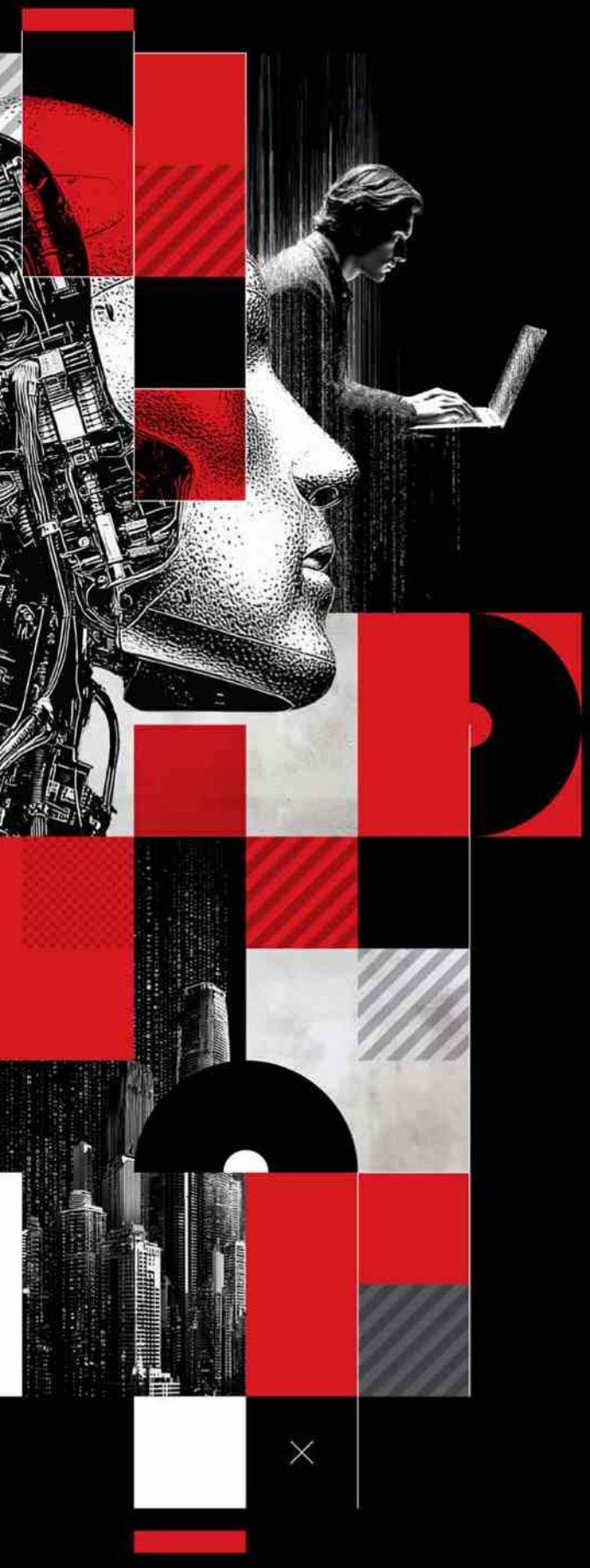
Trend Research

TREND MICRO

THE AI-FICATION OF CYBERTHREATS

TREND MICRO
SECURITY PREDICTIONS
FOR 2026



CONTENTS

As we enter 2026, the cybersecurity landscape is being reshaped by the convergence of AI, automation, interconnected systems, and large-scale operations. The tools, tactics, and procedures that once required coordinated human effort can now be executed rapidly and at scale through highly automated infrastructures. The line between manual and machine-driven activity is increasingly blurred across the threat ecosystem. What was once a digital arms race has evolved into a contest of speed, adaptability, and precision, where the advantage belongs to those who can act and react the fastest.

AI has become a critical component of many core business workflows across industries. But while enterprises have been busy integrating AI into their systems for productivity, threat actors have been doing the same, but for malicious activities. What began as simple automation to assist with phishing and basic intrusion tasks has evolved into large-scale, coordinated operations capable of delivering targeted attacks, fraud campaigns, and system compromises with minimal human input. The barrier to cybercrime has shifted from needing deep technical expertise to simply knowing how to use AI-powered tools. In many ways, AI hasn't just augmented cyberthreats; it has industrialized them.

At the same time, the growing complexity of enterprise systems and the pace of technological change are amplifying existing risks. Modern organizations depend on a web of digital platforms, third-party services, and interconnected supply chains that extend well beyond their direct control. A single weak link — be it a misconfigured system, an exposed API, or a compromised vendor — can provide an entry point for widespread disruption. As organizations pursue efficiency and innovation, maintaining visibility and control across the digital ecosystem has become one of the most pressing challenges for security teams.

In 2026, the threat landscape will be defined not only by what attackers do but also by how efficiently they can scale and adapt their operations. Enterprises will need to shift from reactive defense to anticipatory resilience, embedding security into every stage of AI adoption and automation. The winners of this new era will be those who can innovate securely, combining human oversight, AI safety, and adaptive defenses.

In this report, we explore how cyberthreats are transforming across key attack surfaces, from the escalation of advanced persistent threats (APTs) and the continued evolution of ransomware to increasingly targeted supply chain compromises and cloud-native attacks. Across these areas, a clear trend is emerging: Modern attacks are no longer defined by individual tools or tactics, but rather by the capacity to scale, automate, and coordinate disruption with unprecedented efficiency.

AI

- AI will become both a transformative force and a top attack vector, driving fully autonomous, adaptive, and scalable threats across digital and physical systems.
- Agentic AI will act with growing autonomy, executing multi-step operations and interacting with real systems, turning compromised agents into operational attack vectors.
- The growth of vibe coding will accelerate innovation while simultaneously increasing the risk of unsecure code in organizations that do not implement proper review processes.
- AI-powered deception will reach new heights as deepfakes, hallucinations, and automated social engineering campaigns erode trust and overwhelm traditional defenses.

APTs

- Emerging collaboration models will enable APTs to share access, infrastructure, and payloads, obscuring attribution and accelerating global operations.
- Supply chain and insider threats will converge as state-sponsored operatives infiltrate vendors and enterprises, embedding malicious code or exploiting privileged access from within.
- AI-driven tactics will bypass traditional defenses, while compromised pipelines and open-source repositories will become key attack vectors.
- Geopolitical tensions will drive targeted attacks against critical infrastructure, defense, and strategic industries, heightening risks of espionage, disruption, and cyber conflict.

Enterprise

- Legacy systems, outdated software, and hidden IT debt will remain major enterprise risks, providing attackers with persistent entry points beyond the reach of modern defenses.
- Identity-based and trust-driven attacks will surge as AI automates phishing, session hijacking, and social engineering, making deception more convincing and detection harder.
- AI-driven agents and generative scams will outpace traditional identity and access management (IAM) and phishing defenses, potentially exposing organizations to credential theft, impersonation, and large-scale fraud.
- The line between human and machine insiders will blur as compromised employees, AI agents, and third-party tools alike become vectors for espionage, data theft, and disruption.

Cloud

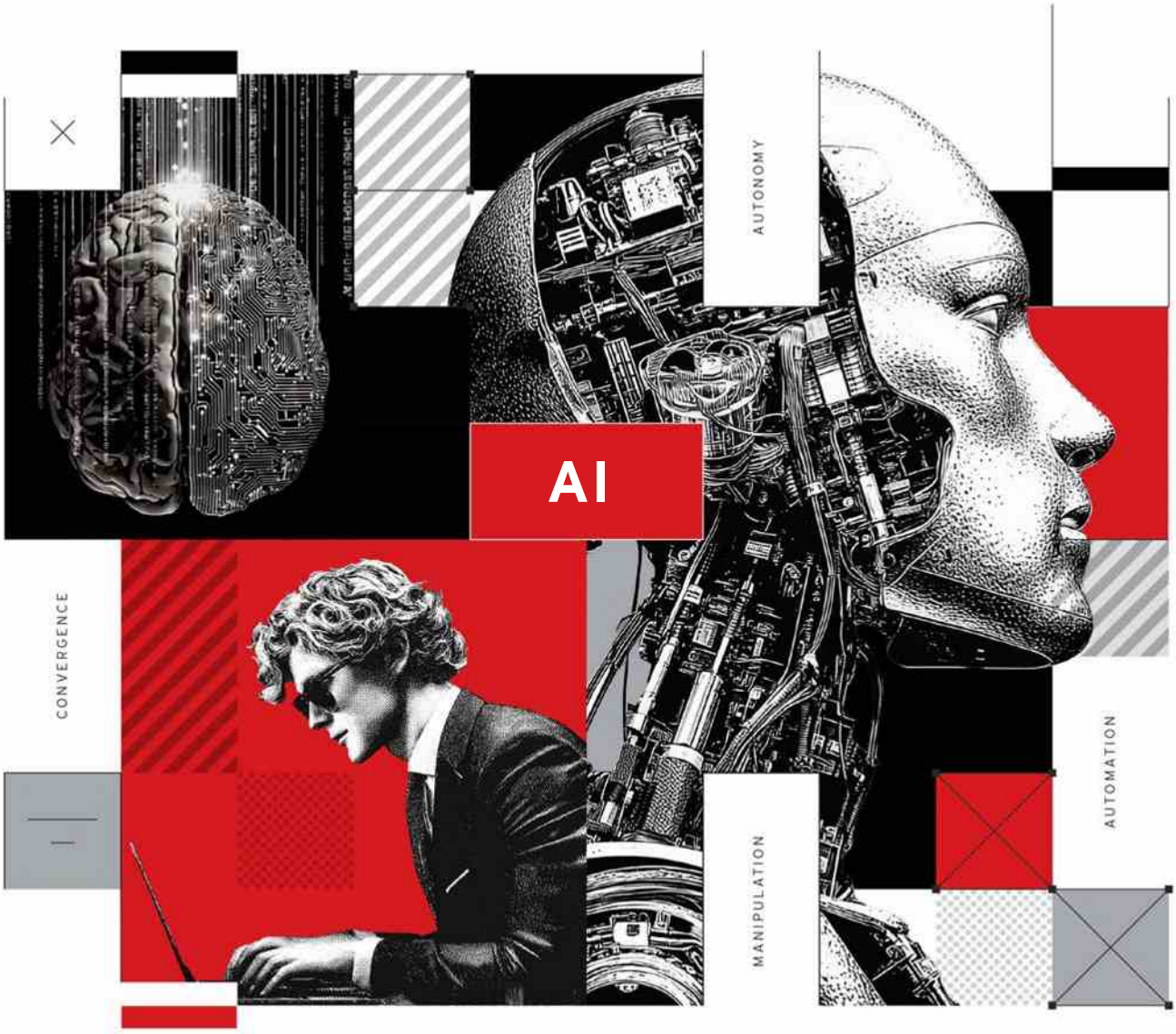
- Cloud environments will remain prime targets as adoption continues to grow, with attackers exploiting high-value workloads, operational dependencies, and hybrid infrastructures.
- Cloud-native phishing campaigns — blending email, SMS, voice, and AI-driven tactics — will become more sophisticated, targeting users and organizations.
- Misconfigurations, overprivileged credentials, exposed APIs, and unsecure containers will remain primary attack vectors, enabling lateral movement, data exfiltration, and supply chain compromise.
- Multi-cloud and hybrid setups will introduce new blind spots, while GPU-based cloud resources will be increasingly exploited for malicious activities.

