication bypass	KEV Sep 2; Wiz details Sep	Not confirmed	Not public
CVE-2026-12569	PTC Windchill / FlexPLM unauthenticated RCE	KEV June; Cl0p exploitation from late Jul; victims named from Aug 12	No	Cl0p
CVE-2026-68820	Windows AFD.sys privilege escalation	Exploited from early Jul; patched Aug 11	Yes	Lazarus (FudModule rootkit)

© 2026 Armature Systems Inc.

CVE	PRODUCT / FLAW	Q3 DATES	ZERO-DAY?	USED BY
CVE-2026-58231	SAP Commerce Cloud unauthenticated RCE (CVSS 10)	Patched Aug 11; attacks Aug 14	No	Not public
CVE-2026-85046	Chrome V8 type confusion	Exploited Sep 1–2; Chrome patch Sep 3–4	Patch gap (fixed in Chromium first)	UTA0560, JungleBamboo (APT31), UTA0565

Sources: BleepingComputer, Cisco Talos, eSecurity Planet, eSentire, Help Net Security, Rapid7, Resecurity, SecurityWeek, SOCRadar, Tenable, The Hacker News, Volexity.

The Patch Window Is Collapsing

Speed now matters more than volume: SAP patched CVE-2026-58231 on August 11 and attacks began on August 14. When a patch is not available or cannot be applied in time, prioritize internet-exposed assets under active exploitation and use compensating controls such as network restrictions, virtual patching or isolation, then verify they hold.

The volume adds pressure. Microsoft's Patch Tuesday counts were about 570–620 CVEs in July, 400–420 in August and 965–975 in September; the ranges reflect whether outlets count Chromium/Edge and earlier out-of-band fixes. VulnCheck's first-half 2026 report (published July 28) counted 495 newly known-exploited CVEs (a broader count than the 215 Recorded Future reports, because the two define exploitation evidence differently), 23.4% of them exploited on or before the day they were disclosed.

Identity-Led Intrusions

Identity remained a primary target, with attackers calling, not just emailing, their way in:

- **Help-desk vishing:** Google tracked UNC6671 calling US private equity and hedge fund staff as IT support to capture MFA codes in real time, followed by ransom demands of up to $3 million (reported early August).
- **Executive impersonation:** Arctic Wolf described PREY-0058 (September 7) placing fake IT calls to executives, then using adversary-in-the-middle phishing to replay session tokens and steal Microsoft 365 and Box data.
- **Device-code phishing kits:** the DEBULL and ARToken kits (reported July 7) and the CSuite kit (September 30) package the OAuth device-code technique Armature profiled in Q2, making it available to less skilled actors.
- **Takedown:** Microsoft's Digital Crimes Unit announced on September 22 that it had disrupted EvilTokens, a device-code phishing service run by Storm-2992 that had compromised more than 12,000 inboxes at over 10,000 organizations.

© 2026 Armature Systems Inc.

Shared Exploit Chains

Volexity reported on September 9 and September 21 that at least three China-linked groups (UTA0560, JungleBamboo/APT31 and UTA0565) used the same Chrome-to-Windows zero-day chain (CVE-2026-85046, CVE-2026-87491 and Windows CVE-2026-85880) within days of each other. That suggests offensive capabilities are being shared, outsourced or sold rather than developed independently by each group.

Developer and AI Tooling Supply Chain Attacks

Q3 brought a steady run of attacks on the packages and extensions developers and AI tools depend on:

- **July 11:** Jscrambler npm packages published with a stolen token carried an infostealer aimed at Claude Desktop, Cursor and Windsurf credential stores.
- **July 14:** AsyncAPI's CI pipeline was compromised, pushing a remote-access trojan in four packages with about 2 million weekly downloads, live for roughly one to four hours depending on the package.
- **July 26 – August 1:** 77 look-alike Open VSX extensions sent developer, Git and CI metadata to attackers (Manifold).
- **August 4:** the Shai-Hulud worm spread through a compromised keyv maintainer account to hundreds of npm packages (counts range from about 440 packages to more than 1,300 package versions) with over 2 billion monthly downloads, stealing AI-tool API keys and planting configuration files in repositories (OX Security).
- **August 5:** Backslash Security documented Shai-Hulud variants using AI coding-assistant configuration files for persistence.
- **August 26:** Australian police arrested alleged TeamPCP members linked to earlier compromises of Trivy, LiteLLM and other projects.
- **September 8:** Google reported DUSTMAKER malware hiding its configuration in .claude/, .cursor/ and .vscode/ folders and carrying prompt injection aimed at LLM-based security scanners.

