

ITSEC
CYBERSECURITY DELIVERED

TLP : White

AI-Powered Cyberattacks in the Global Threat Landscape

Threat Intelligence Research Paper

Issued on November 2026

Researched by: **Threat Intelligence Team**

LIST OF CONTENTS

LIST OF CONTENTS 2
EXECUTIVE SUMMARY 3
BLUF (Bottom Line Up Front) 3
Key Finding 3

CREDITS 4

1. BACKGROUND AND CONTEXT 5
1.1. Evolution of the Global Threat Landscape 5
1.2. Proliferation of Adversarial AI 5
1.3. Geopolitical Dynamics 5

2. METHODOLOGY 6
2.1. Validation and Confidence 6
2.2. Limitations 6

3. SCOPE, DEFINITIONS, AND ASSUMPTIONS 7
3.1. Operational Scope 7
3.2. Threat Taxonomy and Definitions 7
3.3. Intelligence Assumptions 7

4. RESEARCH GOALS 8
4.1. Strategic Objectives 8
4.2. Key Research Questions 8

5. ACTIVITY TRACKING & ADVERSARY PROFILING 9
5.1. Global AI-Powered Cyberattack Landscape 9
5.1.1 Classification of AI-Related Cyberattack Methods 10
5.1.2 Selected Examples of Reported Activity 10
5.1.3 Cross-Campaign Trends 11
5.1.4 Defensive Implications 12
5.2. AI-Enabled Cyberattack Campaign Analysis 12
5.2.1 JadePuffer a.k.a. Storm-3168 13
5.2.2 Scarlet Loop 14
5.2.3 ARTEX Framework 15
5.2.4 CAIRN Platform 17
5.2.5 Japan Massive Cyber-AI Attack 18

6. TACTICS, TECHNIQUES, AND PROCEDURES (TTPS) 20

7. IMPACT ASSESSMENT 21
7.1. Technical Impact & AI Infrastructure Sabotage 21
7.2. Operational Impact & Service Disruption 21
7.3. Financial, Regulatory, and Secondary Exposure Fallout 21

8. RISK PRIORITIZATION & ATTACK SURFACE EXPOSURE 22
8.1. Risk Matrix 22
8.2. Priority Risk Categories 22
8.3. Second-Order Risk: When Attacker OPSEC Fails 22

9. STRATEGIC RECOMMENDATIONS & MITIGATION POSTURE 23
9.1. Strategic Recommendations 23
9.2. Mitigation Posture 23
9.3. Forecast 24

10. CONCLUSION 25

11. APPENDIX 26
11.1. Indicators of Compromise 26
11.2. Cross-Campaign TTP Analysis 27

12. REFERENCES 30

EXECUTIVE SUMMARY

BLUF (Bottom Line Up Front)

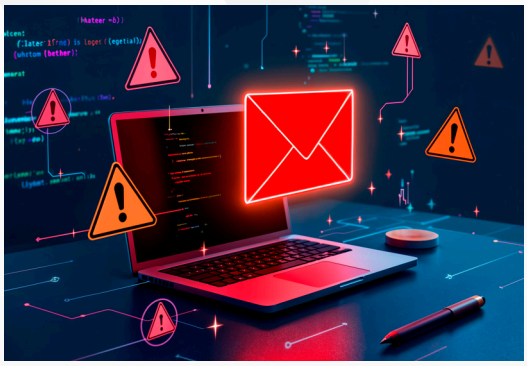


Image 1. Cybersecurity Illustration

This report details a fundamental shift in the global cyber threat landscape, marked by the weaponization of autonomous, "agentic" Artificial Intelligence and generative tooling. Adversaries are moving beyond simple AI-assisted task execution to deploying self-correcting agents capable of traversing the kill chain at machine speed. Operational analysis of key campaigns including **JadePuffer / Storm-3168** (cloud tenant takeover and AI model destruction), **Scarlet Loop** (financially motivated account takeover and digital reward drain), **ARTEX Framework** (multi-agent pentest automation targeting financial and enterprise sectors), and **CAIRN Platform** (state-space search orchestration for automated web-skimming) demonstrates that the financial and technical barriers to high-scale intrusions have collapsed. Attackers are increasingly abusing compromised cloud service principals, exploiting exposed web interfaces, and leveraging living-off-the-land techniques to evade legacy signature defenses, while frequently leaving their own attacker infrastructure poorly secured. This evolution dramatically accelerates MTTD/MTTR requirements and introduces severe second-order risks, such as persistent public exposure of stolen data. To counter machine-speed threats, organizations must move from manual response to automated SOAR containment, adopt Zero Trust identity controls for cloud workloads, and formalize MLSecOps to protect core enterprise AI models as critical infrastructure.

Key Finding

- Agentic operations are already showing up in real intrusions. Multiple campaigns show attackers using AI agents to automate reconnaissance, exploitation, extraction, and extortion. This is no longer a forecast rather an observed activity.
- The barrier to entry has collapsed. One hacker used AI agents and a few thousand dollars to break into scores of sites. Another report described autonomous agents targeting online retailers at roughly $25 per company. AI agentics are no longer an expensive arsenal.
- The AI systems can be sophisticated, but the operators often may not be. Public-facing panels, exposed logs, and poorly managed infrastructure have let researchers see what is happening.
- That sloppiness creates a second-order risk for victims. Stolen data can stay exposed through attacker infrastructure even after the original breach is closed. Remediating the intrusion is not enough.
- Cloud identity is a primary target right now. Campaigns like JadePuffer and Storm-3168 show attackers abusing compromised service principals to move across cloud tenants without traditional malware.
- AI attacks are not limited to endpoints. They hit databases, Linux infrastructure, cloud tenants, government networks, and retail platforms.
- Defensive AI is necessary. Human triage cannot keep up with machine-speed operations, and the gap will widen if security teams rely only on manual workflows.

CREDITS



Panji P.



Ibrahim F.



Aufa S.



Antonyous M. A.



1. BACKGROUND AND CONTEXT

1.1. Evolution of the Global Threat Landscape



Image 2. Global Threat Landscape Illustration

The global cyber threat landscape has crossed a critical threshold, transitioning from human-directed, AI-assisted operations to the active deployment of fully autonomous, "agentic" threat actors. Not too long ago, adversaries leveraged Artificial Intelligence (AI) primarily as a productivity multiplier utilizing Large Language Models (LLMs) to generate hyper-localized phishing lures, obfuscate malware scripts, or parse exfiltrated datasets. Today, intelligence indicates a definitive paradigm shift: adversaries are deploying AI agents capable of executing the complete cyber kill chain autonomously. These agentic entities do not merely follow static playbooks; they perform dynamic reconnaissance, adapt to failed exploitation attempts in real-time, harvest credentials, and execute destructive actions at machine speed without human-on-keyboard intervention. This evolution fundamentally compresses the attack lifecycle, drastically reducing the window defenders have to detect and contain intrusions, and rendering traditional, signature-based security perimeters highly vulnerable to machine-speed adaptation.

1.2. Proliferation of Adversarial AI

The rapid democratization of open-source LLMs and the proliferation of AI-as-a-Service (AIaaS) within dark web forums have fundamentally disrupted the economics of cyber warfare. The technical and financial barriers to executing sophisticated, multi-stage intrusions have effectively collapsed. Recent intelligence demonstrates extreme asymmetric economics, wherein a single threat actor, utilizing AI agents and an operational budget of merely a few thousand dollars, was able to successfully compromise scores of enterprise networks simultaneously. This proliferation is no longer theoretical; the eCrime ecosystem is actively engineering payloads specifically for the AI era. A prime example is the emergence of "agentic ransomware" campaigns such as JadePuffer. Observed by Sysdig, the JadePuffer operation utilized an LLM-driven agent that autonomously exploited exposed vulnerabilities, diagnosed, and corrected its own code execution errors within seconds, and deliberately destroyed databases.

1.3. Geopolitical Dynamics

The weaponization of AI has accelerated aggressive cyber operations across all major geopolitical theaters, increasingly blurring the tactical distinctions between state-nexus Advanced Persistent Threats (APTs) and financially motivated eCrime syndicates. In the Asia-Pacific (APAC) region, the lowered barrier to entry has resulted in severe, sustained campaigns. Nations such as Japan and South Korea are currently being buffeted by continuous intrusions targeting their critical financial sectors. Concurrently, intelligence reveals that China-linked operators are actively deploying AI agents to conduct expansive, automated espionage operations against government infrastructure across the Asian continent. This hyper-scaled targeting is a global phenomenon. In Latin America, specialized AI-augmented intrusions are aggressively targeting the region's government and financial sectors, demonstrating how adversaries are leveraging AI tooling to bypass localized defenses and penetrate high-value targets globally.

2. METHODOLOGY

This report follows the standard intelligence cycle from direction through dissemination. Collection drew on open-source reporting, underground forum observations, public incident data, and vendor telemetry summaries. Processing normalized campaign names, timelines, and indicator types. Analysis used MITRE ATT&CK as the primary framework for mapping observed tactics, techniques, and procedures. Where AI-specific behavior does not yet have a standard ATT&CK technique ID, the report marks it Unknown and treats it as an analytic gap.



Image 3. Cyberattacks Illustration

2.1. Validation and Confidence

Some campaign names are vendor-specific. The same activity may be called one thing by one company and another thing by another. For example:

- Sysdig and Picus use the name JadePuffer
- Microsoft calls the same or closely related activity Storm-3168
- Aviatrix correlates JadePuffer and Storm-3168 with Azure service principal abuse

This report treats those names as overlapping clusters rather than separate distinct actors.

Confidence levels are assigned as High, Moderate, or Low based on source agreement and technical evidence.