Ransomware

- Ransomware will evolve into AI-driven, fully automated operations that scan, exploit, and extort with minimal human input.
- Attackers will shift from pure encryption to intelligent data exploitation, using AI to identify and pressure victims' most sensitive assets.
- Supply chains, open-source components, and AI-integrated workflows will become key entry points, allowing ransomware to infiltrate trusted systems while blending into normal enterprise activity.
- Increasing automation and ransomware-as-a-service (RaaS) tools will democratize attacks, enabling even low-skill actors to launch complex, adaptive campaigns.

Vulnerabilities

- AI will accelerate both the discovery and exploitation of zero-day vulnerabilities, enabling faster reconnaissance, automated exploitation, and broader attack reach.
- New risks will emerge from AI-enabled environments, including prompt injection attacks, model backdoors, and vulnerabilities in inference servers and frameworks.
- Supply chains, open-source libraries, and AI model repositories will remain prime targets for attackers seeking widespread impact.
- Blind spots like unpatched IoT/OT devices, edge appliances, and AI-enabled environments will provide attackers with footholds for lateral movement and exploitation.



Increased Automation, Evolving Systems, and Social Engineering Risks

The availability of large language models (LLMs) and their use in coding tools and integrated development environments (IDEs) gave rise to vibe coding, highlighting the shift toward AI-assisted development, where teams could prototype, iterate, and deploy applications in a significantly shorter amount of time. Based on our telemetry data, vibe coding tools such as Lovable and Vercel have seen significant increases this year in the number of vibe-coded web apps hosted on their platforms, with Vercel seeing a 57% increase and Lovable experiencing a whopping 660% increase from January to September 2025.

However, vibe coding is a double-edged sword. Its ability to rapidly create tools and proof-of-concept (PoC) projects — deploying vibe-coded modules into production software or daily business processes — carries significant risk, especially without knowing what sorts of vulnerabilities might have been inadvertently injected into the code by the vibe-coding AI. The use of vibe coding was found to generate unsecure code 45% of the time,[1] with its widespread adoption potentially leading to the growth of vulnerable and exploitable applications. In fact, we anticipate a commonly generated vulnerability from vibe coding to be widely exploited in the near future.

Aside from generating unsecure code, LLMs can also hallucinate nonexistent libraries. A threat actor can, for example, create libraries using hallucinated library names in an attack called "slopsquatting."[2] While we have not seen attackers exploiting such hallucinations, it's only a matter of time before a threat actor registers commonly hallucinated libraries to infiltrate development codebases and insert the libraries into the software supply chain.

Agentic AI is no longer a distant concept; it's already being deployed in enterprise environments, changing how organizations operate and introducing new cybersecurity challenges. Agentic AI systems make autonomous decisions, execute complex tasks, and interact with digital and physical environments with minimal human oversight.[3] While these capabilities offer significant operational advantages, they also create new risks.[4] High autonomy reduces user oversight, allowing malicious actions to go unnoticed until the damage is done. Deep integration into the enterprise ecosystem amplifies the impact: A single compromise can provide attackers with access to connected systems and accounts. Once compromised, agentic AI systems can manipulate data, disrupt operations, and even control physical devices.[5]

Unlike monolithic state-of-the-art foundation models, these systems are connected to tools, MCP (Model Context Protocol) servers, and the real world. They do not just respond to prompts — they reason, plan and act across multiple steps, integrate with real-world data and tools, and execute workflows in feedback loops.

As the enterprise adoption of agentic AI accelerates, particularly through workflow platforms such as Dify and n8n, new vulnerabilities will emerge. Exposed webhooks and rogue NPM packages can enable workflow compromise or denial-of-service (DoS) attacks.[6, 7]

However, threat actors do not always need to target the agentic system itself; they can also manipulate the surrounding infrastructure, inject poisoned modules, or exploit shared orchestration layers, therefore subverting trusted AI agents into performing malicious actions. Subtle attacks, such as prompt injections, can silently hijack multi-agent workflows and influence downstream behavior without leaving obvious traces.[8] By mapping which services and platforms organizations are adopting, attackers can strategically position themselves to exploit the associated weaknesses.[9]

Agentic AI is evolving toward AI operating systems[10] and swarm intelligence,[11] where multiple agents coordinate to achieve advanced objectives. This evolution means that compromising the orchestrator (or a key agent within the swarm) could result in unintended actions across the system. At the same time, the emergence of agentic edge AI,[12, 13] which are deployed in factories, vehicles, and houses, introduces further risks: As AI agents interact directly with physical systems, they could become vectors for safety hazards or operational disruptions.[14] Some organizations might even adopt agentic AI in sensitive domains without sufficient safeguards, increasing the likelihood of operational, safety, or security incidents.

Agentic capabilities are not just lucrative for business; they also appeal to cybercriminals and nation-state actors. Agentic systems automate routine work, scale decision-making, and execute multistep operations without human intervention. Malicious actors can deploy autonomous agents to automate fraud; run large, highly personalized social engineering campaigns; and orchestrate ransomware or supply chain attacks at scale.

Phishing and social engineering, once constrained by human labor, will become fully automated and hyper-personalized. Traditional attack methods, such as botnets, command-and-control (C&C) servers, and worms, will be reimagined with AI, dynamically generating exploit code, adapting payloads, and propagating autonomously.

Agentic AI shifts cybercrime from a service model to one in which AI acts as the operational "servant," orchestrating attacks at a speed, scale, and complexity previously unimaginable. This shift will force a fundamental rethink of defensive posture, as legacy security solutions struggle to keep up: Defenders will require agentic AI defensive platforms capable of performing triage, acting, and containing attacks at machine speed.

By connecting to tools, APIs, and execution systems, agentic AI itself becomes a more attractive target for threat actors.[15] If an attacker manipulates an autonomous marketing agent into generating and distributing promotional assets containing misused branding or malicious links, the agent might unwittingly execute those tasks at scale, amplifying the impact of a single prompt and turning hallucination or deception into a direct attack vector.

While traditional AI systems rely on human oversight for validation, agentic AI can act autonomously, verifying actions only after execution — or not at all — allowing hallucinated outputs to become operational events.[16] This risk becomes more severe when agents trust one another's outputs by default.

In automated-to-automated (A2A) workflows, one agent's hallucination can quickly propagate into another agent's decision-making, forming a self-reinforcing loop of incorrect actions without any human noticing. For example, a forecasting agent might hallucinate a surge in demand and pass that data to an ordering agent. The ordering agent, assuming the input is valid, then executes a large stock purchase or logistics change. The result is a real-world operational impact caused entirely by automated systems acting on one another's unchecked hallucinations.

In 2026, agentic AI will represent both a transformative opportunity and a formidable security challenge. Its autonomy, deep integration, and capacity to execute complex actions will elevate risks across the enterprise, from cybercrime and supply chain compromise to operational and physical disruption. The next generation of defense must evolve alongside agentic AI systems, treating agents as accountable identities and securing the broader ecosystems in which they operate. Only by doing so can organizations mitigate the rapidly growing threats and fully realize the potential of agentic AI.

We will see the rise of autonomous agents authorized to make significant operational decisions – altering supply chains, issuing refunds, or deploying services – where a single error or hallucination can propagate through interlinked systems.[17] For example, if an inventory management agent hallucinates a stock shortage and automatically initiates a large emergency order, that error could propagate through procurement, logistics, and finance before any human notices the issue. The operational and reputational damage compounds with each automated step.

Autonomous agents will operate with increasing authority across critical systems, but with that rise comes a growing challenge of accountability. When these agents make errors, responsibility becomes difficult to assign.[18] A recent study shows that generative AI increasingly clashes with zero trust architecture principles since verification and audit trails are easily bypassed.[19] A common scenario can be seen in customer service: An agent might hallucinate a nonexistent refund rule and automatically issue payments, leaving no clear trace of who approved them.

AI threats to supply chains and infrastructure are poised to escalate, driven by the extensive use of open-source code and packages, and of interconnected AI ecosystems. An incident involving a malicious MCP server that occurred this year exemplifies these risks.[20] This highlights the risk of open-source code or packages (which can be compromised through malicious updates, hijacked developer accounts, or actors who initially appear benign) setting a foothold in AI ecosystems. Overall, while supply chain and infrastructure compromises are not new to the cybersecurity industry, their expansion into AI ecosystems might result in attacks that use innovative tactics.

The risk is likely extended to A2A, along with other tool-calling protocols. This shift aligns with the wider adoption of agentic AI and agent marketplaces. The developing marketplace ecosystem introduces more complicated dependencies on software and external data, thus broadening the attack surface of AI's supply chain.

AI-generated images and videos, also known as deepfakes, will continue to play a significant role in social engineering schemes.[21] Perhaps the most disturbing trend is the proliferation of "nudifying" applications. These tools represent a new low in digital exploitation, transforming innocent social media photos into sexualized deepfake images. Often available cheaply or free, these services have created an ecosystem of harassment, extortion, and reputational destruction. The psychological toll on victims is severe: long-term trauma, damaged careers, and in extreme cases, tragic outcomes. This isn't a theoretical threat either — it is happening now, and the technology is only becoming more accessible and convincing, making it one of the most dangerous threats of 2026.

Finally, we have observed malicious GGUF models being uploaded to Hugging Face, which is currently working with partner companies for detection and takedowns.[22] In this scenario, attackers can upload tampered models and execute the backdoors, once they are deployed for inference. Even though there are ways to mitigate the risks, we foresee minor incidents of this sort in 2026. We recommend that enterprises set up a policy to reject the installation of non-preapproved AI models, unless they run in a sandbox environment.

For Cybersecurity Professionals

Security teams will face increasingly autonomous and adaptive threats as AI shifts from a supporting tool to a fully operational actor. The rise of vibe coding, agentic AI, and automated multi-agent workflows introduces vulnerabilities across codebases, supply chains, and interconnected systems, allowing errors, vulnerabilities, hallucinations, or malicious manipulation to propagate rapidly. Traditional monitoring and detection methods will struggle to keep pace, as attacks become hyper-personalized, fully automated, and capable of exploiting both digital and physical systems, significantly increasing cybersecurity risks.