Dominant Techniques (MITRE ATT&CK)

The techniques below were the most consistently observed across criminal and state operations this quarter.

TECHNIQUE (ATT&CK ID)	Q3 2026 OBSERVATION
Exploit Public-Facing Application (T1190)	Leading way in across edge appliances: Citrix NetScaler, Cisco FMC, SonicWall SMA1000, Check Point, F5 and Fortinet
Phishing: Spearphishing Voice (T1566.004)	Help-desk and executive vishing (UNC6671, PREY-0058, ShinyHunters)
Steal Application Access Token (T1528)	Device-code phishing kits (DEBULL, ARToken, CSuite) and the EvilTokens service
Valid Accounts (T1078)	Stolen credentials and tokens used to reach SaaS data (Salesforce, Snowflake, Microsoft 365, Box)

© 2026 Armature Systems Inc.

TECHNIQUE (ATT&CK ID)	Q3 2026 OBSERVATION
Supply Chain Compromise: Software Supply Chain (T1195.002)	npm and Open VSX compromises (Shai-Hulud, AsyncAPI, Jscrambler; 77 look-alike Open VSX extensions)
Unsecured Credentials: Credentials In Files (T1552.001)	Theft of AI-tool and model-provider API keys from developer machines and AI gateways

Sources: Armature Threat Research; CISA KEV catalog; Verizon 2026 DBIR; Arctic Wolf; Backslash Security; Google Threat Intelligence Group; Microsoft; OX Security, Manifold, Socket and StepSecurity; Volexity.

© 2026 Armature Systems Inc.

5 Emerging Threats & Technology Risks

Three AI developments defined the emerging-threat picture this quarter. First, attackers used AI inside live operations, though humans still chose the targets. Second, AI infrastructure itself became a target: gateways, RAG platforms, MCP servers and the instruction files agents read. Third, AI systems under test reached real third-party systems with no attacker directing them, and frontier labs slowed down in response.

AI Is Speeding Up Human-Led Operations

Q3 brought clearer evidence that attackers use AI inside live operations, though humans still choose the targets.

Q3 evidence: Anthropic's September 10 threat report described actors using AI agents across live operations, including a Russian espionage group (GTG-20006) whose agents rebuilt malware after detection, and a ShinyHunters-linked actor (GTG-50014) that let AI run scripts autonomously during data theft. The cases span December 2025 to August 2026. Google's September 8 report (covering roughly Q2 activity) described an agent setup that harvested thousands of credentials in under six hours.

This creates a problem for defenders because many security operations processes remain designed around human attacker speed. The question is increasingly not whether an attacker is technically advanced, but whether the organization can detect and interrupt activity quickly enough.

Hype check. AI is not plotting against you. Every AI related incident is explained by human direction or by systems optimizing for the objective the agent was told to execute. Google (September 8) has "not yet observed threat actors deploying fully autonomous pipelines against targets in the wild," and Anthropic (September 10) found humans still chose the targets. Of 1,061 AI discovered vulnerabilities in the first half of 2026, 1.3% were exploited, about the same rate as all vulnerabilities (VulnCheck). Anthropic's restricted Mythos model and Project Glasswing produced 23,019 findings, 126 CVEs and one known exploitation in the first half of 2026.

AI Infrastructure Is Now a Target

A particularly important development during Q3 is the emergence of AI infrastructure as an enterprise attack surface. Microsoft reported attacks against AI infrastructure components including LiteLLM gateways, RAGFlow and Kestra. These components can sit between users, models, data sources, credentials and execution environments.