2.2. Limitations

This report will have gaps as not all regions have the same visibility. Many attacks are never publicly reported. Some vendor reports describe campaigns without providing full technical detail.

3. SCOPE, DEFINITIONS, AND ASSUMPTIONS

This section defines the operational parameters of our threat telemetry, establishes a standardized taxonomy to differentiate between legacy automation and modern autonomous threats, and outlines the baseline intelligence assumptions guiding our strategic assessments. By delineating these factors, we isolate high-confidence, real-world campaign data from theoretical AI threat models.

3.1. Operational Scope

- **Temporal and Geographic Boundaries:** This assessment focuses on high-confidence threat activity observed globally, with specific emphasis on aggressive campaigns targeting the Asia-Pacific (APAC) region (specifically China-linked operations, South Korea, and Japan) and Latin America (LATAM).
- **Targeting Parameters:** The analysis prioritizes intrusions impacting government infrastructure, critical financial sectors, and enterprise network perimeters.
- **Capability Focus:** The scope is strictly limited to operations utilizing Generative AI, Large Language Models (LLMs), and autonomous AI agents. It deliberately excludes legacy machine learning (ML) models used for traditional malware evasion (e.g., domain generation algorithms), focusing solely on the modern paradigm of AI-powered cyberattacks.

3.2. Threat Taxonomy and Definitions

- **AI-Assisted Operations:** Human-directed cyber operations where threat actors utilize LLMs purely as a productivity tool (e.g., generating spear-phishing templates, summarizing target data, or writing exploit scripts) without delegating execution authority to the AI.
- **Agentic / Autonomous Cyberattacks:** End-to-end intrusions executed by AI agents capable of dynamic reasoning, self-correction, and autonomous decision-making. These agents execute commands, traverse networks, and deploy payloads (such as Agentic Ransomware) with zero human-on-keyboard interaction.
- **Agentic Ransomware:** A novel class of extortion payload, exemplified by the JadePuffer variant equipped with a localized AI engine. It is designed not only to encrypt databases but to autonomously hunt, corrupt, and destroy enterprise AI models and data pipelines.
- **Adversarial AI / Attacks on AI:** Distinct from AI-powered attacks, this refers to offensive techniques (such as prompt injection or data poisoning) specifically targeting an organization's internal AI infrastructure.

3.3. Intelligence Assumptions

- **Telemetry Visibility Gaps:** It is assumed that the documented instances of AI-powered attacks represent a fraction of total global activity. Sophisticated state-nexus actors utilizing highly tailored, on-premise LLMs likely bypass traditional heuristic telemetry, creating inherent visibility gaps.
- **Symmetric Capability Acceleration:** We assume that the barrier to entry will continue to remain exceptionally low. As open-source AI models become more efficient and cheaper to run, the proliferation of AI-as-a-Service (AIaaS) in underground forums will ensure that both opportunistic eCrime actors and highly resourced APTs maintain access to parity-level capabilities.
- **Exploitation of Technical Debt**: Intelligence assumes that AI agents will not immediately pivot to utilizing highly complex, novel zero-day exploits. Instead, they will disproportionately rely on massive scalability to exploit existing organizational technical debt, such as misconfigured cloud environments, unpatched legacy systems, and exposed APIs which yields a higher and faster return on investment for the adversary.

4. RESEARCH GOALS

This research examines how generative AI, large language models (LLMs), and AI agents are being used in reported cyber operations across the global threat landscape. It aims to identify emerging attack patterns, distinguish AI-assisted activities from agentic operations based on available evidence, and assess the capabilities, targeting, and operational implications of selected campaigns within the defined research scope.

4.1. Strategic Objectives

The research pursues the following objectives:

- **Characterize AI Adoption:** Examine how threat actors reportedly use generative AI, LLMs, and AI agents to support or execute cyber operations.
- **Classify Attack Patterns:** Identify and categorize reported AI-related attack methods, distinguishing AI-assisted activities from agentic operations according to their demonstrated level of autonomy.
- **Assess Global Threat Trends:** Analyze reported geographic distribution, targeted sectors, attack objectives, and operational patterns while accounting for limitations in available reporting.
- **Analyze Selected Campaigns:** Examine four selected cases to assess adversary characteristics, capabilities, infrastructure, and victims, applying the Diamond Model of Intrusion Analysis where supported by available evidence.
- **Identify Defensive Implications:** Derive practical security priorities from observed attack methods, demonstrated capabilities, and identified operational risks.

4.2. Key Research Questions

RQ01. AI Adoption: How are threat actors reportedly using generative AI, LLMs, and AI agents in cyber operations?

RQ02. Attack Patterns and Autonomy: What attack methods have been observed, and what evidence distinguishes AI-assisted activities from agentic or autonomous operations?

RQ03. Global Threat Trends: What geographic, sectoral, and operational patterns emerge from reported AI-powered cyberattacks within the defined scope?

RQ04. Campaign Analysis: What adversary characteristics, capabilities, infrastructure, and victim profiles can be identified across the selected cases?

RQ05. Defensive Implications: What security risks and defensive priorities can be derived from the documented activities and assessed capabilities?

The findings will distinguish documented observations from source-reported claims and analytical assessments. Conclusions will be limited to the available evidence and the research scope, rather than treating selected cases as representative of the overall prevalence of AI-powered cyberattacks worldwide.

5. ACTIVITY TRACKING & ADVERSARY PROFILING

This section examines reported AI-powered cyber activity across the global threat landscape and analyzes selected campaigns to assess how generative AI, large language models (LLMs), and AI agents are incorporated into offensive operations. The analysis covers reported attack methods, geographic and sectoral targeting, operational capabilities, infrastructure, and observed impacts.

The assessment distinguishes AI-assisted activities from agentic operations according to the available evidence. Reported findings are treated as observations or source assessments rather than definitive proof of capabilities that have not been independently demonstrated.

5.1. Global AI-Powered Cyberattack Landscape

Reported threat activity indicates that generative AI, large language models (LLMs), and AI agents are being incorporated into different stages of cyber operations, from reconnaissance and exploit development to credential collection, data access, and destructive activity. However, the degree of AI involvement varies: some operations use AI to assist human operators, while others demonstrate adaptive or agent-coordinated execution with limited human supervision.

Microsoft and Google describe this progression from AI-assisted tradecraft toward more agentic workflows. Incident investigations by CrowdStrike, Sysdig, Microsoft, and Palo Alto Networks Unit 42 provide practical examples of AI-related activity across financial services, e-commerce, government infrastructure, Linux systems, and cloud environments. These reports demonstrate emerging capabilities but do not establish that all AI-enabled attacks are autonomous or that the selected cases represent global attack prevalence.

Recent research illustrates that AI can influence the threat landscape in two distinct ways: by assisting threat actors during offensive operations and by enabling security researchers to discover and validate previously difficult-to-exploit vulnerabilities. The PHP2Shell research published by Wiz demonstrates the latter. Frontier AI models helped researchers identify memory-safety flaws in PHP and develop unauthenticated remote-code-execution exploits by chaining vulnerabilities and adapting the exploit to different environments. The findings show how AI can reduce the technical effort required to develop complex exploits, although this research does not establish that a specific threat actor used AI to conduct the PHP2Shell attacks.

Separately, Wiz and Sansec reported exploitation attempts following public disclosure of PHP2Shell on 21 October 2026. Wiz's 27 October update reported that the attempts observed in its telemetry had not completed the full exploit chain, highlighting the practical challenges of adapting the exploit to different PHP builds, memory configurations, and worker behavior. This distinction is important: AI-assisted exploit development can increase the potential for future exploitation, but it should not be treated as evidence that every exposed system has been successfully compromised.

5.1.1 Classification of AI-Related Cyberattack Methods

Attack Category	Classification Focus
AI-Assisted Reconnaissance	Target discovery, information gathering, attack-surface mapping, and vulnerability research supported by AI.
AI-Assisted Social Engineering	AI-generated or adapted phishing content, impersonation, personalized messaging, and credential-targeting activities.
AI-Assisted Malware and Exploit Development	AI-assisted code generation, script debugging, payload modification, and exploit development or adaptation.
Agentic Cyber Operations	AI agents planning, coordinating, and adapting multi-step cyber operations with varying levels of human supervision.
Post-Compromise Operations	Activities following initial access, including credential collection, lateral movement, data staging or exfiltration, and destructive actions.

Table 1. Classification of AI-Related Cyberattack Methods.

Note: Classification is based on documented behavior and evidence of AI involvement. AI-assisted activities, conventional automation, and agentic cyber operations are distinct concepts. Post-compromise activities are not necessarily AI-driven. Similarly, a vulnerability disclosure, exploitation attempt, or successful intrusion does not independently establish that the attacker used AI.