For Decision-Makers

The industrialization of AI-driven operations and attacks magnifies organizational exposure, as a single compromised agent or workflow can trigger operational failures, financial losses, and reputational damage across entire ecosystems. The growing autonomy of AI in critical business processes, combined with deep integration into supply chains and infrastructure, means that organizational resilience is, in part, directly tied to AI security, with errors, hallucinations, or malicious manipulations capable of producing large-scale disruption before human oversight can intervene.

Organizations are rapidly adopting AI tools to boost productivity, but this also creates significant cybersecurity risks, effectively giving potential intruders "built-in" attack capabilities within the network. Protecting enterprises in 2026 and beyond will demand a forward-leaning, adaptive security posture that goes well beyond traditional user or network defenses. Security strategies must evolve along with AI-driven and agentic systems, embedding safety guardrails, continuous verification, and accountability mechanisms across every AI workflow and application. Regular adversarial simulations and automated red-team exercises should be conducted to test the durability of LLM- and AI-powered software, transforming the uncertainty of rapid development into a clear view of an organization's security readiness.

The use of agentic AI should require robust trust frameworks, continuous monitoring, and verified audit trails to prevent errors, unchecked hallucinations, and exploitable vulnerabilities. Furthermore, AI-generated code should be evaluated for security risks and unapproved AI models confined to isolated sandbox environments.

As AI-powered social engineering, deepfakes, and fraud automation become more convincing, defenders should move away from content-based detection and instead deploy trust verification systems that authenticate sender identity and communication origins across all channels. Enhanced identity and access management (IAM) with continuous authentication, behavioral analytics, and zero trust principles is essential to detecting anomalous or malicious agent activity. By constantly monitoring AI agents and securing the broader ecosystem, organizations can effectively manage new risks while capitalizing on the operational benefits of AI.

Adaptive Campaigns, Supply Chain Compromise, and Insider Exploitation

In 2026, advanced persistent threat (APT) operations will evolve rather than reinvent itself, driven by sophisticated collaboration models and AI integration that optimizes key attack stages while enabling more adaptive and efficient campaigns. We anticipate that threat actors will share resources and infrastructure to accelerate operations and enhance stealth tactics, with more groups taking advantage of purchasing direct access rather than conducting their own reconnaissance. This evolution toward shared operational hubs means faster attack timelines, reduced effort for initial compromise and lateral movement, and a more efficient attack ecosystem.

We foresee group coordination evolving through the "premier pass-as-a-service" model.[23] Instead of operating alone, threat actors form alliances where one group gains and maintains access to a target, then passes that foothold to another to execute the espionage phase. Typically, APT coordination falls into four categories: type A, shared infection vectors with loose coordination; type B, coordinated supply chain attacks; type C, one group deploying another's payloads; and type D, one group providing provision of an operational box for another. The "premier pass" model fits squarely within types C and D, which require the highest level of coordination, showing how advanced and organized these partnerships have become. By merging access brokering with operational outsourcing, these campaigns obscure attribution, accelerate intrusions, and expand their global reach across different sectors.

APTs are increasingly using AI to scale and enhance their cyberattacks. AI allows APT actors to automate complex operations, reducing the need for human oversight and increasing the speed, scale, and adaptability of their attacks. AI can assist during reconnaissance, lateral movement during intrusions, data theft, and making detection harder. For instance, AI agents can blend malicious activities with normal processes, making detection more challenging. AI-driven reconnaissance enables APT actors to efficiently map targets' infrastructures and identify vulnerable systems like edge devices.

This provides attackers with a significant advantage, as they need only to identify a single vulnerability while defenders must secure all potential entry points. AI-enhanced open-source intelligence (OSINT) enables attackers to better profile their targets, leading to more precise and targeted attacks. This marks a shift from AI-assisted to AI-operated attacks, lowering barriers for less skilled actors while enhancing the effectiveness of their operations.

APTs will rely heavily on AI, machine learning, and automation for reconnaissance, persistence, and evasion. Generative AI will enable context-aware and highly personalized social engineering techniques that closely mimic legitimate communications. Meanwhile, malicious AI chatbots on the dark web (e.g., WormGPT and FraudGPT)[24] will democratize the creation of convincing phishing and other harmful content as crime-as-a-service.

We will likely see cases of APT campaigns that capitalize on leaked enterprise-tuned AI models to obtain internal knowledge of their targets, using a malicious MCP server or another AI component as an unmonitored attack vector. By infiltrating enterprise AI knowledge bases, as observed in cases of exposed and leaked vector databases, APTs can glean proprietary insights into their victims to create targeted exploits, much like in the reconnaissance phase in traditional campaigns. Prepositioning efforts, such as the silent compromise of critical infrastructure and telecommunications, will persist.

Attackers will employ AI-driven and pipeline-based living-off-the-land (LotL) techniques. These include AI-powered malware that, after establishing a foothold, will use LLM connections to generate context-specific commands that rely solely on native host tools and binaries. Since these actions mirror legitimate administrator behavior, endpoint detection tools are more likely to miss them. One prime example of this was the S1ngularity case, where AI command-line tools (specifically Claude, Gemini, and Q) were used to aid in reconnaissance efforts on compromised systems.[25] Threat actors are also likely to exploit legitimate automation and scripts within CI/CD pipelines to execute malicious actions. These pipelines are prime targets because they often store sensitive secrets, such as API keys and cloud credentials, and frequently have excessive privileges that provide direct access to production environments.

Additionally, APT actors are likely to refine their use of AI for disinformation campaigns on social media. According to US law enforcement,[26] Russia-aligned APT actors began using Meliorator as early as 2022. Meliorator is an AI-enabled bot farm generation and management software used to disseminate disinformation to and about some countries, including countries in the EU, Israel, Ukraine, and the US.

Several AI companies closely monitor the usage of their online offerings by APT groups. To evade their monitoring efforts, these threat actors, particularly those aligned with countries like China, Iran, North Korea, and Russia, are expected to shift away from US-based AI products to developing homegrown AI tools. This transition could make it more challenging for Western law enforcement and cybersecurity industries to monitor and counteract their activities effectively.[27]

Supply chain compromise will continue its transition from a high-risk tactic to a critical and sustainable component of APT campaigns. Instead of focusing solely on current software vendors, attackers will increasingly target alternative attack entry points, such as the abandoned software ecosystems and open-source repositories seen in the TAOTH campaign.[28] Combined with collaborative attack models, where one compromised vendor sells or shares access with multiple other APT groups, a single supply chain attack can enable widespread, shared exploitation of the same victim base.

In 2026, we expect nation-state actors to intensify efforts to infiltrate organizations by embedding operatives who pose as legitimate employees, creating the ultimate insider threat.[29] A likely scenario involves the deployment of IT workers overseas who rely on forged identities, deepfake-assisted interviews, and AI-generated personas to evade verification processes. Once inside, these operatives can exploit privileged access for espionage, data theft, and blackmail.

Geopolitical tensions will further drive APTs toward espionage and disruption of AI infrastructure as much as critical infrastructure. Critical infrastructure industries, such as defense, energy, finance, and telecommunications, will be priority targets for both espionage and destructive campaigns, as cyber operations become increasingly linked with conflicts and national-level strategy. This could trigger disruptions to global infrastructure and digital networks.

APT groups will focus on the drone, maritime, aerospace, and telecommunications sectors — all critical to modern defense and secure communications — to steal proprietary technologies, disrupt supply chains, and gather intelligence. Threat actor tactics will also evolve toward advanced infiltration of air-gapped systems and strategic supply chain compromise targeting defense subcontractors and vendors.

We foresee that English-speaking underground threat actors will continue to evolve in sophistication and scale, expanding into non-English forums, and offering "access-as-a-service" business models to strengthen their criminal operations. Their progression from forum participation to offering specialized services such as compromised-account access, AI-driven social engineering, and cashout schemes, signals their growing role as a central force in the cybercrime landscape.[30]

Looking further into the future, as quantum computers approach the ability to break current public-key encryption, the risk of "harvest now, decrypt later" attacks is increasing. Threat actors could intercept and store encrypted data today with the intention of decrypting it once quantum capabilities mature, potentially exposing highly sensitive information long after it was initially transmitted. This growing threat emphasizes the need for organizations to begin migrating long-lifespan and critical data to post-quantum cryptography well in advance of the National Institute of Standards and Technology (NIST) 2030–2035 deprecation timeline.[31]

For Cybersecurity Professionals

Defenders must shift from signature-based detection to behavior monitoring and automated response. CI/CD pipelines, AI agents, and supply-chain tools should now be treated as high-risk assets. Insider threat programs must evolve to catch synthetic or AI-assisted insiders, not just careless employees.

For Decision-Makers

APT campaigns will increasingly threaten critical infrastructure and global supply chains, with attacks amplified by AI-based social engineering. Geopolitical tensions will direct targeting toward sectors essential to national security and commerce. Meanwhile, emerging technologies like quantum decryption have the potential to create long-term strategic risks.

Today's highly connected, AI-powered, hybrid environment is highly interesting for APT actors as it provides multiple areas of vulnerability and entry, as well as multiple data stores that are of interest to them. Organizations should prioritize a layered defense strategy that combines principles such as strict access controls, microsegmentation, continuous device verification, and phishing-resistant multifactor authentication (MFA), with rigorous patch management, especially across edge infrastructure and CI/CD pipelines. Given the ease of generating new fake identities, requiring out-of-band verification for sensitive transactions will greatly help in countering deepfake-enabled fraud. Meanwhile, automation workflows, AI models, and supply chain access points must be secured from the design phase, supported by ongoing monitoring, employee training against AI-generated phishing, and proactive threat hunting with incident response plans to prepare for compromise.

Legacy Systems, Identity Risks, and Operational Blind Spots

The enterprise threat landscape in 2026 will see a combination of innovation, AI automation, and the exploitation of legacy, bespoke, and emerging technologies. One of the major challenges organizations continue to face heading into 2026 is the persistence of legacy systems and outdated IT products. Many businesses still rely on unpatched or unsupported hardware and software that are integrated into daily operations. These systems represent potential hidden vulnerabilities and blind spots that might not be easily addressed by modern security solutions, flaws that threat actors can exploit to launch attacks. Without aggressive modernization or network isolation, what was once manageable IT debt in 2025 could become a major security risk in 2026.