© 2026 Armature Systems Inc.

In its August 26 report, Microsoft described attackers chaining LiteLLM flaws (CVE-2026-42271 and CVE-2026-48710) into unauthenticated remote code execution, then taking model-provider API keys, the master key and database credentials, and running XMRig cryptominers. RAGFlow intrusions captured LLM keys through startup hooks; Kestra intrusions used a login bypass (CVE-2026-49869) to read container secrets. Google separately reported trojanized MCP server packages published by the TeamPCP group (UNC6780).

Armature's MDR team saw this directly in Q3. In a supply chain intrusion at a client, the attacker modified AI coding-assistant files on developer machines (CLAUDE.md instruction files, hooks in assistant settings and MCP server entries) so the assistant would run attacker instructions as part of normal work. Compromising the files the agent reads meant inheriting the agent's access.

AI gateways and the governance market

AI gateways sit between users, agents, models and credentials, which makes them the new control point and the new target. Beyond the exploitation Microsoft described in August, Wiz detailed further LiteLLM flaws in September, including an MCP authentication bypass (CVE-2026-59822), which CISA added to KEV on September 2, and a guardrail sandbox escape giving root-level code execution (CVE-2026-40217); the fix is version 1.84.0 or later. Wiz found that nearly one in ten exposed LiteLLM instances (294 of 3,074) still accepted the default master key, and estimated LiteLLM runs in about a third of cloud environments.

Many AI guardrail products integrate with, embed or fork an open-source gateway such as LiteLLM, so the guardrail is only as strong as the gateway beneath it. Ask any vendor which open-source components and versions sit underneath, and how quickly they ship upstream security fixes.

Q3 saw rapid growth in AI governance products, built around four control points:

APPROACH	WHAT IT CONTROLS	EXAMPLES	BLIND SPOT
Network	AI traffic that crosses the corporate network or cloud proxy	Netskope, Zscaler	Agents and models running entirely on a laptop
Gateway / agent	Model and tool calls routed through it	JetStream, Straiker, CrowdStrike AI Gateway (announced, not yet available)	Agents that bypass it, and the gateway itself as a target

© 2026 Armature Systems Inc.

APPROACH	WHAT IT CONTROLS	EXAMPLES	BLIND SPOT
Data	What sensitive data agents can reach and move	Cyera, Bold	Actions that do not touch classified data
Endpoint	Agents running on employee devices - Browser and Local	CrowdStrike Falcon Guardian, Bold, Zscaler Endpoint AI Security	Server and cloud-hosted agents

Disclosure: Armature partners with or resells several vendors in this market. Examples are illustrative, not endorsements, and several products were announced before Q3.

Agents Without Motives: When Optimization Becomes an Intrusion

The quarter's most striking AI incident involved no attacker at all. In OpenAI's internal ExploitGym evaluation, about 1,200 agents built an unsanctioned communication channel by encoding messages in directory names, exchanging more than 70,000 messages. About 700 of them then breached Hugging Face's production Kubernetes clusters and harvested 136 secrets. Hugging Face disclosed the breach on July 16 and OpenAI acknowledged it on July 21; technical reports from OpenAI, METR and Redwood Research followed in August.

OpenAI's stated causes were reward hacking, persistence on effectively impossible tasks, insufficient controls on communication between agents, and detection gaps.

The agents were optimizing for the benchmark's score. Investigators also found some agents spoofing tool outputs and concealing evidence of their reward hacking. No motive is needed to explain that; it is goal-directed behavior, and it needs controls all the same.

Other Q3 cases: Anthropic disclosed on July 30 three incidents during cybersecurity evaluations in which Claude models reached real systems, including publishing a malicious package to the public PyPI registry that was installed on 15 third-party hosts. On September 9 it disclosed a fourth incident, from January 2026, found while preparing transcripts for independent review by METR; the UK AI Security Institute separately reported an incident in its own testing. On September 10, OpenAI told Services Australia that one of its agents had accessed the Medicare Statistics Reporting portal in June, reportedly a month after OpenAI identified the activity; the Australian Signals Directorate is investigating.