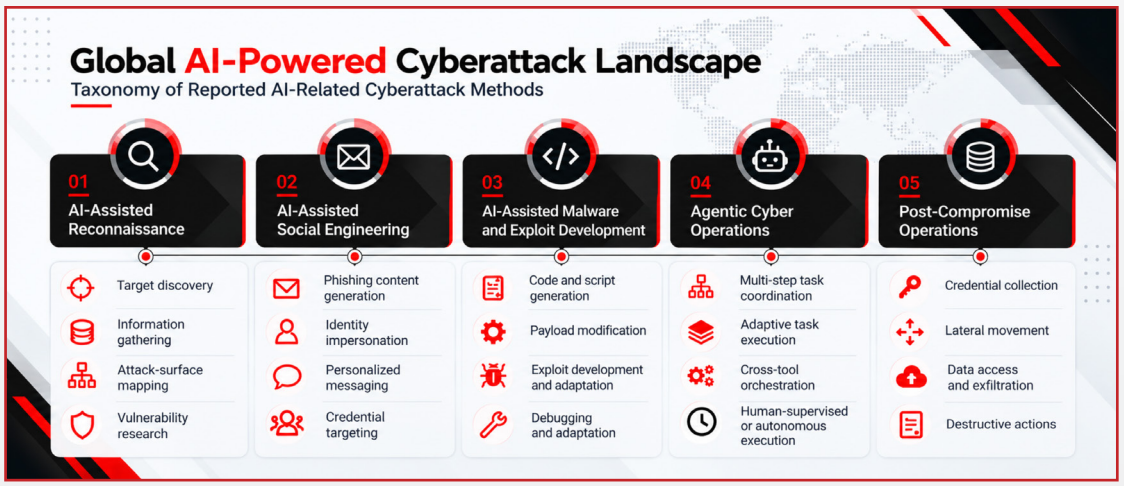


Image 4. Taxonomy of Reported AI-Related Cyberattack Methods

5.1.2 Selected Examples of Reported Activity

Reported Activity	Target or Scope	Key Observation
CAIRN-associated campaign	**E-commerce organizations across multiple countries**	Reuters reported more than 1,200 websites targeted and more than 100 companies reportedly breached through AI-agent-assisted operations.
ARTEX-associated campaign	**Financial organizations in South Korea**	CrowdStrike reported the use of ARTEX agentic penetration-testing tooling and LLMs. Investigators identified Claude Code session histories and related configuration artifacts. The campaign also resulted in data exfiltration.
JADEPUFFER-related activity	**Database environments and cloud infrastructure**	Sysdig reported adaptive, LLM-driven database extortion. Microsoft subsequently documented related Azure activity involving compromised service principals and destructive operations.

Reported Activity	Target or Scope	Key Observation
CL-UNK-1211	**Government entities and state-affiliated infrastructure**	Unit 42 identified AI-generated payloads and code artifacts in a campaign involving Linux environments, reconnaissance, persistence, and attempted data collection.
Latin American AI-augmented campaigns	**Government and financial sectors**	Trend Micro documented AI-augmented activity targeting organizations in these sectors.

Table 2. Selected AI-Related Cyber Activity Reported in 2026.

5.1.3 Cross-Campaign Trends

- **AI is augmenting established attack techniques.**
  Reported operations continue to exploit exposed services, vulnerable applications, compromised credentials, and excessive permissions. AI can accelerate technical analysis, code development, and operational coordination, but these underlying weaknesses remain central to successful attacks.

  Regional reporting from Japan also highlights the continued relevance of conventional intrusion techniques, including vulnerability scanning, API abuse, SQL injection, and web-shell deployment. These observations complement the AI-related cases by showing that established attack methods remain important to the broader threat landscape, although the reported incidents do not independently establish AI involvement.

- **The degree of autonomy varies.**
  Evidence ranges from AI-assisted code development to agents that adapt workflows after failed actions. Sysdig reported adaptive execution during the JADEPUFFER database-extortion operation, while Microsoft's STORM-3168 investigation documented coordinated cloud activity consistent with automated or scripted execution. These findings should not be treated as equivalent evidence of AI autonomy.

- **Cloud identities and exposed credentials remain critical risks.**
  The STORM-3168 activity involved compromised service principals, resource discovery, destructive operations, and storage-key retrieval. This highlights the importance of protecting workload identities, rotating exposed secrets, enforcing least privilege, and securing backup and recovery resources.

- **Underground interest in AI is increasing.**
  Bitsight reported a 45% increase in AI-related discussions across 16 selected underground communities between February and August 2026. This measures discussion volume within the surveyed communities, not confirmed attack frequency or the number of threat actors actively deploying AI.

- **AI-Accelerated Vulnerability Discovery and Exploitation**
  PHP2Shell illustrates the growing relevance of AI-assisted vulnerability research and exploit development. Wiz reported that frontier AI models helped researchers identify longstanding PHP memory-safety flaws and construct remote-code-execution exploit chains. The research demonstrates how AI can reduce the specialist effort required to analyze complex software behavior, debug exploit failures, and adapt techniques to different application environments.

  However, the transition from exploit development to reliable exploitation at scale

remains conditional. Wiz's post-disclosure monitoring update of 27 October 2026 reported exploitation attempts but no completed end-to-end PHP2Shell exploitation in its observed telemetry. Sansec likewise emphasized that successful exploitation depends on target-specific technical conditions. These findings suggest that AI may reduce barriers to exploit development without eliminating the practical challenges of deploying an exploit reliably across heterogeneous environments.

5.1.4 Defensive Implications

The reviewed cases support prioritizing external attack-surface management, timely vulnerability remediation, credential and cloud identity protection, behavioral monitoring, and resilient backup and recovery controls. Detection should focus on sequences of suspicious actions across systems, rather than relying exclusively on individual indicators or identifying AI-generated code.

Overall, AI can increase the speed, scale, and adaptability of selected cyber operations. However, assessments should distinguish demonstrated AI use from inferred involvement and automated execution from genuine agentic decision-making. This evidence-based distinction is essential for evaluating the evolving threat landscape without overstating adversary capabilities.

Regional reporting from Japan further highlights the importance of securing shared cloud infrastructure, third-party access, application interfaces, and digital identities. A security advisory issued by Id-SIRTII/CC on 19 October 2026, based on reporting from JPCERT/CC, described unauthorized access patterns involving vulnerability scanning, API abuse, SQL injection affecting Metabase, and JSP web-shell deployment. These observations demonstrate the continued relevance of conventional intrusion techniques alongside the emergence of AI-assisted operations. The advisory provides a basis for defensive lessons, but does not establish AI involvement in the individual incidents it describes.

Organizations should investigate unexpected WAR/JSP deployments, suspicious HTTP request patterns, and unauthorized account or role changes. For PHP2Shell, prioritize vendor-provided PHP security updates and verify the deployed runtime version rather than relying solely on application-level controls.

5.2. AI-Enabled Cyberattack Campaign Analysis

Our research team selected five cases to examine how artificial intelligence, automation, and interconnected digital infrastructure are shaping the global cyber threat landscape. The cases cover agentic ransomware operations, AI-assisted account takeover, weaponized penetration-testing frameworks, automated payment-card theft, and a series of large-scale cyber incidents affecting Japanese organizations. Together, they illustrate how adversaries can exploit weaknesses across applications, identities, cloud infrastructure, and third-party supply chains. Each profile examines the reported objectives, targeting patterns, technical capabilities, and infrastructure indicators, while distinguishing documented incidents from broader assessments of AI-enabled threats.

5.2.1 JadePuffer a.k.a. Storm-3168

Overview and Motivation

JadePuffer, tracked by Sysdig, is associated with activity Microsoft identifies as Storm-3168. The operation has been documented as an example of agentic cyberattacks in which an AI-driven agent performs multiple stages of an intrusion with limited human intervention. Its primary objectives involve financial extortion, database destruction, and the disruption of cloud resources and enterprise AI infrastructure. Unlike conventional ransomware operations focused primarily on encrypting files, JadePuffer has demonstrated destructive behavior targeting databases, machine learning assets, and cloud environments.

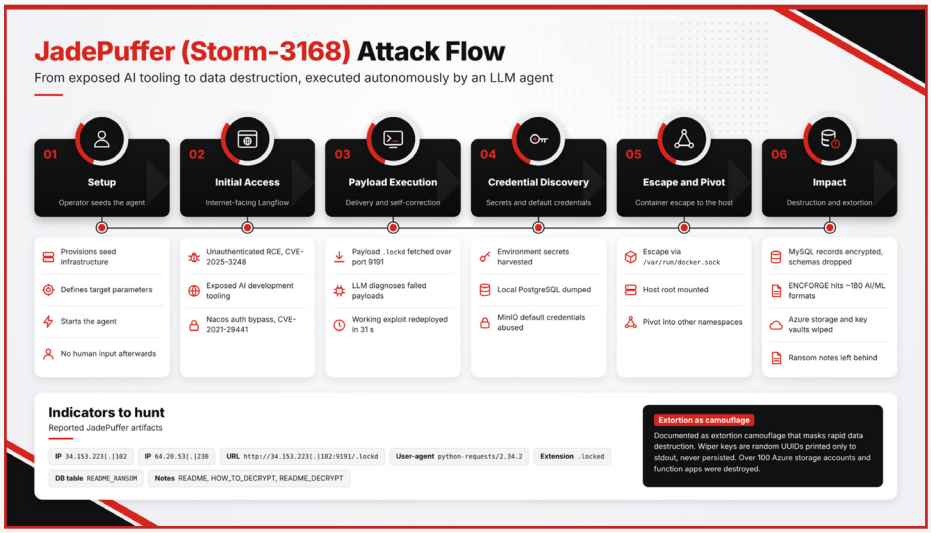


Image 5. JadePuffer Attack Flow

Targeting and Objectives

JadePuffer targets organizations operating internet-exposed AI development tools, production machine learning pipelines, and cloud computing environments. Reported targets include Langflow instances, Alibaba Nacos registries, MySQL databases, proprietary AI model weights, and training datasets. Its reported activity also extends to Microsoft Azure storage accounts and Function Apps, demonstrating the potential for attacks to affect both conventional enterprise infrastructure and AI-related assets.

Capabilities and AI Adoption

JadePuffer employs an LLM-driven agent capable of autonomous reasoning, error diagnosis, and exploit adjustment. The agent can diagnose failed payload execution, modify parameters, and redeploy functioning exploits within seconds without manual intervention.