We expect Identity-based attacks to gain momentum, with phishing-as-a-service, adversary-in-the-middle (AiTM), and session hijacking techniques evolving through AI automation. Current IAM systems were designed for humans and long-lived service accounts, not for agents that spin up, call tools, delegate tasks, and disappear. These agents often rely on broad API keys and static secrets, making them prime targets for credential theft and lateral movement. As AI agents increasingly coordinate across services, inter-agent trust will emerge as a new attack vector for privilege escalation and the weaponization of other agents. Phishing, identity theft, and watering-hole attacks will persist as core tactics, enhanced by AI tools capable of scraping social media and crafting personalized lures using the gathered information.[32]

IAM will become a central concern for enterprises. Since AI agents often rely on broad API keys and static secrets that persist through sessions, they create new opportunities for credential theft, lateral movement,[33] and privilege escalation.[34] Impostor agents – whether hijacked or maliciously injected – will emerge, impersonating users or systems and operating under legitimate privileges. Without continuous checks, delegation logs, and behavior monitoring, these agents might go undetected by legacy authentication systems,[35] leaving organizations vulnerable to unauthorized access.

In 2026, phishing and online scams will evolve into more natural and convincing forms, powered by advances in generative AI for text and voice synthesis. Traditional detection models that relied on unnatural phrasing or typographical errors will lose their effectiveness, as AI-generated messages become indistinguishable from human-written ones. Threat actors are beginning to train AI systems on their victims' social media and public data, allowing messages to be tailored to each target's organization, role, and personal communication style.

Even basic personalization and "natural tone" generation are already possible using publicly available AI tools. In 2026, many attackers will adopt these capabilities as standard. This will especially impact business email compromise (BEC) schemes and phishing attacks by blending seamlessly into normal business workflows (for instance, disguised as legitimate cloud service notifications or corporate communications).

Voice- and chat-based social engineering will also become a more realized threat. With voice synthesis and automated conversational AI, semiautonomous "fraud bots" can now impersonate customer service or financial institution staff and engage with large numbers of potential victims. Fully autonomous fraud systems have yet to emerge, but AI is already amplifying human operators, dramatically increasing the efficiency of scam operations.

This evolution is fueling the rise of scam-as-a-service, where AI tools for generating persuasive texts or voice scripts are sold cheaply on the dark web. As a result, individuals with little technical expertise can now launch linguistically sophisticated, large-scale fraud campaigns.

On the defensive side, traditional awareness training that focuses on spotting "unnatural" messages is becoming obsolete. Future defenses will need to shift toward trust verification architectures, systems that validate who is sending a message and where the communication originates. Organizations must establish multilayered mechanisms that verify sender authenticity and communication integrity, extending beyond AI-content detection alone.

For the longer term, AI systems capable of analyzing the emotional responses of victims and dynamically adjusting their tone or wording — so-called autonomous fraud AI — could emerge. However, this remains a mid-term threat beyond 2026. In the short term, AI will not reinvent scams, but rather amplify their quality, precision, and efficiency.

In 2026, increasingly sophisticated attacks will target authentication mechanisms, with AI enabling threat actors to process stolen credentials more efficiently and identify high-value targets within compromised datasets. Edge devices, often less secure than core infrastructure, will continue to serve as attractive entry points for initial access.

The challenge is compounded by AI agents creating an entirely new class of identity risk. These autonomous systems operate with delegated permissions, making traditional identity verification models inadequate. When an AI agent acts on behalf of a human, verifying intent and detecting compromise or manipulation become critical challenges.

It is important to note, however, that 2026 will not be the year when AI will completely replace human fraudsters, but rather the year when AI will expand and enhance fraud. At the core of every scam remains human trust. AI will serve as a powerful tool to mimic and exploit that trust.

Attackers are evolving toward session-focused, trust-based techniques. Adversary-in-the-browser[36] payloads hijack active sessions, while stolen or replayed tokens can bypass MFA. The objective will increasingly shift from simply stealing credentials to controlling authenticated sessions and human trust. Threat actors will also continue using mobile devices as initial infection vectors, exploiting popular messaging apps like Signal and WhatsApp to deliver payloads, harvest credentials, or establish persistent access.

In 2026, we will see threat actors focusing on cloud accounts, software-as-a-service (SaaS) platforms, IT management tools, and software supply chains, all of which are areas with broad impact potential for enterprises. Misconfigurations, unpatched servers, and compromised updates will remain primary attack vectors, while disabling endpoint detection and response (EDR) systems will be a key persistence tactic.[37]

Insider threats, both intentional and accidental, will remain a constant risk. As previously noted, insider threats from foreign IT workers, such as nation-state operatives, are expected to intensify, with AI likely enabling other threat actors to replicate tactics. Remote work and the increasing reliance on automated systems expand the attack surface, while AI agent "insiders" emerge as new vectors. These AI-driven assistants, often granted access to sensitive data, can be manipulated to expose secrets or credentials if compromised.[38] While difficult to predict, widespread layoffs driven by cost-cutting and efficiency initiatives might also increase the likelihood of employee dissatisfaction within the workforce, leading to potential insider threats, data exfiltration, and increased susceptibility to social engineering or recruitment by malicious actors.

Meanwhile, the rise of AI-powered web browsers introduces a new and largely uncharted attack surface. These browsers, capable of running agentic workflows and integrating with enterprise data, are susceptible to prompt injection, plugin supply chain attacks, and session hijacking. Organizations should treat AI browsers as high-risk assets, applying the same scrutiny as they would to other privileged endpoints.

Many enterprises continue to underestimate the growing risks embedded within everyday tools and software updates. Threat actors are shifting their focus to capturing trusted sessions, tokens, and development pipelines, which grant them persistent and often undetected access. At the same time, overlooked issues, such as shadow SaaS adoption and misconfigured cloud automation, expose sensitive data and credentials. In 2026, the greatest danger will not stem from novel exploits, but from the subtle, ongoing misuse of trusted systems that operate beyond the scope of traditional monitoring.

For Cybersecurity Professionals

In 2026, defenders will face increasingly complex threats as legacy systems, automation, and cloud dependencies combine for a riskier threat landscape. Attackers are shifting from credential theft to session hijacking, token replay, and AI-driven social engineering that manipulates both human and machine trust. The rise of autonomous, AI-enabled malware and insider-like agents will blur the boundaries between internal and external threats, forcing security teams to adapt to faster, stealthier, and more context-aware attacks that exploit everyday business operations.

For Decision-Makers

Exploits hidden within trusted software updates, AI tools, or SaaS platforms can silently undermine entire ecosystems, disrupting operations and damaging trust. As enterprises race to integrate AI and automation for efficiency, they risk amplifying vulnerabilities and becoming too reliant on systems that could be manipulated or even fail altogether. The true danger lies not in novel exploits, but in the invisible misuse of trusted technologies, turning innovation, if unchecked, into a liability. This underscores the growing importance of robust compliance frameworks and evolving regulatory oversight, ensuring that the deployment and governance of AI-driven systems in enterprises align with security, transparency, and accountability standards.

Balancing innovation and security in 2026 will require integrating protection seamlessly into every stage of technological adoption. Rather than relying solely on traditional frameworks, enterprises should focus on resilience, visibility, and continuous validation as core security principles. Security should act as an enabler of innovation, ensuring that automation, AI integration, and modernization efforts are designed with defense in mind.

Escalating Risks, Misconfigurations, and Multi-Cloud Threats

In 2026, cloud adoption will continue to expand across industries, motivating threat actors to prioritize cloud environments for both attacks on high-value targets and as part of their own operational infrastructure. The US government has noted the evolution of threat actor tactics, especially from advanced groups, in a bid to gain initial access to cloud resources.[39]

Over the past year, there has been an increase in sophisticated phishing attacks, such as AiTM[40] and MFA phishing, malicious Azure Entra ID applications,[41] device code phishing,[42] and voice phishing,[43] with threat actors even commercializing these tools as AiTM SaaS services.[44] In 2026, threat actors, driven by the growing reliance on cloud services, will develop increasingly sophisticated cloud-native phishing campaigns, blending traditional tactics like email phishing, smishing (SMS + phishing), and vishing (voice + phishing) with cloud-specific techniques to better evade detection and target cloud users. Furthermore, threat actors will chain vulnerabilities in exposed services to breach cloud environments (for example, combining API flaws with container escapes to gain footholds and pivot to sensitive workloads.).[45]

Major supply chain attacks on package repositories[46] demonstrate how such incidents can indirectly compromise cloud environments. In 2026, these attacks will become even more impactful for cloud systems, even without direct compromise of the environment.[47]

Threat actors have increasingly exploited credentials exposed in public services and artifact repositories. Since nearly every cloud service provider (CSP) now hosts its own repositories for containers and software packages, these platforms have become valuable targets. In 2026, threat actors will intensify their focus on artifact repositories to harvest exposed source code, credentials, and other sensitive data that could fuel future attacks.[48]

In 2026, we will see threat actors targeting SaaS applications with multi-cloud presence, as these platforms often represent weak points in cross-cloud security. By compromising a single SaaS application, attackers can move laterally across multiple CSPs, enabling access and deeper infiltration within interconnected environments.

While most organizations are now operating in multi-cloud environments, our data shows that nearly half (approximately 47%) struggle to maintain full visibility of their cloud assets.[49] This leaves exploitable gaps for attackers who can launch cross-platform attacks or move laterally between on-premises and cloud resources. Integration mechanisms within hybrid setups are sometimes undocumented or unaudited, further exposing organizations to risks.

Data theft and exfiltration, as well as stealing compute resources, remain primary objectives for threat actors, who are likely to develop sophisticated exfiltration techniques that can hinder incident response. With cloud infrastructure software now included in vulnerability bounty programs, there will be both an uptick in security research and a corresponding risk of threat actors abusing these new techniques for malicious purposes.[50]

Misconfigurations will remain a leading cause of cloud breaches in 2026, as attackers continue to exploit common mistakes such as overprivileged credentials. A publicly exposed bucket or overly permissive storage credentials can open the door to large-scale data exposure and, in some cases, supply chain attacks that take advantage of implicit trust relationships. It's no surprise that around three quarters of organizations have faced a serious cloud security incident stemming from misconfiguration.[51]

A single administrator role or API key with overprivileged credentials can inadvertently provide attackers full control if compromised through phishing, leaked credentials, or exposed software.[52] The complexity of cloud IAM systems often leads to overly permissive configurations, which threat actors can exploit to access enterprise data. In 2026, attackers will increasingly target these misconfigured cloud applications using zero-day vulnerabilities or phishing as initial compromise vectors.

