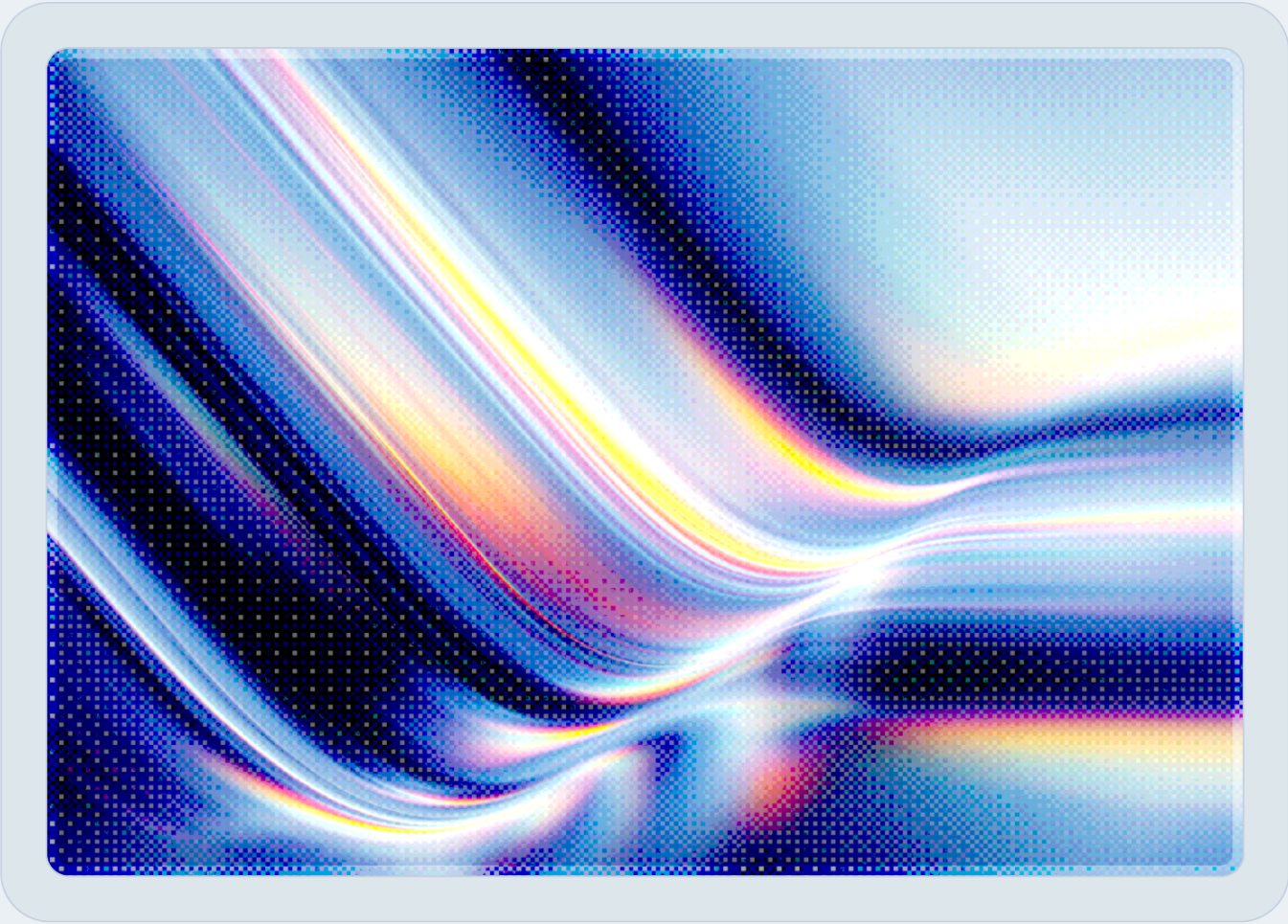
CSAI Foundation | Cloud Security Alliance

Japan's H1 2026 Ransomware Surge Hits SMEs, Qilin Uses AI Scripts

2026-09-18

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

© 2026 Cloud Security Alliance. Some rights reserved.

You may download, store, display, view, print, redistribute, and link to this document in its original, unmodified form, provided that attribution to the Cloud Security Alliance is maintained and all trademark and copyright notices remain intact.