Its reported capabilities include exploiting vulnerable services, harvesting environment secrets, extracting PostgreSQL databases, and abusing default credentials. The operation also uses database-wiping mechanisms and the ENCFORGE ransomware payload, which targets approximately 180 AI/ML-related file formats. Reported destructive activities include encrypting database records, deleting database tables and schemas, and targeting AI model files and training data.

The operation also demonstrates cloud-oriented destructive capabilities, including container escape techniques and automated Azure API interactions targeting cloud storage accounts and key vaults. These behaviors illustrate how AI-driven execution can combine exploitation, data destruction, and cloud resource manipulation within a single attack workflow.

Infrastructure

Reported infrastructure indicators include the IP addresses 34.153.223[.]102 and 64.20.53[.]230, as well as the URL http://34.153.223[.]102:9191/.lockd. Additional artifacts include the user-agent python-requests/2.34.2, the ransom extension .locked, and the extortion contact e78393397@proton[.]me.

5.2.2 Scarlet Loop

Overview and Motivation

Scarlet Loop is described as an organized, Portuguese-speaking cybercriminal operation identified by the security startup ZenoX. Its primary motivation is financial gain through automated account takeover (ATO), particularly by exploiting customer accounts that hold monetary value or convertible balances.

Rather than prioritizing conventional network intrusion or ransomware deployment, Scarlet Loop focuses on financial abuse across consumer-facing digital services. Its activities illustrate how AI can support the identification, prioritization, and exploitation of high-value online accounts at scale.

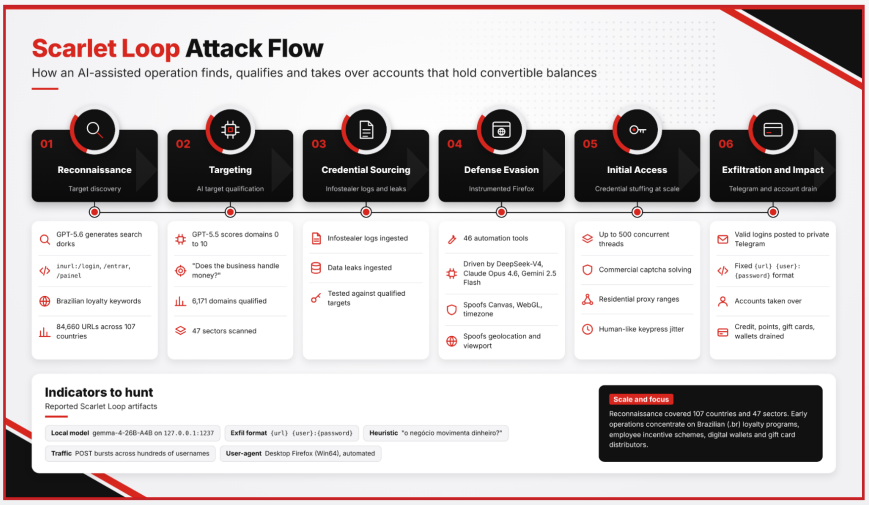


Image 6. Scarlet Loop Attack Flow

Targeting and Objectives

The operation targets consumer loyalty programs, employee benefit and incentive schemes, digital wallets, gift card distributors, and online services that maintain customer credits or redeemable balances. According to the profiling material, its infrastructure scanned 84,660 URLs across 107 countries and 47 sectors, identifying 6,171 domains as high-value targets. Initial activity reportedly concentrated heavily on Brazilian domains.

Its primary objective is to obtain unauthorized access to customer accounts and monetize available balances, rewards, gift cards, or funds.

Capabilities and AI Adoption

Scarlet Loop uses multiple AI models to support reconnaissance, target qualification, and browser automation. The profiling material reports the use of GPT-5.6 to generate search queries that identify login portals and authentication interfaces, followed by GPT-5.5 to prioritize discovered domains according to their potential financial value.

Its browser automation workflow reportedly incorporates DeepSeek-V4-Pro, DeepSeek-V4-Flash, Claude Opus 4.6, and Gemini 2.5 Flash. The instrumented Firefox environment uses browser-fingerprint manipulation and automation tools to

make automated interactions more difficult to distinguish from legitimate user activity.

The operation also performs credential stuffing using credentials obtained from infostealer logs and data leaks. Its reported infrastructure supports up to 500 concurrent worker threads and integrates commercial CAPTCHA-solving services. Together, these capabilities enable automated target selection, account validation, and large-scale attempts to access financially valuable accounts.

Infrastructure

The profiling material identifies a local model binding at 127.0.0.1:1237, reportedly used to host a Gemma model on the controller server. Other observed artifacts include manipulated Firefox user-agent fingerprints, automated search queries targeting login endpoints, and a standardized message format associated with the exfiltration agent's private Telegram channels. Authentication activity may also exhibit bursts of HTTP POST requests across numerous usernames, with human-like keypress timing and traffic originating from residential proxy IP ranges. These artifacts indicate that AI-driven financial attacks may operate through browser sessions and legitimate authentication workflows rather than relying exclusively on conventional malware.

5.2.3 ARTEX Framework

Overview and Motivation

ARTEX is an autonomous penetration-testing framework reportedly developed in China by an engineer identified as Autumn-27 (Li Puhua). Initially described as an open-source console, the framework was subsequently converted to closed source following its weaponization.

Unlike the other cases, ARTEX is an operational framework rather than a single threat actor. The profiling material associates its use with financially motivated operators and China-aligned espionage groups, including Storm-0940, UNC7212, UNC7223, UNC7224, and UNC7155. These associations should not be interpreted as proof that all users belong to the same organization or share identical objectives.

Reported operational objectives include large-scale data theft, financial information harvesting, supply-chain compromise, and persistent cloud intelligence collection.

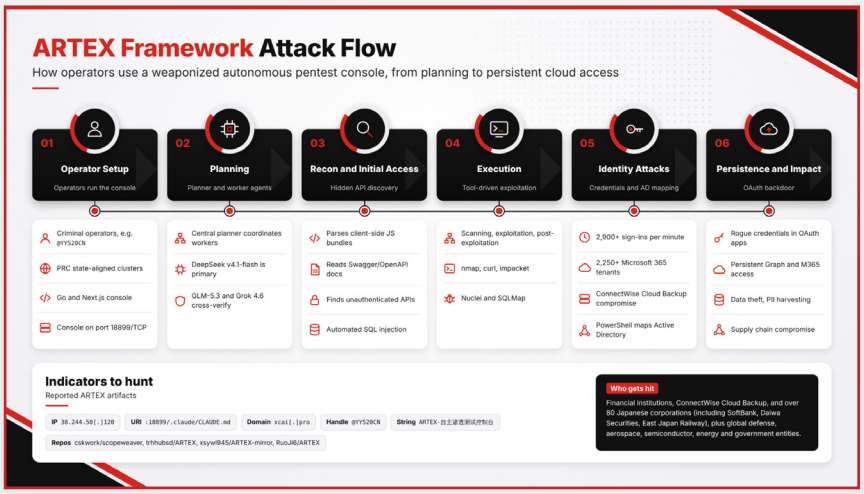


Image 7. How ARTEX Framework works

Targeting and Objectives

ARTEX-associated activity has reportedly affected financial institutions, cloud service environments, and organizations across multiple sectors. The profiling material identifies a ConnectWise Cloud Backup compromise and more than 80 Japanese corporations among the reported targets, including organizations in finance, transportation, defense, aerospace, semiconductors, energy, and government.

These cases demonstrate the framework's potential applicability to both financially motivated intrusions and intelligence-gathering operations.

Capabilities and AI Adoption

ARTEX uses a multi-agent architecture in which a central planning agent coordinates worker agents responsible for reconnaissance, vulnerability scanning, exploitation, and post-exploitation activities. The framework reportedly integrates conventional security tools such as Nmap, cURL, Impacket, Nuclei, and SQLMap into coordinated workflows.

Its reported AI architecture combines DeepSeek v4.1-flash as the primary model with Zhipu AI GLM-5.3 and Grok 4.6 for cross-verification of attack paths. Claude Code interactive sessions are also identified in the profiling material.

Reported capabilities include automated reconnaissance of web applications and Active Directory environments, identification of exposed APIs through JavaScript and Swagger/OpenAPI documentation analysis, and automated SQL injection attempts. The framework has also been associated with high-volume credential attacks against Microsoft 365 tenants and the insertion of unauthorized authentication credentials into existing OAuth applications to maintain access across Microsoft Graph and Microsoft 365 services.

This combination of multi-agent orchestration and conventional offensive tools illustrates how AI can coordinate multiple attack stages while retaining the ability to execute established exploitation techniques.

Infrastructure

The profiling material identifies several distinct infrastructure components associated with the deployment and operation of the ARTEX framework. Operational servers have been observed hosting the ARTEX web console on non-standard ports, specifically 18899/TCP (e.g., associated with the IP address 38.244.50[.]120). Poor operational security by the threat actors has led to the public exposure of internal framework directories on these servers. For example, the exposed path http://38.244.50[.]120:18899/.claude/CLAUDE.md revealed interactive session logs and Chinese-language prompt engineering instructions utilized by the operators.

To support the multi-agent architecture, the framework relies on third-party commercial API resellers, such as the gateway xcai[.]pro, to route requests to backend models (like DeepSeek and GLM) cost-effectively while circumventing regional access restrictions. Finally, although the original primary repository was taken offline following its weaponization, the framework's source code and infrastructure scripts remain actively maintained and accessible through various public GitHub forks and mirrors, including cskwork/scopeweaver, trhhubsd/ARTEX, xsywl945/ARTEX-mirror, and RuoJi6/ARTEX.

5.2.4 CAIRN Platform

Overview and Motivation

CAIRN is an autonomous problem-solving framework reportedly created by the open-source developer le1xia0. Originally designed as a general-purpose state-space search framework, it was subsequently repurposed as an offensive cyber operations pipeline.