APIs and containers continue to provide direct avenues of attack in cloud-native environments. As APIs often serve as the entry point to applications, misconfigurations, such as missing authentication or exposed gateways, can grant unauthorized access and data exposure.[53] Unsecure defaults, such as hardcoded credentials, missing authentication, or exposed sensitive information, further expand the attack surface and, in some cases, could enable large-scale compromises.[54]

Another growing concern is the rise of poisoned container images,[55] in which attackers inject malicious code into widely used or trusted images, allowing the malware to spread across a large number of deployments. Container orchestration platforms like Kubernetes and Docker Swarm remain attractive targets, as attackers can exploit misconfigurations, default credentials, or known vulnerabilities to seize control of entire clusters.

Incentives around access to AI and cloud GPU resources will help drive a rise in cloud attacks. Cloud GPUs have become valuable cloud assets, and attackers will target them to capitalize on powerful compute capabilities for their own purposes or to broker unauthorized access. These threats include parties embargoed from accessing the latest GPU hardware.[56]

In multi-cloud and hybrid environments, the biggest risks are breakdowns in tenant isolation. There have been many exploits discovered that show GPU-specific data leakage risks that undermine tenant isolation, such as NVBleed,[57] LeftoverLocals,[58] and the Nvidia container escape.[59] It's only a matter of time before threat actors find new GPU vulnerabilities that enable sensitive data theft across multi-tenant environments. While the occurrence of this type of attack is not a given in 2026, the possibility of such exploitation in the coming years is real and should shape risk appetite and control design for organizations.

For emerging techniques against cloud native services, we expect threat actors to keep seeding infected items at scale, including consumer grade internet-of-things (IoT) devices, and to migrate these footholds into enterprise IoT device ecosystems. Against this backdrop, attacks that target cloud GPUs will be among the top cloud-focused threats of 2026, be they for direct compute theft, for the resale of compute access, or to steal sensitive data from GPU memory. Organizations should be aware of research analyzing attacks against multi-tenant cloud GPU environments to set control priorities and test plans for 2026.

For Cybersecurity Professionals

Cloud defenders will confront increasingly sophisticated attacks driven by misconfigurations, multi-cloud complexity, and AI-assisted exploitation. Threat actors are chaining API flaws, poisoned containers, and overprivileged credentials to move laterally across hybrid environments with speed and precision. The growing use of SaaS integrations and GPU-based workloads adds new surfaces for data theft, compute hijacking, and cross-tenant breaches. The growing complexity of interconnected systems means that a single overlooked flaw can now impact entire operations.

For Decision-Makers

Cloud risks in 2026 will directly affect business continuity, intellectual property, and reputation. Compromises in multi-cloud or hybrid environments, supply chains, or GPU resources could lead to widespread operational disruption, regulatory exposure, and loss of competitive advantage.

Threat actors are constantly scanning cloud environments for exposed resources to steal valuable data. Organizations should carefully monitor and audit large-scale cloud migrations before production, involving experienced red teams with cloud experience. Audits should continue on a regular basis post-migration to account for ongoing changes, and organizations should prioritize deploying state-of-the-art cloud security solutions to maintain robust protection.

Automated Extortion, Supply Chain Infiltration, and Targeted Exploitation

Ransomware attack volumes in 2026 will be consistent with current trends, but payment rates will decline, leading to more public data leaks as groups seek alternative leverage to get their victims to pay.[60] As with the other threat categories, AI-driven automation will define this next phase, from autonomous vulnerability discovery and penetration testing to automated data analysis that helps attackers prioritize victims, craft targeted coercion tactics, and increase the impact of secondary extortion. Automation will enable ransomware groups to both analyze stolen data for second-order extortion opportunities and rapidly identify and exploit high-value vulnerabilities,[61] allowing them to penetrate deeper within the network and extract maximum opportunities from each breach.[62]

Ransomware groups will continue to rapidly exploit newly disclosed, high-value vulnerabilities for mass infections, a tactic exemplified by groups like ClOp.[63] While the core ransomware model remains effective and largely unchanged, the creation and deployment process is becoming increasingly automated and adaptive, with accelerated vulnerability discovery enabling deeper, more precise payload delivery across target environments.

Ransomware tactics will shift focus from pure data encryption to data exploitation and intelligence-driven extortion. AI's ability to analyze non-text media, such as images, voice, and video, allows attackers to identify a victim's most sensitive assets and apply targeted pressure. As financial and data protection regulations strengthen, breaches involving confidential or proprietary information, particularly enterprise knowledge bases, will carry greater impact, making data theft and exposure more valuable than traditional encryption-based disruption.

Ransomware campaigns will become more seamless, blending into legitimate activity and exploiting the complexity of enterprise supply chains and digital ecosystems. As organizations accelerate their digital transformation and AI adoption, attackers will adapt accordingly, integrating these same technologies into their attacks for greater precision, persistence, and reach.

In 2026, ransomware groups will increasingly make use of agentic AI to manage large portions of the attack lifecycle without human oversight. As previously mentioned, these AI-powered tools will be capable of scanning for vulnerabilities, adapting attack methods in real time, and executing complete ransomware operations end to end. What began as AI-assisted malware development in 2025, as seen in attacks involving FunkSec,[64] will mature into operationally independent malware such as LameHug,[65] marking a major leap in automation.

The continued rise of AI-powered ransomware-as-a-service (RaaS)[66] will allow even inexperienced operators to conduct complex attacks with minimal skill, reducing reliance on traditional RaaS affiliates and making independent ransomware operations increasingly more common. We predict that this democratization of offensive capability will greatly expand the threat landscape.

Ransomware groups will escalate coercion tactics beyond traditional data theft and encryption. AI-driven extortion bots will engage victims directly in ransom negotiations. Some groups, such as the Global Group Ransomware syndicate,[67] have already begun experimenting with these automated negotiation agents. Other expected tactics will exert reputational and regulatory pressure: contacting regulators, leaking manipulated media, or fabricating scandals to coerce victims into paying.

Supply chain attacks will remain one of the most efficient ransomware delivery mechanisms in 2026.[68] As businesses integrate AI models, agents, and open-source tools into daily workflows, these assets will become lucrative attack vectors. Supply chain compromises are a key entry point for ransomware as attackers exploit the open-source components organizations depend on. By tampering with popular packages, development pipelines, or container images, threat actors can silently deliver malicious updates at scale.

Modern ransomware operators are increasingly mirroring legitimate enterprises. These threat actors make use of cloud infrastructure, rent computing resources, and form shell companies to disguise operations. Constant rebranding, driven by law enforcement crackdowns and sanctions, helps them avoid attribution, reset their reputation, and recruit new affiliates. This agility forces defenders to shift focus from tracking specific ransomware to identifying persistent tactics, techniques, and procedures that transcend individual group identities.

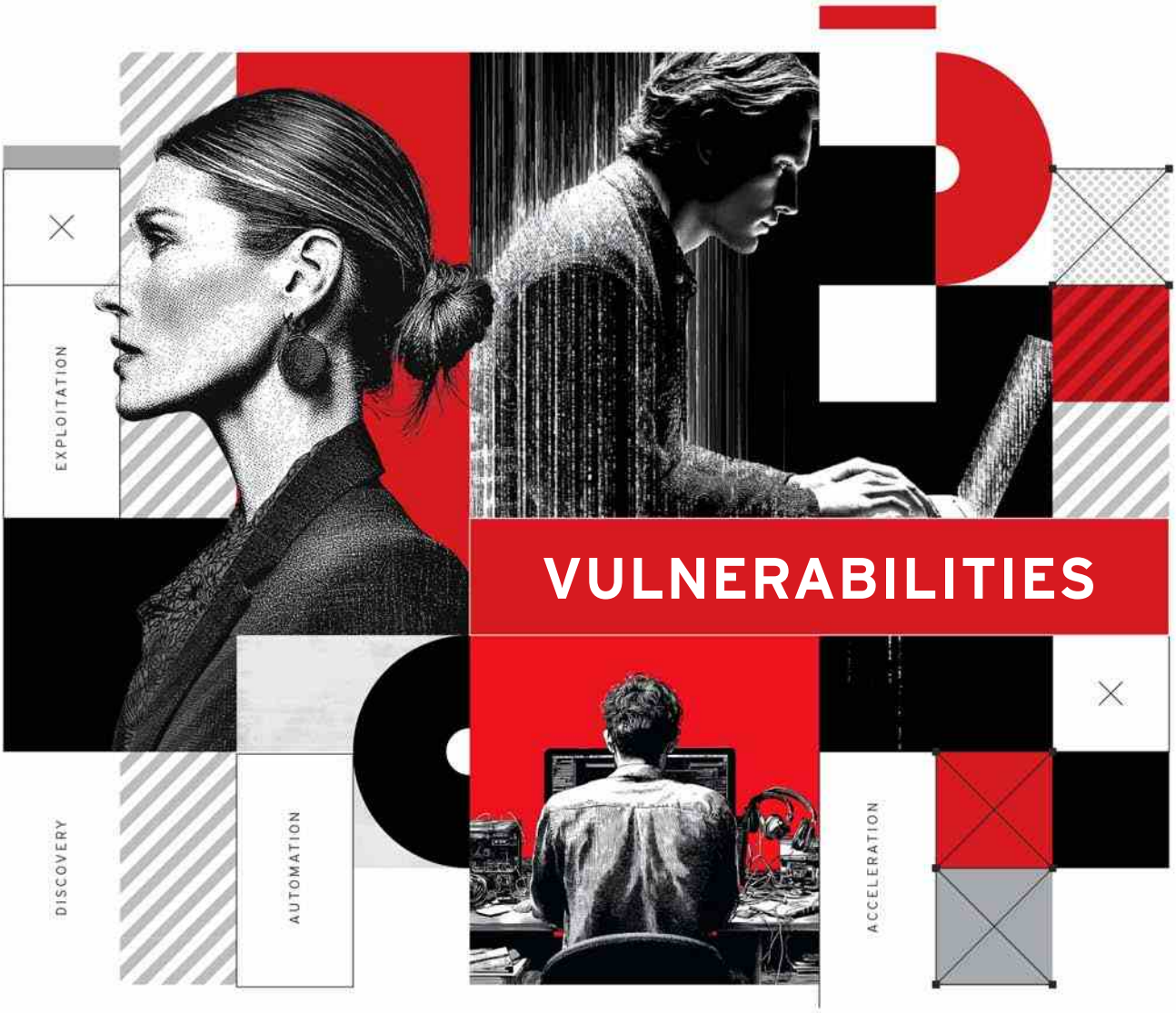
For Cybersecurity Professionals

Ransomware in 2026 will be faster, smarter, and more adaptive. Agentic AI will automate every stage of the attack lifecycle, from vulnerability discovery to ransom negotiation, leaving defenders little response time. Meanwhile, attackers will use automated data analysis to identify possible leverage points, extending pressure long after initial breaches. Constant rebranding and tool reuse will blur attribution, forcing defenders to focus on behavioral detection over static indicators.