This document may not be modified or altered. You may quote portions of the document as permitted by the Fair Use provisions of the United States Copyright Act, provided that attribution is given to the Cloud Security Alliance.

This document may be shared on professional and social media platforms in its original form with attribution.

This document was generated with AI assistance and has not undergone official CSA review and approval processes.

Key Takeaways

Cisco Talos has documented 90 ransomware incidents affecting Japanese organizations between January and July 2026, a 4.7 percent increase over the 86 incidents recorded in the same period of 2025, with small and medium-sized enterprises absorbing the overwhelming majority of the damage [1]. Organizations with capital below JPY 1 billion (roughly $6.7 million) accounted for 78 percent of victims, up from 69 percent a year earlier, and firms capitalized below JPY 100 million alone made up 48 percent of all incidents [1]. The threat landscape has also reshuffled: The Gentlemen, a ransomware-as-a-service operation that only began operating in mid-2025, led Japan-specific activity with 14 confirmed incidents, while Qilin and SafePay followed with seven each, and Talos noted that "very few groups active during the same period last year" remained active in 2026 [1]. Most significantly, Talos assessed with medium-to-high confidence that several Python tools recovered from a Qilin-linked open directory — including scripts designed to destroy Active Directory infrastructure and disable Veeam backups — were generated with the assistance of a large language model, based on structured documentation-style comments, redundant step-by-step logging, and command history referencing an "llm_chatbot" directory [1]. The finding extends a pattern CSA has tracked since mid-2026 in which ransomware operators use AI coding assistants to accelerate tool development rather than to invent fundamentally new attack techniques, a dynamic that shifts the defensive burden toward behavioral detection and away from signature matching [2][3].

Background

Talos's semiannual review of the Japanese ransomware landscape found that the country's threat profile in the first half of 2026 diverged sharply from the prior year's, even as the overall volume of attacks held roughly steady [1]. The near-total turnover in which groups were most active — with legacy operators largely absent and newer entrants like The Gentlemen dominating incident counts — is consistent with, though not confirmed by, a ransomware-as-a-service ecosystem in which affiliates migrate toward platforms offering the most reliable payouts and technical support; law enforcement takedowns and voluntary rebranding among criminal operators could produce a similar pattern of turnover without that dynamic being the cause. The Gentlemen itself illustrates how fast such migration can happen: the group first appeared in July 2025 and by the second quarter of 2026 had become one of the most active ransomware operations tracked globally by claimed-victim count, claiming credit for more than 328

victims in the first five months of the year alone and expanding across manufacturing, construction, healthcare, government, and IT targets on nearly every continent [4][5]. Qilin, by contrast, is a more established operation that rebranded from an earlier group called Agenda in 2022 and has since built one of the largest ransomware-as-a-service affiliate programs by claimed-victim count in the criminal ecosystem, reportedly claiming more than 1,400 victims worldwide between mid-2025 and mid-2026 [6].

The sectoral and size distribution of Japanese victims is consistent with either deliberate targeting of under-resourced firms or an opportunistic pattern that disproportionately affects the most numerous category of target; the data alone does not distinguish between the two. Manufacturing accounted for 34 percent of all incidents, more than three times the next-largest category [1]. This concentration is plausibly linked to manufacturing's weight in Japan's export-oriented economy and to smaller subcontractors' tendency to run older, less-monitored industrial and IT infrastructure, though Talos's report does not itself quantify this mechanism. Information and communications firms (11 percent) and services businesses (9 percent) rounded out the top three sectors [1]. The concentration on smaller firms, summarized in the table below, is consistent with a pattern described elsewhere in ransomware-as-a-service research, in which affiliates favor targets with weaker security operations and less capacity to negotiate or absorb a prolonged outage over targets offering a larger but less certain ransom.

Capital size (JPY)	Share of H1 2026 victims	Share of H1 2025 victims
Under 100 million	48%	Not separately reported
Under 1 billion (cumulative)	78%	69%
1 billion and above	22%	31%

Source: Cisco Talos [1].