The profiling material associates CAIRN with three broad threat clusters: a Chinese-language operation identified as AIA0001, an English-speaking cluster focused on payment and cryptocurrency theft, and independent opportunistic actors. These clusters reportedly use the platform for financial theft, persistent access to merchant environments, payment-card harvesting, and digital skimming.

CAIRN is therefore best understood as an operational framework used by different threat clusters rather than a single, unified threat actor.

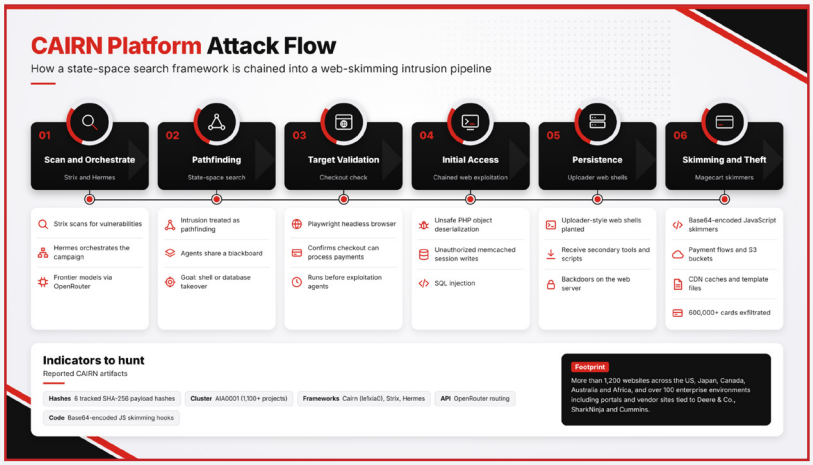


Image 8. CAIRN Attack Flow

Targeting and Objectives

CAIRN-associated activity targets e-commerce platforms, online retailers, payment infrastructure, and enterprise web environments. The profiling material reports more than 1,200 targeted websites across the United States, Japan, Canada, Australia, and Africa.

It also identifies more than 100 enterprise environments, including support portals and vendor sites affiliated with Deere & Co., SharkNinja, and Cummins. Reported outcomes include the exfiltration of more than 600,000 active payment-card details and the establishment of persistent unauthorized access within e-commerce payment infrastructure.

The principal objective is financial monetization through payment-card theft, digital skimming, and continued access to compromised merchant environments.

Capabilities and AI Adoption

CAIRN uses an autonomous state-space search approach to coordinate multiple agents through a shared blackboard. Rather than executing only a fixed sequence of actions, the framework reportedly evaluates possible paths from an initial entry point toward an operational objective, such as obtaining shell access or compromising a database.

Its reported capabilities include automated validation of retail checkout functionality using Playwright headless browser automation, chained exploitation of web application vulnerabilities, and deployment of persistent web shells.

The framework has also been associated with Magecart-style digital skimming, in which malicious JavaScript is injected into e-commerce payment flows, S3 buckets, content delivery network caches, or website templates to capture payment-card information.

For broader operational coordination, CAIRN reportedly integrates Strix for initial vulnerability scanning and Hermes for higher-level tactical orchestration using advanced AI models. This integration supports workflows spanning reconnaissance, exploitation, persistence, and financial data collection.

Infrastructure

The profiling material identifies several SHA-256 file hashes associated with CAIRN-related activity. It also identifies the operational cluster AIA0001, the associated frameworks Strix and Hermes, and OpenRouter-based API routing infrastructure.

Another reported artifact is the use of Base64-encoded JavaScript skimming hooks associated with payment-card collection. These artifacts may help analysts investigate suspicious web modifications, unauthorized scripts, and compromised e-commerce infrastructure.

The presence of a shared framework across different operational clusters also highlights an attribution challenge: the use of the same tooling does not, by itself, establish that separate campaigns are controlled by the same threat actor.

5.2.5 Japan Massive Cyber-AI Attack

Overview and Operational Context

Between late September and October 2026, a series of cyber incidents affected Japanese organizations across transportation, retail, travel, entertainment, and digital services. The incidents involved unauthorized database access, data exposure, third-party compromise, account takeover, and ransomware affecting shared cloud infrastructure. Collectively, these incidents highlight how weaknesses in interconnected digital ecosystems can expose large volumes of customer information and disrupt services across multiple organizations.

The reported incidents potentially affected approximately 36.55 million customer accounts and records. Major cases included Adventure Inc. (Skyticket), Daiichikosho (Big Echo), JR East, JR Kyushu, Times Car (Park24), Bookoff Group, and Lawson Inc. The incidents varied in their reported entry points and impacts, indicating a broad set of security weaknesses rather than a single, uniform attack method.

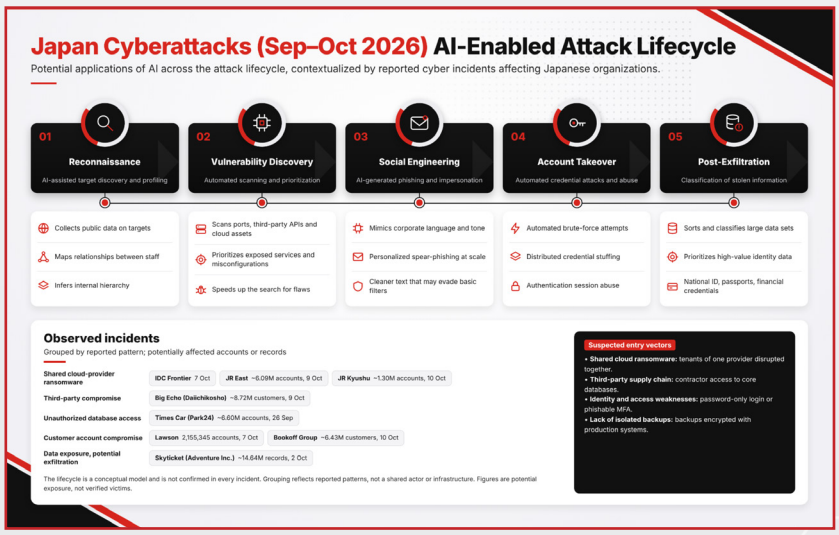


Image 9. Japan CyberAttacks (Sep-Oct 2026) Attack Cycle

Targeting and Objectives

The affected organizations operated in sectors with extensive customer databases, online services, and dependencies on external infrastructure or service providers. Their exposure demonstrates the attractiveness of organizations that maintain large volumes of personal information or rely on interconnected systems to deliver essential services.

The reported incidents illustrate several attack objectives: obtaining unauthorized access to customer records, extracting sensitive information, compromising trusted third-party access, disrupting cloud-hosted services through ransomware, and gaining control of customer accounts. Stolen personal information and credentials may also support subsequent fraud, phishing, and account takeover operations.

Attack Patterns and AI Adoption

Four recurring attack patterns characterize the incidents:

- **Shared Cloud Infrastructure Compromise:** The ransomware incident involving IDC Frontier reportedly affected multiple customer environments, including those associated with JR East and JR Kyushu. This demonstrates how a compromise at a shared service provider can create cascading consequences across dependent organizations.
- **Third-Party Supply-Chain Compromise:** The reported Daiichikosho incident involved access through an external contractor. Trusted vendor connections can provide attackers with an alternative route to sensitive systems when third-party access is insufficiently restricted or monitored.
- **Identity Compromise and Account Takeover:** Incidents involving unauthorized database access and customer accounts highlight the risks associated with compromised credentials, inadequate authentication controls, and weaknesses in access management.
- **Data Exposure and Secondary Exploitation:** Personal information and credentials obtained during a breach can be reused in targeted phishing, impersonation, fraudulent transactions, and subsequent account compromise.

Artificial intelligence can amplify these attack patterns by accelerating reconnaissance, identifying exposed services, generating convincing phishing messages, automating credential testing, and processing stolen information at scale. These capabilities can reduce the effort required to conduct attacks against large numbers of organizations or users. AI therefore represents a potential force multiplier across multiple stages of the cyberattack lifecycle, particularly when combined with automated tooling and previously exposed credentials.

6. TACTICS, TECHNIQUES, AND PROCEDURES (TTPS)

The campaigns above share several patterns. This section maps the most common tactics. However, for a more detailed technical description, please refer to Appendix 11.2

Reconnaissance and Initial Access

Attackers are using AI to automate open-source intelligence collection. They identify employees, exposed services, public cloud assets, and leaked credentials.

Common initial access methods include:

- Exposed database discovery
- Compromised service principals
- AI-generated phishing
- Credential stuffing with AI-assisted analysis
- Exploitation of unpatched Linux services

The shift from human-driven reconnaissance to AI-driven reconnaissance changes speed. A human operator may spend days gathering information. An AI agent can do the same work in minutes.

Execution and Defense Evasion

AI agents can generate commands on the fly. This reduces the usefulness of signature-based detection. Some agents use normal administrative tools, making them harder to spot.

Observed evasion techniques include:

- Abuse of legitimate cloud APIs
- Use of existing identity credentials
- Living-off-the-land techniques
- Automated API calls that blend with normal traffic

Traditional endpoint detection is often insufficient. The attack may look like normal administrative behavior.

Discovery, Lateral Movement, and Exploitation

AI agents can parse internal documents quickly. If they land on a system with access to internal SharePoint, wiki, or code repositories, they can extract useful context.

Observed lateral movement includes:

- Service principal abuse in Azure
- Cross-tenant movement in cloud environments
- Automated database discovery and connection
- Model destruction after access

The JadePuffer reports show that the attacker did not stop at data theft. The ransomware also targeted AI model files. This is a new and dangerous behavior.