For Decision-Makers

We will see ransomware operators weaponizing corporate knowledge bases, AI models, and supply chain dependencies, turning business assets into attack surfaces. Attackers will use automation to identify high-impact targets, exploit interdependencies, and apply tailored pressure on organizations, while turning everyday business tools and partnerships into entry points, exploiting the same networks and integrations companies rely on to operate. Even well-secured organizations could face major disruptions as attackers use stolen information, leaked data, and misinformation to damage trust and reputation.

While zero trust frameworks are important for strengthening security against threats such as ransomware, they must be supported by broader resilience measures under a comprehensive Cyber Risk and Exposure Management (CREM)[69] approach. This includes regular awareness training to counter increasingly convincing AI-assisted social engineering, and the use of threat intelligence platforms to track known ransomware threats and assist in developing effective defensive strategies. At the same time, fundamentals such as comprehensive business impact analyses, offline and immutable backups, tested recovery playbooks, and fallback workflows remain essential to maintaining continuity even when core systems are disrupted.



AI-Driven Exploits, Zero-Day Acceleration, and Supply Chain Risks

Consistent with the broader themes of this report, the evolution of zero-day vulnerabilities in 2026 will be influenced by the integration of AI into both the discovery and exploitation processes. As organizations increasingly integrate AI into their workflows, new vulnerabilities will emerge, such as prompt injection flaws and bugs in AI-enabled environments (as seen in the CurXecute vulnerability in Cursor IDE),[70] offering fresh entry points for exploitation.

The use of AI-driven cyber reasoning systems (CRSs) will further accelerate vulnerability identification in open-source and enterprise environments, enhancing both the speed and scope of zero-day exploitation. Moreover, AI-enhanced reconnaissance capabilities will empower threat actors to conduct more sophisticated and efficient information-gathering operations, employing both network-level intelligence and OSINT to refine targeting and attack planning. If an LLM's coding style contains a flaw, every codebase it produces could share that vulnerability, allowing AI-driven tools to exploit it instantly while defenders take time to patch and redeploy.[71] Again, vibe coding plays a factor in the vulnerability landscape, with some tests showing that approximately 45% of AI-generated code introduces security bugs.[72]

We expect threat actors to use AI to generate and refine SQL and command injection attacks. While the associated vulnerabilities are well known, automation will enable faster, more precise, and large-scale exploitation across diverse software and services where such flaws persist. Additionally, vulnerabilities in the expanding AI ecosystem, including inference servers, MCP servers, and AI frameworks, will introduce new attack surfaces.

Microarchitectural flaws will continue to pose challenges in securing next-generation chipsets, including GPUs used for large-scale AI inference. Meanwhile, edge appliances such as VPNs and security systems, which have been heavily exploited in 2025, will remain high-value targets.

Approaching 2026, organizations continue to underestimate certain blind spots that attackers are likely to exploit. One major concern is the persistence of unpatched IoT and operational technology (OT) devices,[73] which provide attackers with reliable footholds for lateral movement.

Supply chain vulnerabilities remain a critical enterprise risk. Open-source dependencies and third-party code libraries are under increasing scrutiny from threat actors, who recognize that compromising these repositories can result in widespread downstream exploitation, similar to the Log4j incident.[74]

On the regulatory side of things, governments are beginning to enforce patching deadlines based on severity scores,[75] but these do not always align with real-world exploitability — a high CVSS score doesn't necessarily warrant an emergency patch response. This misalignment might strain enterprise resources and lead to misplaced priorities.

For Cybersecurity Professionals

The vulnerability landscape in 2026 will be defined by speed and complexity. AI is accelerating both the discovery and exploitation of weaknesses. Zero-days will emerge faster than patching cycles can keep up, and automation will make exploitation more targeted and large-scale. Persistent issues such as unpatched IoT and OT devices, as well as misconfigurations in AI-enabled environments, will continue to provide attackers with entry points for lateral movement and persistence. For defenders, visibility and prioritization will become the hardest challenges as the number of exploitable vulnerabilities grow.

For Decision-Makers

As organizations deepen their reliance on AI, the IoT, and open-source software, the risks for businesses increase accordingly. A single compromised AI model, library, or third-party dependency could cause harm across supply chains and customer ecosystems, amplifying both operational and reputational damage. Managing vulnerabilities is no longer just an IT problem; it now includes the AI systems and workflows that support everyday business. The challenge is to keep pushing innovation forward while making sure these new technologies don't become the weakest link in the organization.

To strengthen vulnerability management strategies in 2026, we recommend reframing vulnerability management through the CREM model,[76] prioritizing overall risk reduction rather than simply patching flaws. Enterprises should dynamically prioritize remediation based on not only vulnerability severity but also asset criticality, exposure, and exploit likelihood. Beyond simply accelerating patching, organizations need to factor in misconfigurations and access control gaps when measuring their overall exposure. This approach provides a clearer, data-driven understanding of what vulnerabilities pose the greatest business risk, helping leaders make smarter, organization-wide decisions to reduce exposure and build lasting resilience.

Conclusion

Cyberthreats will continue to evolve in speed, complexity, and effectiveness for the foreseeable future, with the threat actors behind them becoming more organized, industrialized, and resourceful — employing automation, shared infrastructure, and social engineering to maximize reach and impact.

AI is increasingly becoming central to both enterprise operations and the threat landscape as a whole. Threat actors are using AI to automate attacks, craft highly convincing social engineering schemes, and exploit vulnerabilities at unprecedented speed and scale. They are able to coordinate complex campaigns with minimal effort, lowering the barrier for executing high-impact cyber operations.

The modern attack surface is expanding rapidly across hybrid and multi-cloud environments, remote endpoints, AI infrastructure, edge devices, bespoke code, and complex digital supply chains. Traditional perimeter defenses are no longer sufficient. Instead, security must be embedded across every layer of technology and process. Implementing a risk-based approach, one that emphasizes asset discovery, exposure management, and context-driven prioritization, enables enterprises to devote resources to the vulnerabilities and risks that matter most.

Businesses cannot afford to focus merely on emerging threats while neglecting foundational security practices. The balance between innovation and security will remain critical. Organizations that prioritize visibility, patch management, identity security, and governance, while also investing in threat intelligence and operational resilience, will be the ones that are best positioned to withstand both old and novel forms of attack. The goal is not just to detect or respond, but also to anticipate and minimize impact through layered defense and continuous improvement.

As threat actors exploit growing interconnectivity, organizations should ensure that teams, tools, and data work together. Continuous monitoring, proactive threat hunting, and well-practiced incident response plans will remain essential components of a resilient security posture well into the future. Training and awareness programs must also evolve, empowering employees to recognize and respond to modern threats that use emerging forms of technologies.

Looking ahead to 2026 and beyond, effective cybersecurity will be measured by an organization's resilience against ever-changing threats. Attackers might shift tactics, but their objectives — disruption, theft, and control — remain constant. Defenders must prioritize building systems and teams that can adapt, recover, and strengthen after each incident. Security is no longer a static target, but a continuous, evolving factor that needs to be aligned with the threat landscape.

References

1 Jean-Pierre Joosting. (July 30, 2025). *EENewsEurope.* "Report Finds AI-Generated Code Poses Security Risks." Accessed on Oct. 29, 2025, at https://www.eenewseurope.com/en/report-finds-ai-generated-code-poses-security-risks/.

2 Sean Park. (June 5, 2025). *Trend Micro.* "Slopsquatting: When AI Agents Hallucinate Malicious Packages." Accessed on Oct. 30, 2025, at https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/slopsquatting-when-ai-agents-hallucinate-malicious-packages.

3 Salvatore Gariuolo and Vincenzo Ciancaglini. (June 18, 2025). Trend Micro. "The Road to Agentic AI: Defining a New Paradigm for Technology and Cybersecurity." Accessed on Oct. 30, 2025, at https://www.trendmicro.com/vinfo/us/security/news/security-technology/the-road-to-agentic-ai-defining-a-new-paradigm-for-technology-and-cybersecurity.

4 Joanna England. (Aug. 11, 2025). *Intelligent CISO.* "AI's Silent Threat: Navigating the Risks of Autonomous Agents." Accessed on Oct. 30, 2025, at https://www.intelligentciso.com/2025/08/11/ais-silent-threat-navigating-the-risks-of-autonomous-agents.

5 Salvatore Gariuolo and Vincenzo Ciancaglini. (June 18, 2025). *Trend Micro.* "The Road to Agentic AI: Defining a New Paradigm for Technology and Cybersecurity." Accessed on Oct. 30, 2025, at https://www.trendmicro.com/vinfo/us/security/news/security-technology/the-road-to-agentic-ai-defining-a-new-paradigm-for-technology-and-cybersecurity.

6 Nnenna Ndukwe. (Aug. 26, 2025). *Qodo.* "The Rise of Agentic Workflows in Enterprise AI Development." Accessed on Oct. 30, 2025, at https://www.qodo.ai/blog/agentic-workflows-in-ai-development/.

7 eSynergy. (n.d.). *eSynergy.* "The rise of Agentic AI Workflows: A game-changer in digital efficiency." Accessed on Oct. 30, 2025, at https://www.esynergy.co.uk/blogs/rise-of-agentic-ai-workflows/.

8 Dorian Schultz. (April 1, 2025). *SplxAI.* "Exploiting Agentic Workflows: Prompt Injections in Multi-Agent AI Systems." Accessed on Oct. 30, 2025, at https://splx.ai/blog/exploiting-agentic-workflows-prompt-injections-in-multi-agent-ai-systems.

9 Umesh Sachdev. (July 31, 2025). *World Economic Forum.* "Enterprise AI is at a tipping Point, here's what comes next." Accessed on Oct. 30, 2025, at https://www.weforum.org/stories/2025/07/enterprise-ai-tipping-point-what-comes-next/.

10 AGI Research. (2025). *GitHub.* "AGI Research Organization – GitHub Repository." Accessed on Oct. 30, 2025, at https://github.com/agiresearch.

11 Archana Vaidheeswaran. (Feb. 20, 2025). *Tribe AI.* "The Agentic AI Future: Understanding AI Agents, Swarm Intelligence, and Multi-Agent Systems." Accessed on Oct. 30, 2025, at https://www.tribe.ai/applied-ai/the-agentic-ai-future-understanding-ai-agents-swarm-intelligence-and-multi-agent-systems/.

12 Barbara. (Feb. 14, 2025). *IoT For All.* "Edge AI in 2025: Bold Predictions and a Reality Check." Accessed on Oct. 30, 2025, at https://www.iotforall.com/edge-ai-2025-predictions-reality-check.