This targeting pressure on smaller organizations may matter beyond Japan's borders. Many of Japan's small and medium manufacturers sit inside multinational supply chains as parts suppliers, subcontractors, or logistics partners, which raises the possibility — though this report does not document a specific instance — that a ransomware outage at a single undercapitalized firm could cascade into production delays or data exposure for far larger customers that never appear in the incident's own headline statistics.

Security Analysis

The most consequential finding in Talos's report is not the incident count but the evidence that Qilin-linked actors are now using AI coding assistance to build their operational tooling. During an investigation into a Qilin-affected environment, Talos researchers gained access to an open directory containing several Python scripts and assessed, with medium-to-high confidence, that the code exhibited hallmarks of AI generation rather than conventional handwritten malware [1]. One script, identified as `deadman.py`, was built to deploy destructive wiper payloads through Active Directory Group Policy Objects, and contained detailed, structured comments describing the tool's workflow in a style Talos characterized as documentation-like rather than typical of hand-rolled offensive tooling [1]. A second script, `veeam_kill.py`, was designed to disable Veeam backup infrastructure — a common precursor step ransomware operators take to prevent victims from restoring data without paying — and was organized into four clearly labeled sequential steps with consistent, redundant logging at each stage [1]. A third tool, `deploy_locker.py`, went further still, including prose-style documentation describing the script's purpose, prerequisites, and usage examples in a manner closely resembling output a developer would request from a coding assistant rather than notes a criminal operator would typically bother writing for internal tooling [1]. Talos also recovered bash history entries showing commands that inspected directories named "llm_chatbot," providing corroborating, if circumstantial, evidence that the operators were interacting directly with a large language model during development [1].

These characteristics — systematic step-by-step structuring, redundant progress logging at a uniform level of detail, and comments that read as explanatory documentation rather than terse operational notes — are consistent with patterns CSA has documented in other AI-assisted intrusions this year, including a June 2026 case in which an attacker used an LLM to "vibe-code" an Active Directory enumeration script that carried similar tells: an unedited placeholder value, over-engineered fallback logic, and a debugging-session-style filename [3]. In both cases, the underlying attack techniques — wiper deployment via Group Policy, backup-service disruption, and AD reconnaissance — are not novel; ransomware operators have used all three for years. What AI assistance appears to change is the speed and ease with which an operator, potentially with limited scripting experience, can produce working code to execute them, and the extent to which smaller or less technically sophisticated RaaS affiliates can now field capabilities that previously required a more experienced in-house developer [2][3]. This mirrors the "friction reduction, not new technique" framing CSA has used to describe AI-assisted ransomware development more broadly: based on the cases documented so far, large language models appear to be lowering the cost of building functional attack tooling more quickly than corresponding defensive techniques are maturing [2].

The defensive implication is that syntactic or signature-based detection of these specific scripts is likely to have a short shelf life, because the underlying LLM can regenerate functionally equivalent code with different comments, variable names, and structure on request. What should remain stable, and therefore worth building detection around, is the behavioral sequence each script executes: unauthorized modification of Group Policy Objects, particularly ones that create or alter startup scripts or scheduled tasks across many machines simultaneously; abrupt termination or disabling of Veeam or other backup agent processes; and bulk enumeration of Active Directory objects inconsistent with an account's normal activity. Organizations that build detection logic around these behavioral fingerprints rather than around any single sample's code signature are likely to retain coverage even as Qilin or other AI-assisted operators iterate on their tooling.

Recommendations

Immediate Actions

Organizations, and particularly small and medium-sized enterprises in manufacturing and adjacent sectors that this data shows are disproportionately targeted, should inventory and disable any internet-accessible devices, VPN concentrators, or remote administration services that are not actively required, since Talos found that initial access in these campaigns centered on VPN vulnerabilities, misconfigurations, and stolen credentials rather than specific zero-days [1]. Multi-factor authentication should be enforced without exception on all VPN, remote desktop, and administrative accounts as a baseline control against credential-based intrusion, though the specific initial-access vector varies by actor: The Gentlemen's intrusions have been linked to exploitation of exposed edge infrastructure such as FortiGate and Cisco appliances [5], while Talos's Japan data points more broadly to VPN vulnerabilities, misconfigurations, and stolen credentials [1]. Security teams should also alert on any unauthorized or unexpected changes to Group Policy Objects and on the abrupt termination of backup agent processes such as Veeam, since both are late-stage indicators that a ransomware deployment is imminent or already underway and represent one of the last practical windows for intervention before encryption or destruction begins.