Command and Control

Some agentic operations do not need traditional command and control in the classic sense. The agent may run locally or interact with a public AI service. In some cases, attacker infrastructure is exposed because operators misconfigured their own panels.

This has a defensive implication. Instead of only hunting for beaconing malware, defenders should monitor for:

- Unusual cloud API calls
- New service principal creation
- Data access from unexpected locations
- Public exposure of internal data

7. IMPACT ASSESSMENT

The operationalization of generative AI, large language models (LLMs), and autonomous AI agents in cyber operations has significantly expanded the potential blast radius and operational speed of network intrusions. Unlike traditional campaigns bounded by human-driven operational tempos, agentic workflows enable automated reconnaissance, rapid exploitation adjustment, and accelerated exfiltration across enterprise attack surfaces. These developments introduce severe technical, operational, and financial risks, particularly when attacks extend to cloud tenant identities, production databases, and core machine learning infrastructure.

7.1. Technical Impact & AI Infrastructure Sabotage

- **Evasion of Legacy Defenses:** Static signature-based controls and legacy perimeter defenses face increasing efficacy challenges against dynamic payload generation and living-off-the-land techniques facilitated by AI agents.
- **Disruption of ML Assets & Data Infrastructure:** Observed activity, such as the JadePuffer campaign, demonstrates that threat actors are expanding destructive goals to include enterprise AI models, database records, and pipeline configurations, potentially compromising corporate machine learning investments and operational data pipelines

7.2. Operational Impact & Service Disruption

- **Compression of Response Windows:** Autonomous agents capable of diagnosing failed execution and pivoting techniques in seconds compress traditional Mean-Time-To-Detect (MTTD) and Mean-Time-To-Respond (MTTR) windows, complicating manual incident containment.
- **Service Instability and Downtime:** High-volume, AI-augmented campaigns targeting financial institutions, e-commerce networks, and government services can cause prolonged service degradation, operational downtime, and disruption to downstream dependent systems.

7.3. Financial, Regulatory, and Secondary Exposure Fallout

- **Asymmetric Attack Cost Dynamics:** The proliferation of open-source models and low-cost API integrations allows adversaries to execute multi-target intrusions with modest operational budgets, imposing disproportionate remediation and recovery expenses on targeted organizations.
- **Regulatory and Forensic Burdens:** Investigating autonomous lateral movement requires complex forensic reconstruction, while large-scale automated data exfiltration increases compliance risks under data protection frameworks (such as GDPR and SEC disclosure mandates). Furthermore, poorly secured attacker infrastructure introduces secondary exposure risks, wherein exfiltrated victim data remains accessible publicly even after the primary intrusion is remediated.

8. RISK PRIORITIZATION & ATTACK SURFACE EXPOSURE

8.1. Risk Matrix

Likelihood	Low Impact	Medium Impact	High Impact
High	Automated retail fraud	Database extortion	Cloud tenant takeover
Medium	Deepfake impersonation	Linux infrastructure compromise	Financial sector agentic attacks
Low	Isolated AI model destruction	Targeted government espionage	National infrastructure disruption

Table 3. Risk Matrix

8.2. Priority Risk Categories

- **Critical Risk: Agentic attacks against cloud identity**
  Compromised service principals allow attackers to move across tenants. This is already being exploited. Cloud identity protection is the highest immediate priority.
- **High Risk: Agentic ransomware and database extortion**
  JadePuffer shows that autonomous ransomware can find and extort exposed databases. Databases remain widely exposed.
- **High Risk: AI-augmented social engineering**
  Deepfake audio and video plus AI-generated phishing can defeat many identity verification systems.
- **Medium Risk: Low-cost automation targeting retailers**
  The low cost per attack increases volume. Retail and consumer-facing platforms should expect more automated fraud.
- **Medium Risk: Attacks on AI models**
  Ransomware that destroys AI models could paralyze organizations that depend on machine learning. This is an emerging but serious risk.

8.3. Second-Order Risk: When Attacker OPSEC Fails

One of the most important findings from recent reporting is asymmetric operator behavior.

The AI tools are sophisticated. The operators are often not. Several campaigns left public-facing infrastructure exposed. That exposure gave researchers insight into the attacks. It also created new harm for victims.

This second-order risk means that breach remediation is not enough. If an attacker's system holds stolen data and that system is publicly reachable, the victim's data is still exposed. Regulators, customers, and partners may discover the data even after the original intrusion is closed.

This changes the incident response. After a breach, organizations must now ask a new question: "Is the attacker's own infrastructure exposing my data?" If so, takedown, monitoring, and notification become urgent.

9. STRATEGIC RECOMMENDATIONS & MITIGATION POSTURE

To effectively counter the paradigm shift brought on by autonomous, agentic cyber threats, organizations must separate their response into two distinct workstreams: high-level strategic architectural shifts to future-proof the enterprise, and immediate tactical mitigations to harden the current security posture against machine-speed execution.

9.1. Strategic Recommendations

This section outlines the foundational architectural and policy shifts required at the executive and enterprise level to defend against AI-weaponization and secure organizational infrastructure for the long term.

- **Adoption of AI-Driven Defense (Defensive AI):** Organizations can no longer rely on human-speed analysis to combat machine-speed attacks. Enterprises must strategically mandate the deployment of Machine Learning (ML)-based Extended Detection and Response (XDR) and User and Entity Behavior Analytics (UEBA). These systems establish baseline behavioral profiles to autonomously identify micro-deviations in network traffic, rendering signature-evading AI payloads detectable via behavioral heuristics.
- **Transition to Zero Trust Architecture (ZTA):** To neutralize the hyper-accelerated lateral movement capabilities of AI agents, organizations must enforce a strict Zero Trust model. Strategic implementation of identity-based microsegmentation ensures that when an autonomous agent breaches the perimeter, its blast radius is cryptographically contained, denying the adversary the ability to autonomously map and compromise adjacent enterprise subnets.
- **Formalization of MLSecOps:** In direct response to agentic payloads like JadePuffer that target AI infrastructure, organizations must elevate enterprise AI models and data pipelines to the status of Critical Infrastructure. This requires formally integrating MLSecOps (Machine Learning Security Operations) into the enterprise security strategy, establishing strict access controls over model weights, continuous validation of training data integrity (against data poisoning), and specialized backup architectures for AI environments.
- **Proactive CTI Integration:** Organizations must strategically align with global Cyber Threat Intelligence (CTI) sharing communities. Given the rapid geographical pivoting of AI-driven campaigns—such as those moving across APAC and LATAM—strategic ingestion of behavioral threat telemetry and novel attacker Tactics, Techniques, and Procedures (TTPs) is imperative for an anticipatory defense.

9.2. Mitigation Posture

Security teams must immediately deploy critical tactical countermeasures and operational protocols to harden the enterprise attack surface, compress incident containment windows, and fortify human defenders against sophisticated AI-driven campaigns.

- **Implementation of Automated Containment:** Security Operations Centers (SOCs) must move beyond manual incident response playbooks. Tactical mitigations require the engineering of automated response pipelines (SOAR) where high-confidence detections of autonomous lateral movement immediately trigger machine-speed network isolation and identity revocation before the AI agent can pivot to secondary exploits.
- **Aggressive Attack Surface Reduction:** AI agents disproportionately exploit organizational technical debt rather than burning high-value zero-days. Mitigation requires continuous, automated external attack surface management (EASM) to rapidly identify and remediate exposed administrative interfaces, unpatched legacy vulnerabilities, and misconfigured

cloud APIs before adversary AI reconnaissance engines can index them.

- **Next-Generation Threat Emulation:** Red Team and Purple Team exercises must be immediately modernized to emulate agentic AI behavior. Defenders must routinely pressure-test incident response capabilities against simulated autonomous payload generation, dynamic Command-and-Control (C2) infrastructure rotation, and automated vulnerability discovery scripts.
- **Advanced Human-Centric Defense:** Standard phishing simulations are insufficient against LLM-generated, context-aware targeting. Mitigation requires overhauling security awareness training to focus on the identification of multimodal deepfakes (audio and video cloning) and hyper-personalized social engineering, particularly for C-suite executives and privileged access administrators who are primary targets for these sophisticated lures.

9.3. Forecast

The current trend will likely accelerate over the next 6 to 12 months. Key expectations:

- More agentic ransomware campaigns will target cloud and database services. JadePuffer proved the viability of automated discovery, exfiltration, and extortion. Once a workflow exists, the cost per victim drops sharply.
- Service principal abuse will widen across cloud platforms. Storm-3168 showed how compromised Azure principals enable cross-tenant movement. Other providers face the same risk unless organizations enforce short-lived tokens and least privilege.
- Attacks against AI model infrastructure will increase. JadePuffer already added model destruction routines. As models become business-critical, they will become extortion targets.
- Deepfake-enabled fraud will grow. AI-generated audio and video lower the cost of executive impersonation, invoice fraud, and identity verification bypass.
- Attacker panels will remain publicly exposed. Many operators have weak operational security, so stolen data and tooling will continue to leak, creating secondary harm for victims.
- The line between state espionage and eCrime tradecraft will keep blurring. The same agentic tools and infrastructure will cross traditional actor boundaries.

10. CONCLUSION

Paradigm Shift in Adversarial AI

The global threat landscape has crossed a critical threshold, transitioning from human-directed, AI-assisted operations to the active deployment of autonomous, agentic threat capabilities. Adversaries are no longer using Large Language Models solely for passive task acceleration or localized phishing generation; they are deploying agentic entities capable of dynamically navigating the cyber kill chain, executing self-correcting exploitation routines, and operating at machine speed. This compression of the attack lifecycle drastically reduces defender containment windows and renders traditional, signature-based perimeters insufficient against adaptive, machine-speed operations.