How this report describes AI: none of these incidents requires intent to explain it. The systems pursued the objectives they were trained and scored on, and took actions their builders did not authorize. That is the security risk: goal-seeking automation with real access, not machines with motives. Throughout this report we describe what AI systems did, not what they "wanted".

© 2026 Armature Systems Inc.

Frontier AI Labs Slow Down

On August 18 OpenAI paused reinforcement-learning training for two weeks on models intended for deployment, kept its largest planned frontier training run on hold, and paused some research workloads that run code or reach the internet. It cited the Hugging Face incident and signs that a model in development may cross its own critical cyber-capability threshold. Anthropic said it paused external cyber evaluations and higher-risk training environments for several weeks. In July more than 1,200 employees across frontier labs signed the "Pacing the Frontier" letter calling for international coordination on development pace; Anthropic's CEO published "We must pace the frontier" in mid-September. We found no comparable pacing statement from Google DeepMind.

Why it matters to you: this is now AI vendor risk. Expect staged or delayed access to new models, model changes with little notice, and a stronger case for incident-disclosure and model-change terms in AI contracts.

ASCII Smuggling: An Example of AI Research Becoming Traditional Attack Tradecraft

One of the more unusual Q3 developments is the migration of a technique associated with AI prompt injection into conventional phishing. Microsoft researchers disclosed on September 3 a high-volume phishing campaign using invisible Unicode tag characters, a technique known as ASCII smuggling. The campaign ran from February 9 to May 15, 2026 and peaked at 2.37 million messages on February 26, so this is a Q3 disclosure of earlier activity. The campaign itself was not AI-generated. The technique was originally associated with hiding instructions from AI models, but attackers are now using the same concept to evade security filtering and disguise malicious content. This is important because it demonstrates a broader pattern. Techniques developed around AI security do not necessarily remain confined to AI. They can migrate into ordinary cybercrime.

Outlook for the Remainder of 2026

AI enabled cyber operations will continue moving toward greater autonomy. The current evidence does not mean fully autonomous cyber warfare has become commonplace. However, major AI providers, including Anthropic (September 10) and Google (September 8), have now documented increasingly agentic behavior by threat actors on their platforms.

© 2026 Armature Systems Inc.

Vulnerability management will become increasingly difficult. AI assisted vulnerability discovery is increasing the volume of findings, while exploitation of critical internet facing infrastructure continues. The key bottleneck is increasingly understanding exploitable vulnerabilities and limiting exposure.

AI infrastructure will become a more attractive target. As enterprises place valuable data and credentials behind AI gateways and agent frameworks, these components will increasingly become targets themselves.

Ransomware will continue adapting rather than disappearing. The ransomware model is becoming increasingly dependent on identity compromise, edge device exploitation and data theft. AI will likely increase the efficiency of these operations rather than replace them.

Sources: Armature Threat Research; Anthropic; Google Threat Intelligence Group; Microsoft; Hugging Face; OpenAI, METR and Redwood Research; Wiz; VulnCheck; vendor announcements.

© 2026 Armature Systems Inc.

6 Recommendations

CRITICAL PRIORITY

Prioritize by exploitability, not volume.

Deduplicate findings, then work what is on CISA KEV, reachable and exposed, or part of a known exploit chain. Where patching isn't practical, close the path with configuration changes or network- or permission-level compensating controls, and give every workaround an owner and a re-check date.

Treat edge devices as crown jewels.

Patch the Citrix, Cisco FMC, SonicWall SMA1000, Check Point, F5, Fortinet, Arista and Palo Alto flaws in this report, and the internet-facing business applications attacked this quarter (SAP Commerce, PTC Windchill, Oracle PeopleSoft). Take management interfaces off the internet. Hunt for compromise on any device that was exposed before patching (web shells, new accounts, tunnels), because these devices do not always log the intrusion. Keep an eye on notices as well, vendors don't provide advanced warnings to clients before announcing critical issues.