Short-Term Mitigations

Security patches should be applied promptly across internet-facing systems, and any unsupported or end-of-life operating systems and applications — often found among smaller manufacturers running legacy industrial or line-of-business software — should be prioritized for replacement or, where replacement is not immediately feasible, network isolation [1]. Organizations should extend access

control reviews and monitoring to third-party vendors, subsidiaries, and supply-chain partners; given Japan's concentration of undercapitalized manufacturing subcontractors, a compromise at one node could plausibly propagate to larger partners through shared network access or data flows, even though this report does not document a specific such cascade. Endpoint detection and response tooling should be tuned to flag suspicious remote access patterns and privilege escalation attempts rather than relying primarily on file-hash or signature-based rules, a posture that is increasingly necessary given the demonstrated ability of AI-assisted operators to regenerate functionally equivalent but syntactically novel tooling on demand [1][2].

Strategic Considerations

The concentration of ransomware impact on smaller organizations in Japan is consistent with reporting on RaaS targeting patterns elsewhere, in which affiliates appear to favor victims with weaker security operations over those offering the largest potential ransom, and defenders should expect this dynamic to persist and likely intensify as AI-assisted tooling lowers the technical bar for less experienced affiliates to operate effectively. Organizations that have historically treated their limited size as a reason to under-invest in security monitoring should reconsider that calculus in light of data showing SMEs now represent the clear majority of victims in at least one mature ransomware market. More broadly, security leaders should treat the appearance of AI-generated tooling in an active ransomware campaign as further evidence — following two prior CSA-documented cases [2][3] — that this is a recurring rather than isolated phenomenon, and should incorporate AI-assisted attack development into threat models and detection engineering roadmaps accordingly.

CSA Resource Alignment

CSA's The Attacker's Coding Partner: AI-Assisted Ransomware Development [2] is the most directly relevant prior CSA analysis, having documented in June 2026 how large language models are lowering the technical barrier to functional ransomware development and EDR evasion across multiple threat actors; the Qilin findings in this note are a concrete, named-actor confirmation of the friction-reduction dynamic that report described in general terms, and its recommendation to prioritize behavioral over signature-based detection applies directly to the Group Policy and backup-disruption tooling documented here. CSA's AI-Generated PowerShell Malware Hits Active Directory [3], published in July 2026, analyzed a separate June 2026 intrusion in which an attacker used an LLM to build an Active Directory enumeration script bearing similar AI-authorship tells — structured documentation-style comments and redundant logging — and its argument that AI functions as a force multiplier rather than a source of new attack primitives is directly borne out by the Qilin case, since none of the destructive

techniques Talos observed (GPO-based wiper deployment, backup disruption, AD enumeration) are themselves new. Both CSA notes point to the AI Controls Matrix (AICM) v1.1 [7] as the relevant control framework, particularly its Threat and Vulnerability Management and Application and Infrastructure Security domains, which set expectations for behavioral detection capability and backup and recovery resilience that map directly to this incident's Immediate Actions and Short-Term Mitigations above.

References

[1] Cisco Talos. "Ransomware incidents in Japan in the first half of 2026: Investigation of The Gentlemen's infrastructure and evidence of Qilin's AI use." Cisco Talos Blog, September 2026.

[2] Cloud Security Alliance. "The Attacker's Coding Partner: AI-Assisted Ransomware Development." Cloud Security Alliance, June 2026.

[3] Cloud Security Alliance. "AI-Generated PowerShell Malware Hits Active Directory." Cloud Security Alliance, July 2026.

[4] The Cyber Express. "The Gentlemen Ransomware Group's Global Rise In H1 2026." The Cyber Express, 2026.

[5] Check Point Research. "Thus Spoke...The Gentlemen." Check Point Research, 2026.

[6] SOCRadar. "Dark Web Profile: Qilin (Agenda) Ransomware." SOCRadar, 2026.

[7] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.1." Cloud Security Alliance, 2026.