13 Numaan Huq and Roel Reyes. (Oct. 10, 2025). *Trend Micro.* "Agentic Edge AI: Autonomous Intelligence on the Edge." Accessed on Oct. 30, 2025, at https://www.trendmicro.com/vinfo/gb/security/news/cybercrime-and-digital-threats/agentic-edge-ai-autonomous-intelligence-on-the-edge.

14 Dr. Jagreet Kaur. (April 22, 2025). *XenonStack.* "Edge AI with Agentic AI for Distributed Intelligence." Accessed on Oct. 30, 2025, at https://www.xenonstack.com/blog/edge-ai-autonomous-systems.

15 Michael Hendricks. (Sept. 26, 2025). *Rippling.* "Agentic AI Security: Complete Guide to Threats, Risks & Best Practices 2025." Accessed on Oct. 30, 2025, at https://www.rippling.com/blog/agentic-ai-security.

16 McKinsey. (Oct. 16, 2025). *McKinsey.* "Deploying Agentic AI with Safety and Security: A Playbook for Technology Leaders." Accessed on Oct. 30, 2025, at https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/deploying-agentic-ai-with-safety-and-security-a-playbook-for-technology-leaders.

17 McKinsey. (Oct. 16, 2025). *McKinsey.* "Deploying Agentic AI with Safety and Security: A Playbook for Technology Leaders." Accessed on Oct. 30, 2025, at https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/deploying-agentic-ai-with-safety-and-security-a-playbook-for-technology-leaders.

18 Lumenova AI. (Sept. 19, 2025). *Lumenova AI.* "AI Agents: Potential Risks." Accessed on Oct. 30, 2025, at https://www.lumenova.ai/blog/ai-agents-potential-risks/.

19 Dan Xu et al. (2025). *MDPI*. "The Erosion of Cybersecurity Zero-Trust Principles Through Generative AI: A Survey on the Challenges and Future Directions." Accessed on Oct. 30, 2025, at https://www.mdpi.com/2624-800X/5/4/87.

20 Ravie Lakshmanan. (Sept. 29, 2025). *The Hacker News*. "First Malicious MCP Server Found Stealing Emails in Rogue Postmark-MCP Package." Accessed on Oct. 30, 2025, at https://thehackernews.com/2025/09/first-malicious-mcp-server-found.html.

21 David Sancho, Salvatore Gariuolo, and Vincenzo Ciancaglini. (July 9, 2025). *Trend Micro*. "Deepfake It Till You Make It: A Comprehensive View of the New AI Criminal Toolset." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/deepfake-it-til-you-make-it-a-comprehensive-view-of-the-new-ai-criminal-toolset.

22 David Cohen. (March 4, 2025). *JFrog*. "JFrog and Hugging Face Join Forces to Expose Malicious ML Models." Accessed on Oct. 30, 2025, at https://jfrog.com/blog/jfrog-and-hugging-face-join-forces/.

23 Daniel Lunghi and Leon M Chang. (Oct. 22, 2025). *Trend Micro*. "The Rise of Collaborative Tactics Among China-aligned Cyber Espionage Campaigns." Accessed on Oct. 29, 2025, at https://www.trendmicro.com/en_us/research/25/j/premier-pass-as-a-service.html.

24 Bayu Anggorojati, Arif Perdana, and Derry Wijaya. (July 24, 2024). *Asia Times*. "FraudGPT, WormGPT and the rise of dark LLMs." Accessed on Oct. 23, 2025, at https://asiatimes.com/2024/07/fraudgpt-wormgpt-and-the-rise-of-dark-llms/.

25 Merav Bar and Rami McCarthy. (Aug. 27, 2025). *Wiz*. "s1ngularity: supply chain attack leaks secrets on GitHub: everything you need to know." Accessed on Oct. 30, 2025, at https://www.wiz.io/blog/s1ngularity-supply-chain-attack.

26 Federal Bureau of Investigation et al. (July 9, 2024). "State-Sponsored Russian Media Leverages Meliorator Software for Foreign Malign Influence Activity." Accessed on Oct. 30, 2025, at https://www.ic3.gov/CSA/2024/240709.pdf.

27 OpenAI. (June 2025). *OpenAI*. "Disrupting Malicious Uses of AI: June 2025." Accessed on Oct. 30, 2025, at https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-ai-june-2025.pdf.

28 Nick Dai and Pierre Lee. (Aug. 28, 2025). *Trend Micro*. "TAOTH Campaign Exploits End-of-Support Software to Target Traditional Chinese Users and Dissidents." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/en_us/research/25/h/taoth-campaign.html.

29 Mercedes Cardona. (Aug. 21, 2025). *Dark Reading*. "Fake Employees Pose Real Security Risks." Accessed on Oct. 30, 2025, at https://www.darkreading.com/cyberattacks-data-breaches/fake-employees-pose-real-security-risks.

30 Stephen Hilt and Mayra Rosario Fuentes. (Jan. 7, 2025). *Trend Micro*. "Bridging Divides, Transcending Borders: The Current State of the English Underground." Accessed on Nov. 6, 2025, at https://documents.trendmicro.com/assets/white_papers/wp-bridging-divides-transcending-borders.pdf.

31 Dustin Moody et al. (Nov. 12, 2024). *National Institute of Standards and Technology (NIST)*. "Transition to Post-Quantum Cryptography Standards (NIST IR 8547)." Accessed on Oct. 23, 2025, at https://csrc.nist.gov/pubs/ir/8547/ipd.

32 Matthew Kosinski and Stephanie Carruthers. (n.d.). *IBM Think*. "With generative AI, social engineering gets more dangerous—and harder to spot." Accessed on Oct. 23, 2025, at https://www.ibm.com/think/insights/generative-ai-social-engineering.

33 Dan Kaplan. (Aug. 29, 2025). *Security Boulevard*. "How AI Agents Are Creating a New Class of Identity Risk." Accessed on Oct. 30, 2025, at https://securityboulevard.com/2025/08/how-ai-agents-are-creating-a-new-class-of-identity-risk.

34 Eric Olden. (Sept. 10, 2025). *Strata*. "A New Identity Playbook for AI Agents: Securing the Agentic User Flow." Accessed on Oct. 30, 2025, at https://www.strata.io/blog/agentic-identity/new-identity-playbook-ai-agents-not-nhi-8b/.

35 Rabia Noureen. (Aug. 18, 2025). *Petri*. "AI Agents Expose a New Identity Security Gap, Okta Warns." Accessed on Oct. 30, 2025, at https://petri.com/ai-agents-identity-security-gap/.

36 MITRE Corporation. (n.d.). *MITRE*. "Adversary in the Browser (AiTB) – CAPEC-662." Accessed on Oct. 23, 2025, at https://capec.mitre.org/data/definitions/662.html.

37 Maristel Policarpio et al. (Aug. 14, 2025). *Trend Micro*. "Crypto24 Ransomware Group Blends Legitimate Tools with Custom Malware for Stealth Attacks." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/en_us/research/25/h/crypto24-ransomware-stealth-attacks.html.

38 Vincenzo Ciancaglini et al. (Dec. 6, 2024). *Trend Micro*. "AI Assistants in the Future: Security Concerns and Risk Management." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/vinfo/us/security/news/security-technology/looking-into-the-future-risks-and-security-considerations-to-ai-digital-assistants.

39 Cybersecurity and Infrastructure Security Agency (CISA). (Dec. 17, 2024). *CISA*. "BOD 25-01: Implementing Secure Practices for Cloud Services." Accessed on Oct. 23, 2025, at https://www.cisa.gov/news-events/directives/bod-25-01-implementing-secure-practices-cloud-services.

40 Arjun Trivedi. (Nov. 30, 2023). *Microsoft*. "Identifying Adversary-in-the-Middle (AiTM) Phishing Attacks through 3rd-Party Network Detection." Accessed on Oct. 23, 2025, at https://techcommunity.microsoft.com/blog/microsoftsentinelblog/identifying-adversary-in-the-middle-aitm-phishing-attacks-through-3rd-party-netw/3991358.

41 Shweta Sharma. (Aug. 1, 2025). *CSO Online*. "Cybercrooks faked Microsoft OAuth apps for MFA phishing." Accessed on Oct. 23, 2025, at https://www.csoonline.com/article/4032743/cybercrooks-faked-microsoft-oauth-apps-for-mfa-phishing.html.

42 Charlie Gardner, Steven Adair, and Tom Lancaster. (Feb. 13, 2025). *Volexity*. "Multiple Russian Threat Actors Targeting Microsoft Device Code Authentication." Accessed on Oct. 23, 2025, at https://www.volexity.com/blog/2025/02/13/multiple-russian-threat-actors-targeting-microsoft-device-code-authentication/.

43 Google Threat Intelligence Group. (June 4, 2025). *Google Cloud Blog*. "The Cost of a Call: From Voice Phishing to Data Extortion." Accessed on Oct. 23, 2025, at https://cloud.google.com/blog/topics/threat-intelligence/voice-phishing-data-extortion.

44 Quentin Bourgue, Grégoire Clermont, and Sekoia TDR. (Jan. 16, 2025). *Sekoia.io*. "Sneaky 2FA: exposing a new AiTM Phishing-as-a-Service." Accessed on Oct. 23, 2025, at https://blog.sekoia.io/sneaky-2fa-exposing-a-new-aitm-phishing-as-a-service/.

45 Trend Micro. (Dec. 16, 2024). *Trend Micro*. "The Artificial Future – Trend Micro Security Predictions for 2025." Accessed on Oct. 23, 2025, at https://documents.trendmicro.com/assets/white_papers/The-Artificial-Future-Trend-Micro-Security-Predictions-for-2025.pdf.

46 Jeffrey Francis Bonaobra and Joshua Aquino. (Sept. 18, 2025). *Trend Micro*. "What We Know About the NPM Supply Chain Attack." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/en_us/research/25/i/npm-supply-chain-attack.html.

47 Cybersecurity and Infrastructure Security Agency (CISA). (Sept. 23, 2025). *CISA*. "Widespread Supply Chain Compromise Impacting npm Ecosystem." Accessed on Oct. 23, 2025, at https://www.cisa.gov/news-events/alerts/2025/09/23/widespread-supply-chain-compromise-impacting-npm-ecosystem.

48 Cloud Security Alliance. (April 28, 2025). *Cloud Security Alliance*. "Top Threats to Cloud Computing – Deep Dive 2025." Accessed on Oct. 23, 2025, at https://cloudsecurityalliance.org/artifacts/top-threats-to-cloud-computing-2025.