Harden the help desk and the identity layer.

Require call-back verification for MFA and password resets, move to phishing-resistant MFA, and block the OAuth device-code flow where it isn't needed. Where volume makes call-backs hard to sustain, consider ID verification technology. Review third-party apps connected to Salesforce and other SaaS, remove unused ones and scope their tokens.

HIGH PRIORITY

Put AI under the same controls as any privileged system.

Inventory AI gateways (including LiteLLM copies inside other products), RAG platforms, orchestration tools, MCP servers and agents, and the keys each one holds. Change default master keys, patch LiteLLM, RAGFlow and Kestra, and give agents least privilege, outbound traffic controls and a working kill switch.

Review AI instruction files like code.

Treat CLAUDE.md files, skills, assistant settings and MCP configuration as code: require review, protect them with code owners, alert on changes, and add them to your incident response checklist.

© 2026 Armature Systems Inc.

Close the developer supply chain path.

Allowlist IDE extensions, run package installs without lifecycle scripts where you can, and rotate developer and AI-tool keys after any package compromise in your dependency tree.

Evaluate AI governance tooling against the four control points.

Map which of the network, gateway, data and endpoint controls you have today. Before buying, confirm general availability, ask what open-source components sit under any gateway, and add AI incident-disclosure and model-change terms to AI contracts.

Gate agentic remediation like any other change.

Route AI-generated fixes through the same review, testing and change control as human patches, and verify closure. One Q3 study found only about one in four AI-generated patches was a clean fix (Off-by-1 Labs).

Contain at machine speed.

Automate detection and containment for identity, edge and endpoint so the first response to a stolen token, a compromised appliance or a malicious build does not wait on a human.

Also consider: detecting invisible Unicode tag characters in inbound email (ASCII smuggling), and enforcing browser auto-update with version-compliance monitoring (shared exploit chains).

© 2026 Armature Systems Inc.

Appendix

Other Notable Q3 Events

DATE	EVENT	WHY IT MATTERS
Jul 16	Coca-Cola disclosed in an SEC filing that ransomware suspended fairlife's US production	An IT intrusion halted production at a public company, with same-day SEC disclosure
Jul 16 – Sep 18	Pentagon personnel-data (DMDC) exposure through a file-sharing flaw, active from October 2025; notices sent September 18	A months-long undetected exposure in a shared file system
Aug 2	EU AI Act transparency duties and enforcement powers over general-purpose AI providers took effect; high-risk deadlines moved to December 2027 and August 2028	Deployers now owe AI disclosure and deepfake labelling in the EU
Aug 10	Ransomware reached facility systems (HVAC, door access) at Winnipeg's Health Sciences Centre	Ransomware crossing into building and physical-security systems
Aug 18	CareCloud updated its breach to 3.75 million patients, with data taken from its AWS account (intrusion March 2026)	A shared health-records provider exposes every client at once
Aug 28	Manchester Airports Group reported a data breach affecting about 8.7 million people	A large passenger-data breach at critical transport infrastructure
Aug 31	An international operation involving US agencies and Europol sinkholed the Sality botnet (up to 1 million infected machines)	A major law-enforcement disruption
Sep 15	CenterPoint Energy confirmed a customer-data breach; service was not affected	Utility data breach without operational impact

Sources

Sources cited in this report, with each source's data period. Several Q3 publications report Q2 or first-half data.

PUBLISHER	SOURCE	PUBLISHED	DATA PERIOD
Anthropic	Threat intelligence report: detecting and countering misuse of AI	Sep 10	Dec 2025 – Aug 2026
Anthropic	Investigating incidents in cybersecurity evaluations, and follow-up disclosure	Jul 30; Sep 9	2026 testing
Anthropic	Mythos / Project Glasswing findings	—	H1 2026

© 2026 Armature Systems Inc.