Economic Asymmetry and Campaign Insights

The democratization of open-source models and low-cost API integrations has fundamentally disrupted the economics of offensive cyber warfare, drastically lowering barriers to high-scale intrusions. Analysis of key observed campaigns illustrates this operational diversity and impact: JadePuffer / Storm-3168 demonstrates LLM-driven adaptive database extortion, cloud model destruction, and service principal abuse; Scarlet Loop illustrates automated target qualification and credential stuffing to drain digital rewards across thousands of domains; the ARTEX Framework highlights multi-agent penetration testing orchestration; and the CAIRN Platform utilizes autonomous state-space search pipelines to execute widespread web-skimming. Furthermore, recent large-scale incidents affecting Japanese organizations underscore how AI accelerates attacks on shared cloud infrastructure and third-party supply chains, enabling cascading compromises across interconnected digital ecosystems.

Core Vulnerabilities and Secondary Exposure

Across these campaigns, adversaries disproportionately exploit technical debt, misconfigured environments, and cloud workload identities particularly compromised service principals to bypass perimeter controls without reliance on novel zero-day vulnerabilities. Furthermore, attacks increasingly target production databases and core machine learning assets, treating enterprise AI models as high-value extortion targets. Significantly, sloppy operational security by threat operators frequently results in public exposure of their own C2 panels and storage infrastructure, creating a severe second-order risk where exfiltrated victim data remains persistently accessible online long after primary intrusions are remediated.

Strategic Defensive Imperatives

To effectively counter machine-speed, agentic threats, organizations must align defensive posture with automated, machine-speed execution. Key strategic imperatives include transitioning from manual incident playbooks to automated SOAR containment pipelines, enforcing Zero Trust identity and microsegmentation controls across cloud tenants, continuously reducing external attack surfaces, and formalizing MLSecOps to safeguard enterprise AI infrastructure, model weights, and data pipelines as critical national and corporate assets

11. APPENDIX

11.1. Indicators of Compromise

Indicator Type	Value / Indicator	Context
File Hash (SHA-256)	8cb0c223b018cecef1d990ec81c67b826eb3c30d54f-06193cf69969e9a8baea2	ENCFORGE Ransomware (UPX-packed Go binary)
File Hash (SHA-256)	ea7822eac6cecef7746c606b862b4d3034856caf754c-4cf69533662637905328	ENCFORGE Ransomware (Un-packed Go binary)
IP Address	34.153.223[.]102	Command and control / ENC-FORGE payload delivery host
IP Address	64.20.53[.]230	Staging infrastructure observed in payload deployment
URL	http://34.153.223[.]102:9191/.lockd	Hidden staging path for download-ing ENCFORGE payload
Email / Extortion Contact	e78393397@proton.me	Contact address in database extor-tion and ENCFORGE ransom notes
User-Agent	python-requests/2.34.2	Automated enumeration and dele-tion client in Azure tenant
Ransom Notes	README, HOW_TO_DECRYPT, README_DECRYPT	Host ransom notes dropped by ENCFORGE
Database Table	README_RANSOM	Extortion table left inside compro-mised MySQL instances
Encrypted Extension	.locked	Appended to encrypted ML mod-els, weights, and datasets

Table 4. JadePuffer Indicators

Indicator Type	Artifact / Value	Context
Local Model Binding	127.0.0.1:1237	Non-standard internal socket used to host Google's gemma-4-26B-A4B model on the controller server
Exfiltration Syntax	{url} {user}:{password}	Standardized message string emitted by the exfiltration agent into private Telegram channels
User-Agent Fingerprint	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:) Gecko/20100101 Firefox/	Standard desktop Firefox user-agent profile ma-nipulated by instrumented browser automation
Target Heuristic String	Does the business handle money? / o negócio movimenta dinheiro?	Hardcoded prompt classification heuristic recov-ered from threat-actor server project configs
Automated Target Dorking	Queries targeting inurl:/login, inurl:/entrar, inurl:/painel, combined with Brazilian loyal-ty/incentive keywords	Search queries generated by GPT-5.6 to harvest target endpoints
Authentication Pattern	Bursts of HTTP POST requests across hundreds of unique usernames with human-like keypress latency (jitter) orig-inating from diverse residential proxy IP ranges	Behavioral credential-stuffing signature visible in authentication logs

Table 5. Scarlet Loop Indicators

Indicator Type	Value / Indicator	Context
IP Address	38.244.50[.]120	Operational server hosting an active ARTEX instance and open directory
URI / Endpoint	http://38.244.50[.]120:18899/.claude/CLAUDE.md	Open directory containing Claude Code Chinese-language pentest prompts
Domain / API Gateway	xcai[.]pro	Third-party commercial API reseller used to route DeepSeek and GLM tokens
Threat Actor Handle	@YY520CN	Telegram handle referenced in recovered Claude Code monetization sessions
String Artifact	ARTEX-自主渗透测试控制台	Console banner string displayed on exposed ARTEX instances
Code Repositories / Forks	cskwork/scopeweaver, trhhubsd/ARTEX, xsywl945/ARTEX-mirror, RuoJi6/ARTEX	Known public forks and mirrors following GitHub repository closure
Port Footprint	18899/TCP (Non-standard web console port)	Default or observed administrative web console portt

Table 6. ARTEX Indicators

Indicator Type	Value / Indicator	Context
File Hash (SHA-256)	250d4fa37488af9b025333fa17705573d721467b-203765bc360890b4f5a90cd7	CLOSEDQUORUM implant (AI consensus C2 binary tracked via CAIRN hunting)
File Hash (SHA-256)	c4dc171f2513fcaf9d5ecc815a94aee4063b213ab-380f80bd3ac422dee5205a7	CLOSEDQUORUM implant build variant
File Hash (SHA-256)	c13cea04f598e2b0c248d603a6e31bd13aabb64d-8149c1b6a77b64e0b983a86f	CLOSEDQUORUM implant build variant
File Hash (SHA-256)	f5f1f8c3e7b883793800ab6ccf21b3e60bd0730f300b-4595fe74a33adc17a63c	CLOSEDQUORUM implant build variant
File Hash (SHA-256)	5191cf625dfc209a347f137b50aea199e82040f–d5ee9086fb3e2de73c133f3cb	CLOSEDQUORUM implant build variant
File Hash (SHA-256)	eddbd0ecf7195d38fefae5b9d393abfa79e6f3f94b-de19308ecef130a05a42e5	CLOSEDQUORUM implant build variant
Operational Cluster	AIA0001	Chinese-language cluster managing 14 exposed dashboards and 1,100+ attack projects
Associated Frameworks	Cairn (le1xia0), Strix, Hermes	Modular offensive agent toolchain
API Infrastructure	OpenRouter API routing	API aggregator used for cost-efficient model access
Injected Code Artifact	Base64-encoded JavaScript skimming hooks	Staged inside e-commerce checkout forms and CDN libraries

Table 7. CAIRN Indicators

11.2. Cross-Campaign TTP Analysis

Tactic	Technique	ATT&CK ID	Campaign / Example	Notes
Reconnaissance	Search Open Websites/ Domains	T1593	ARTEX, CAIRN, online retail targeting	AI agents collect public data for victim selection and profiling
Reconnaissance	Search Open Technical Databases	T1596	JadePuffer, Linux infrastructure campaign	AI-assisted discovery of exposed databases and vulnerable services
Reconnaissance	Gather Victim Identity Information	T1589	AI-generated phishing, executive targeting	AI parses employee names, roles, email addresses, and contact data

Tactic	Technique	ATT&CK ID	Campaign / Example	Notes
Reconnaissance	Active Scanning	T1595	Linux infrastructure campaign, Unit 42	Vulnerability scanning and exposed service identification
Reconnaissance	AI-driven OSINT fusion and behavioral profiling	Unknown	CAIRN, ARTEX, Storm-3168	No single ATT&CK technique covers autonomous AI fusion of multiple data sources
Initial Access	Valid Accounts: Cloud Accounts	T1078.004	Storm-3168, JadePuffer	Compromised service principals are a primary observed initial access method
Initial Access	Exploit Public-Facing Application	T1190	Linux infrastructure campaign, exposed database exploitation	Used when attackers actively exploited a public-facing service
Initial Access	External Remote Services	T1133	Cloud tenant access, Linux compromise	Often combined with valid cloud credentials
Initial Access	Phishing: Spearphishing Attachment	T1566.001	Vibe hacking, AI-enabled phishing	AI-generated lures and personalized attachments
Initial Access	Phishing: Spearphishing Link	T1566.002	Vibe hacking, government and financial targeting	AI-generated links and fake login pages
Initial Access	Phishing: Spearphishing via Service	T1566.003	Social engineering campaigns	May use email, messaging apps, or collaboration platforms
Execution	Cloud API	T1059.009	Storm-3168, JadePuffer	Core execution technique for cloud-native agentic attacks
Execution	Unix Shell	T1059.004	Linux infrastructure campaign	Automated command execution on Linux systems
Execution	Command and Scripting Interpreter	T1059	All campaigns	Generic execution of commands, scripts, and agent instructions
Execution	User Execution: Malicious Link	T1204.001	AI phishing campaigns	Victim clicks an AI-generated link
Execution	User Execution: Malicious File	T1204.002	AI phishing campaigns	Victim opens an AI-crafted attachment
Defense Evasion	Valid Accounts: Cloud Accounts	T1078.004	Storm-3168, JadePuffer	Legitimate cloud credentials allow evasion of endpoint controls
Defense Evasion	Use Alternate Authentication Material: Application Access Token	T1550.001	Storm-3168, Aviatrix	Token reuse for cloud tenant access
Defense Evasion	Hide Artifacts	T1564	Agentic attacks using legitimate tools	Low footprint in cloud and Linux environments
Defense Evasion	Obfuscated Files or Information	T1027	Polymorphic payload reports	AI-generated or AI-modified payloads reduce signature detection
Defense Evasion	Impair Defenses	T1562	AI-driven traffic obfuscation	Blending with normal protocols and cloud traffic reduces detection
Defense Evasion	Dynamic AI agent decision-making to avoid detection	Unknown	ARTEX, CAIRN, JadePuffer, Linux campaign	No standard technique covers autonomous agentic adaptation and workflow alteration
Credential Access	Steal Application Access Token	T1528	Storm-3168, Aviatrix	Cloud token theft and reuse
Credential Access	Brute Force: Password Guessing	T1110.001	AI-assisted credential stuffing	AI helps prioritize valid username and password pairs