49 Alfredo Oliveira. (Aug. 6, 2024). *Trend Micro*. "Today's Cloud and Container Misconfigurations Are Tomorrow's Critical Vulnerabilities." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/vinfo/us/security/news/virtualization-and-cloud/todays-cloud-and-container-misconfigurations-are-tomorrows-critical-vulnerabilities.

50 Dustin Childs. (Jan. 16, 2024). *Trend Micro Zero Day Initiative*. "Pwn2Own Vancouver 2024: Bring Cloud-Native/Container Security to Pwn2Own." Accessed on Oct. 23, 2025, at https://www.zerodayinitiative.com/blog/2024/1/16/pwn2own-vancouver-2024-bring-cloud-nativecontainer-security-to-pwn2own.

51 Alfredo Oliveira. (Aug. 6, 2024). *Trend Micro*. "Today's Cloud and Container Misconfigurations Are Tomorrow's Critical Vulnerabilities." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/vinfo/us/security/news/virtualization-and-cloud/todays-cloud-and-container-misconfigurations-are-tomorrows-critical-vulnerabilities.

52 Alfredo Oliveira and David Fiser. (Sept. 23, 2025). *Trend Micro*. "AI-Powered App Exposes User Data, Creates Risk of Supply Chain Attacks". Accessed on Oct. 23, 2025, at https://www.trendmicro.com/en_us/research/25/i/ai-powered-app-exposes-user-data.html.

53 Nitesh Surana. (July 11, 2024). *Trend Micro*. "Leaky Labels: Bypassing Traefik Proxy Leveraging cAdvisor Metrics." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/vinfo/us/security/news/virtualization-and-cloud/leaky-labels-bypassing-traefik-proxy-leveraging-cadvisor-metrics.

54 Nitesh Surana. (May 2, 2024). *Trend Micro*. "Observability Exposed: Exploring Risks in Cloud-Native Metrics." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/vinfo/gb/security/news/virtualization-and-cloud/observability-exposed-exploring-risks-in-cloud-native-metrics.

55 Alfredo de Oliveira and David Fiser. (Dec. 4, 2024). *Trend Micro*. "Silent Sabotage: Weaponizing AI Models in Exposed Containers." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/silent-sabotage-weaponizing-ai-models-in-exposed-containers.

56 Numaan Huq et al. (May 23, 2024). *Trend Micro*. "Navigating the Threat Landscape for Cloud-Based GPUs." Accessed on Oct. 30, 2025, at https://www.trendmicro.com/vinfo/us/security/news/threat-landscape/navigating-the-threat-landscape-for-cloud-based-gpus.

57 Yicheng Zhang et al. (March 22, 2025). *arXiv*. "NVBleed: Covert and Side-Channel Attacks on NVIDIA Multi-GPU Interconnect." Accessed on Oct. 30, 2025, at https://arxiv.org/pdf/2503.17847.

58 Tyler Sorensen and Heidy Khlaaf. (Jan. 16, 2024). *Trail of Bits*. "LeftoverLocals: Listening to LLM Responses Through Leaked GPU Local Memory." Accessed on Oct. 30, 2025, at https://blog.trailofbits.com/2024/01/16/leftoverlocals-listening-to-llm-responses-through-leaked-gpu-local-memory/.

59 Shir Tamari, Ronen Shustin, and Andres Riancho. (Feb. 11, 2025). *Wiz*. "How Wiz Found a Critical NVIDIA AI Vulnerability: Deep Dive into a Container Escape (CVE-2024-0132)." Accessed on Oct. 30, 2025, at https://www.wiz.io/blog/nvidia-ai-vulnerability-deep-dive-cve-2024-0132.

60 Vladimir Kropotov et al. (Feb. 23, 2023). *Trend Micro*. "Understanding Ransomware Using Data Science." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/understanding-ransomware-using-data-science.

61 FBI Support Cyber Knowledge Base. (n.d.). *FBI Support Cyber Knowledge Base*. "What is the Potential for AI to Automate Vulnerability Discovery and Exploitation?" Accessed on Oct. 23, 2025, at https://fbisupport.com/potential-ai-automate-vulnerability-discovery-exploitation.

62 Anas AlMajali et al. (Oct. 13, 2024). *Multidisciplinary Digital Publishing Institute (MDPI)*. "Automated Vulnerability Exploitation Using Deep Reinforcement Learning." Accessed on Oct. 23, 2025, at https://www.mdpi.com/2076-3417/14/20/9331.

63 Janus Agcaoili et al. (June 15, 2021). *Trend Micro*. "Ransomware Double Extortion and Beyond: REvil, Clop, and Conti." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/ransomware-double-extortion-and-beyond-revil-clop-and-conti.

64 Ravie Lakshmanan. (Jan. 10, 2025). *The Hacker News*. "AI-Driven Ransomware FunkSec Targets 85 Victims Using Double Extortion Tactics." Accessed on Oct. 23, 2025, at https://thehackernews.com/2025/01/ai-driven-ransomware-funksec-targets-85.html.

65 Bill Toulas. (July 17, 2025). *BleepingComputer*. "LameHug malware uses AI LLM to craft Windows data-theft commands in real-time." Accessed on Oct. 30, 2025, at https://www.bleepingcomputer.com/news/security/lamehug-malware-uses-ai-llm-to-craft-windows-data-theft-commands-in-real-time/.

66 Alan Willie. (February 2025). "The Evolution of Ransomware-as-a-Service (RaaS): AI's Role in Cybercrime and Countermeasures." Accessed on Oct. 30, 2025, at https://www.researchgate.net/publication/388928559_The_Evolution_of_Ransomware-as-a-Service_RaaS_AI%27s_Role_in_Cybercrime_and_Countermeasures.

67 Mandvi. (July 22, 2025). *Cyber Press*. "GLOBAL GROUP's Golang Ransomware Expands Attacks Across Major Operating Systems." Accessed on Oct. 23, 2025, at https://cyberpress.org/global-groups-golang-ransomware-expands-attacks/.

68 Trend Micro. (Sept. 20, 2022). *Trend Micro*. "The Risk of Ransomware Supply Chain Attacks." Accessed on Oct. 23, 2025, at https://www.trendmicro.com/en_us/research/22/i/ransomware-supply-chain-attack-stats.html.

69 Trend Micro. (n.d.). *Trend Micro*. "Cyber Risk Exposure Management." Accessed on Oct. 24, 2025, at https://www.trendmicro.com/en_us/business/products/cyber-risk-exposure-management.html.

70 Chetan Seripally. (Aug. 4, 2025). *Medium*. "CurXecute Vulnerability in Cursor IDE: A Wake-Up Call for Developer Security." Accessed on Oct. 24, 2025, at https://medium.com/@seripallychetan/curxecute-vulnerability-in-cursor-ide-a-wake-up-call-for-developer-security-25b607ba960e.

71 Shivani Shukla, Himanshu Joshi, and Romilla Syed. (Sept. 26, 2025). "Security Degradation in Iterative AI Code Generation: A Systematic Analysis of the Paradox." Accessed on Oct. 24, 2025, at https://arxiv.org/pdf/2506.11022.

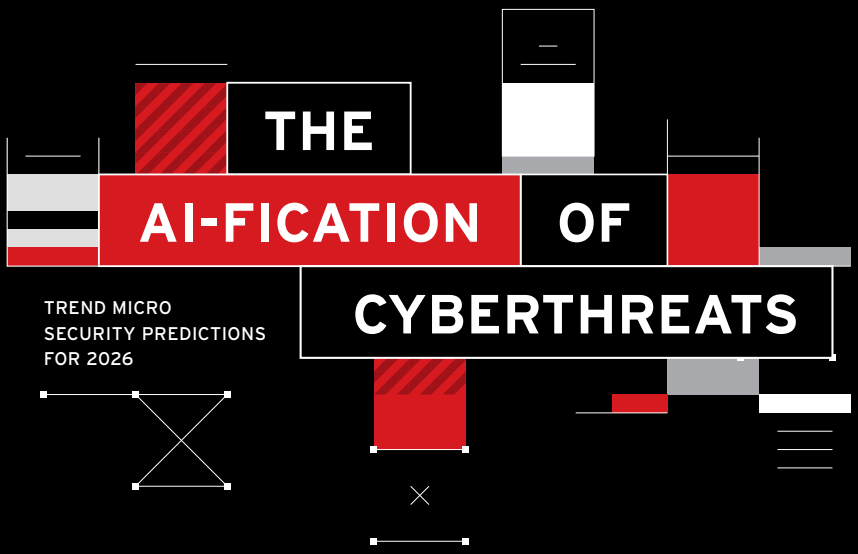
72 Jean-Pierre Joosting. (July 30, 2025). *EE News Europe*. "Report Finds AI-Generated Code Poses Security Risks." Accessed on Oct. 24, 2025, at https://www.eenewseurope.com/en/report-finds-ai-generated-code-poses-security-risks/.

73 Trend Micro. (Jan. 17, 2025). *Trend Micro*. "IoT Botnet Linked to Large-Scale DDoS Attacks Since the End of 2024." Accessed on Oct. 24, 2025, at https://www.trendmicro.com/en_us/research/25/a/iot-botnet-linked-to-ddos-attacks.html.

74 Trend Micro. (n.d.). *Trend Micro*. "Log4j (Log4Shell) Vulnerability - What To Know." Accessed on Oct. 24, 2025, at https://www.trendmicro.com/en_us/apache-log4j-vulnerability.html.

75 Gerbrand ten Napel, Michel van Eeten, and Simon Parkin. (2025). *IEEE*. "Speedrunning the Maze: Meeting Regulatory Patching Deadlines in a Large Enterprise Environment." Accessed on Oct. 30, 2025, at https://www.computer.org/csdl/proceedings-article/sp/2025/223600a081/21B7RnD4zWU.

76 Trend Micro. (n.d.). *Trend Micro*. "Cyber Risk Exposure Management." Accessed on Oct. 24, 2025, at https://www.trendmicro.com/en_us/business/products/cyber-risk-exposure-management.html.



TREND MICRO™ | Proactive Security Starts Here

Trend Micro, a global cybersecurity leader, helps make the world safe for exchanging digital information between people, governments, and enterprises.

Trend leverages security expertise and AI to protect more than 500,000 enterprises and millions of individuals across clouds, networks, endpoints, and devices worldwide.

At the core is Trend Vision One™, the only AI-powered enterprise cybersecurity platform that centralizes cyber risk exposure management and security operations, delivering layered protection across on-premises, hybrid, and multi-cloud environments.

The unmatched threat intelligence delivered by Trend empowers organizations to proactively defend against hundreds of millions of threats every day.

Proactive security starts here. TrendMicro.com

Trend Research