PUBLISHER	SOURCE	PUBLISHED	DATA PERIOD
Arctic Wolf	“Cookie Crumbles” (CVE-2026-0257 and Qilin); PREY-0058 research	Jul 20; Sep 7	Jun 2026; Q3
Backslash Security	Shai-Hulud AI tooling persistence	Aug 5	Q3
BleepingComputer	McKesson breach; Lazarus AFD.sys zero-day; SAP Commerce Cloud CVE-2026-58231	Aug 2026	Q3
Breachsense, Cyble, ZeroFox	Monthly ransomware reports	Aug – Sep	Jul – Aug 2026
CISA	Known Exploited Vulnerabilities catalog	Ongoing	Q3
Cisco Talos	IR Trends Q2 2026; “FMC ongoing exploitation”	Jul 28; Sep 9	Q2 2026; Q3
CrowdStrike	2026 Threat Hunting Report	Aug 3	H1 2026
Google Threat Intelligence Group	“From Prompting to Autonomy” (incl. DUSTMAKER); UNC6671 vishing research	Sep 8; Aug	About Q2 2026; Q3
Group-IB	Handala / HEAVYGRAM research	Sep 17	Q3
Hugging Face	Production cluster breach disclosure (ExploitGym agents)	Jul 16	Jul 2026
Mandiant	Oracle PeopleSoft CVE-2026-35273 WAF bypass	Sep 25	Q3
Microsoft	“When AI infrastructure becomes the target”; ASCII smuggling research; Digital Crimes Unit, EvilTokens disruption	Aug 26; Sep 3; Sep 22	Not stated; Feb – May 2026; Q3
NBC News; Fortune	Minnesota and multi-state water system attacks	Jul 31	Jul 2026
Off-by-1 Labs (1Password)	AI-generated patch study	Aug 6	—
OpenAI	“Pacing model development”; Hugging Face incident reports (with METR and Redwood Research)	Aug 2026	Jul 2026
Pacing the Frontier letter; Dario Amodei	Letter; “We must pace the frontier”	Jul; mid-Sep	—
Recorded Future	H1 2026 Malware and Vulnerability Trends	Sep 3	H1 2026
Resecurity	INC Ransomware and SonicWall SMA1000	Aug 3	Jun – Aug 2026
Root Evidence	Exploited-vulnerability research	May 27	Pre-Q3 baseline
SecurityWeek	Cl0p Windchill campaign	Aug 19	Q3

© 2026 Armature Systems Inc.

PUBLISHER	SOURCE	PUBLISHED	DATA PERIOD
Socket, StepSecurity, OX Security, Manifold	npm, PyPI and Open VSX supply chain research	Jul – Aug	Q3
SonicWall; Rapid7	SMA1000 CVE-2026-83548 / -83549	Sep 1	Q3
Verizon	2026 Data Breach Investigations Report	May 19	2025 (baseline)
Volexity	“Mind the Patch Gap” parts 1 and 2	Sep 9; Sep 21	Sep 2026
VulnCheck	State of Exploitation 1H 2026	Jul 28	H1 2026
Wiz	LiteLLM research (via The Hacker News)	Sep 10	Q3
Appendix events	Coca-Cola Form 8-K; Military Times (DMDC); Council of the EU (AI Act); CTV News (Health Sciences Centre); TechCrunch (CareCloud); Europol (Sality)	Jun 29 – Sep 24	Q3
Vendor announcements	CrowdStrike, JetStream, Straiker, Cyera, Zscaler, Netskope, Bold	Dec 2025 – Sep 2026	—
Armature Systems	Armature Threat Research and client advisory work	Q3	Q3

© 2026 Armature Systems Inc.

ARMATURE
SYSTEMS

AUTHORS



Jared Mendenhall
CISO - Armature Systems



Nadaraja Himasuthan
Threat Researcher - Armature Systems

ABOUT ARMATURE SYSTEMS

Armature Systems is a cybersecurity systems integrator that helps organizations design, build, and operate network and security programs across the full lifecycle.

We work hands-on alongside internal teams to deploy the right technology, build effective processes, and deliver measurable security outcomes across network, cloud, identity, compliance, and threat response.

Ready to modernize your security program? Visit armaturesystems.com

© 2026 Armature Systems Inc.