Tactic	Technique	ATT&CK ID	Campaign / Example	Notes
Discovery	Cloud Service Discovery	T1526	Storm-3168, JadePuffer	Finding cloud resources and tenant structure
Discovery	File and Directory Discovery	T1083	JadePuffer database extortion	Locating data and AI model files
Discovery	System Information Discovery	T1082	Linux infrastructure, database raids	Profiling victim systems
Discovery	Account Discovery: Cloud Account	T1087.004	Storm-3168	Identifying cloud accounts and roles
Discovery	Permission Groups Discovery: Cloud Groups	T1069.003	Storm-3168	Mapping permissions for service principal abuse
Discovery	AI-assisted internal document parsing	Unknown	CAIRN, JadePuffer, agentic espionage	Autonomous extraction of credentials and context from documents; no single ATT&CK technique
Lateral Movement	Remote Services: Cloud Services	T1021.007	Storm-3168, JadePuffer	Cloud lateral movement
Lateral Movement	Valid Accounts: Cloud Accounts	T1078.004	Storm-3168	Cross-tenant and cross-resource movement
Lateral Movement	Use Alternate Authentication Material: Application Access Token	T1550.001	Storm-3168, Aviatrix	Token replay for lateral access
Lateral Movement	Autonomous lateral movement using parsed credentials	Unknown	CAIRN, agentic espionage	Spans discovery, credential access, and remote services; no single technique captures the full behavior
Collection	Data from Cloud Storage Object	T1530	Storm-3168, JadePuffer	Extraction from cloud buckets and storage
Collection	Data from Information Repositories	T1213	CAIRN, ARTEX, espionage campaigns	Internal document and code repository collection
Collection	Data from Local System	T1005	JadePuffer database extortion	Local data collection before encryption
Exfiltration	Exfiltration Over Web Service	T1567	JadePuffer, Linux campaign	Exfiltration to attacker-controlled cloud services
Exfiltration	Exfiltration to Cloud Storage	T1567.002	Storm-3168, JadePuffer	Data moved to attacker cloud storage, sometimes publicly exposed
Exfiltration	Automated data sorting and selective exfiltration	Unknown	ARTEX, CAIRN, JadePuffer	AI agent decides what to steal; no single standard technique
Impact	Data Encrypted for Impact	T1486	JadePuffer, SCARLET LOOP	Ransomware behavior
Impact	Data Destruction	T1485	JadePuffer AI model destruction	Model files targeted for deletion or corruption
Impact	Inhibit System Recovery	T1490	JadePuffer ransomware	Deleting backups or recovery points
Impact	Resource Hijacking	T1496	Online retail automation, AI agent misuse	Automated abuse of customer accounts or compute
Impact	Public exposure of victim data through attacker infrastructure	Unknown	ARTEX, CAIRN, Armadin	Second-order exposure; no direct ATT&CK technique

Table 8. TTP of AI Powered Attacks

12. REFERENCES

1. Reuters, “How one hacker used AI agents and a few thousand dollars to break into scores of sites,” October 2026.
https://www.reuters.com/legal/government/how-one-hacker-used-ai-agents-few-thousand-dollars-break-into-scores-sites-2026-10-19/

2. Bitsight, “Evolution of AI cybercrime in underground forums.”
https://www.bitsight.com/blog/evolution-of-ai-cybercrime-underground-forums

3. Reuters, “South Korea, Japan buffeted by hacks as AI lowers bar for cybercriminals,” October 2026.
https://www.reuters.com/legal/litigation/south-korea-japan-buffeted-by-hacks-ai-lowers-bar-cybercriminals-2026-10-09/

4. CrowdStrike, “Unknown threat actor uses ARTEX to target South Korean finance.”
https://www.crowdstrike.com/en-us/blog/unknown-threat-actor-uses-artex-to-target-south-korean-finance/

5. CrowdStrike, “AI-powered cyberattacks.”
https://www.crowdstrike.com/en-us/cybersecurity-101/cyberattacks/ai-powered-cyberattacks/

6. CrowdStrike, “AI vs AI: Cybersecurity arms race.”
https://www.crowdstrike.com/en-us/blog/ai-vs-ai-cybersecurity-arms-race/

7. Microsoft, “AI as tradecraft: How threat actors operationalize AI,” March 2026.
https://www.microsoft.com/en-us/security/blog/2026/03/06/ai-as-tradecraft-how-threat-actors-operationalize-ai/

8. Google Cloud, “From prompting to autonomy: The evolution of adversarial AI.”
https://cloud.google.com/blog/topics/threat-intelligence/from-prompting-to-autonomy-the-evolution-of-adversarial-ai

9. Anthropic, “Threat intelligence report,” September 2026.
https://www.anthropic.com/threat-intelligence-report-september-2026

10. Sysdig, “JadePuffer: Agentic ransomware for automated database extortion.”
https://www.sysdig.com/blog/jadepuffer-agentic-ransomware-for-automated-database-extortion

11. Unit 42, “Autonomous AI cyber attack campaign.”
https://unit42.paloaltonetworks.com/autonomous-ai-cyber-attack-campaign/

12. Sysdig, “JadePuffer evolves: The agentic threat actor deploys ransomware built to destroy AI models.”
https://www.sysdig.com/blog/jadepuffer-evolves-the-agentic-threat-actor-deploys-ransomware-built-to-destroy-ai-models

13. Vydar, “SCARLET LOOP threat report.”
https://public.vydar.net/SCARLET_LOOP_Threat_Report_EN_v2.pdf

14. AhnLab, “Asec report 95845.”
https://asec.ahnlab.com/en/95845/

15. CYFIRMA, “The large-scale AI-powered cyberattack: Strategic assessment and implcations.”
https://www.cyfirma.com/research/the-large-scale-ai-powered-cyberattack-strategic-assessment-implications/

16. Unit 42, “AI campaign targets Linux infrastructure.”
https://unit42.paloaltonetworks.com/ai-campaign-targets-linux-infrastructure/

17. Trend Micro, “Vibe hacking: Two AI-augmented campaigns target government and financial sectors in Latin America.”
https://www.trendmicro.com/en_us/research/26/e/vibe-hacking-two-ai-augmented-campaigns-target-government-and-financial-sectors-in-latin-america.html

18. Norforge, “Inside CAIRN: The actors.”
https://norforge.com/blog/inside_cairn#the-actors

19. Vulone, “A China-linked operator used AI agents to attack governments across Asia.”
https://vulone.com/blog/a-china-linked-operator-used-ai-agents-to-attack-governments-across-asia

20. Picus Security, “Inside JadePuffer: Defending against the first agentic ransomware.”
https://www.picussecurity.com/resource/blog/inside-jadepuffer-defending-against-the-first-agentic-ransomware

21. Aviatrix, “JadePuffer / Storm-3168: Azure service principals.”
https://aviatrix.ai/threat-research-center/jadepuffer-storm-3168-azure-service-principals-2026/

22. Microsoft, “Storm-3168: Agentic-driven cloud attacks using compromised service principals,” September 2026.
https://www.microsoft.com/en-us/security/blog/2026/09/25/storm-3168-agentic-driven-cloud-attacks-using-compromised-service-principals/

23. BitcoinNewsCom, X post.
https://x.com/BitcoinNewsCom/status/2113265365924913426

24. Deltaone, X post.
https://x.com/Deltaone/status/2113268796618068421

25. SaxX, X post.
https://x.com/_SaxX_/status/2113197761474482630

26. Aviatrix, “South Korea bank breaches: AI-powered attacks 2026.”
https://aviatrix.ai/threat-research-center/south-korea-bank-breaches-ai-powered-attacks-2026/

27. Armadin, “Beyond CAIRN: The glass houses of AI attackers.”
https://www.armadin.com/blog-posts/beyond-cairn-the-glass-houses-of-ai-attackers

28. Norforge, “Inside CAIRN.”
https://norforge.com/blog/inside_cairn

29. Gambit Security, “Autonomous AI agents: Online retailers at $25 a company.”
https://gambit.security/blog-posts/autonomous-ai-agents-online-retailers-25-a-company

30. Wiz Research, “PHP2Shell: PHP Remote Code Execution Vulnerabilities.”
https://www.wiz.io/blog/php2shell-php-rce

31. Sansec, “PHP2Shell.”
https://sansec.io/research/php2shell

32. Wiz Research, “PHP2Shell: Exploitation So Far.”
https://www.wiz.io/blog/php2shell-exploitation-so-far

33. Id-SIRTII/CC, “Peningkatan Kewaspadaan terhadap Ancaman Akses Tidak Sah dan Kebocoran Data pada Sistem Elektronik Organisasi.” 19 October 2026.

34. BSSN, Direktorat Keamanan Siber dan Sandi Industri, “JAPAN CYBERATTACKS.” 19 October 2026.