CSAI Foundation

Cloud Security Alliance AI Safety Initiative

The Collapsing Exploit Window

How AI Is Compressing Mean-Time-to-Exploit Across Enterprise Infrastructure

Unofficial AI-assisted Research

2026-04-11

Table of Contents

Executive Summary

For more than a decade, the security industry operated on an implicit assumption: that organizations possessed a grace period between the public disclosure of a vulnerability and the moment adversaries could realistically exploit it. That assumption is no longer sound. Artificial intelligence has fundamentally altered the economics and timelines of exploit development, and the empirical record of 2025 documents the consequences with uncomfortable precision.

While organizations historically took a median of 32 days to apply patches to known vulnerabilities—a window that once roughly corresponded to the time available before exploitation began—that window has collapsed to approximately 5 days for median time-to-exploit in 2025. [1][3][10] Rapid7's 2026 Global Threat Landscape Report found a 105 percent year-over-year increase in exploited high and critical severity vulnerabilities—146 CVEs actively exploited in 2025 compared to 71 in 2024—while the median time from vulnerability disclosure to inclusion in CISA's Known Exploited Vulnerabilities catalog compressed from 8.5 days to 5.0 days. [1] Rapid7's 2026 analysis also found that a significant and growing fraction of Known Exploited Vulnerabilities showed exploitation preceding or coinciding with CVE publication, a pattern indicating that some campaigns were active before formal vulnerability identifiers were assigned. [1] Against this backdrop, the Verizon 2025 Data Breach Investigations Report found that only 54 percent of vulnerable devices were fully remediated within the year. [3] The defender gap—the distance between how fast attackers weaponize and how fast organizations remediate—has become a structural feature of the threat landscape rather than a temporary operational shortfall.

Artificial intelligence is the principal accelerant. Research by the University of Illinois Urbana-Champaign demonstrated that GPT-4, operating as an autonomous agent, could exploit 87 percent of tested one-day vulnerabilities when provided with CVE descriptions—a success rate that no other model or automated scanner achieved in the same evaluation. [5] Independent research has shown AI-enabled frameworks generating functional proof-of-concept exploit code in 10 to 15 minutes at costs approaching one dollar per exploit. [20] Automated patch diffing tools, augmented by large language models, now allow adversaries to analyze binary changes, isolate the patched code path, and construct working exploits within hours of a vendor's patch release—before most enterprise patch testing cycles have begun. [9]

This whitepaper examines the mechanisms driving MTTE compression, documents the structural defender gap it creates, and offers organizations a framework for response that acknowledges the new operational reality. The era of the forgiving internet—in which slow patching reliably carried manageable risk—is over for critical, externally facing infrastructure, and narrowing for enterprise environments broadly. [19]

Introduction and Background

The Historical Model and Its Failure

The concept of a patch window has shaped enterprise vulnerability management for nearly two decades. When a vendor disclosed a vulnerability and released a corresponding patch, security teams could reasonably expect days to weeks before that vulnerability appeared in active exploit chains targeting production infrastructure. This window reflected the genuine effort required to transform a disclosed vulnerability into a reliable exploit: understanding the flaw's mechanics, developing proof-of-concept code, testing it against realistic target configurations, and packaging it for distribution. Nation-states and sophisticated criminal actors could compress this timeline, but the mass-market exploitation ecosystem typically required time.

Enterprise security programs were designed around this assumption. Patch management policies based on thirty-day or sixty-day remediation targets for critical and high-severity vulnerabilities were built not on complacency but on a reasonable estimate of attacker timelines. Organizations invested in testing cycles to verify that patches did not break production systems, accepting the measured risk that the delay created. Security teams prioritized based on CVSS severity scores and threat intelligence feeds that operated on human timescales.

That model has encountered a discontinuity. The combination of AI-assisted exploit research, automated binary analysis, and agentic exploitation frameworks has not merely shortened the window—in some cases it has eliminated it entirely. The CISA KEV data for 2025 shows that exploitation of newly disclosed vulnerabilities is no longer unusual before the patch has been widely deployed, and in a significant fraction of cases, exploitation precedes even the formal CVE publication. [6][7] Understanding how this happened—and what it means for enterprise security architecture—is the purpose of this paper.

Scope and Methodology

This analysis synthesizes publicly available threat intelligence from industry sources, academic research, and incident response findings published through April 2026. It focuses on enterprise infrastructure vulnerabilities—software and hardware deployed in organizational environments, as distinct from individual consumer devices or cloud platform provider vulnerabilities. Statistical claims are drawn from sources identified in the References section; where precise numbers are unavailable, the paper uses qualified language to convey directional assessments. The paper does not rely on internal CSA assessments or unpublished research in its referenced citations.

Understanding the Exploit Window

Defining Mean-Time-to-Exploit

Mean-time-to-exploit (MTTE) describes the interval between the public disclosure of a vulnerability and confirmed evidence of that vulnerability's active exploitation in the wild. The metric has several important nuances. MTTE distributions are not uniform: some vulnerabilities are never exploited, while others see exploitation within hours of disclosure. The median is a more informative statistic than the mean for most practical purposes, because a small number of rapidly exploited, high-profile vulnerabilities can dramatically distort the average upward, while the mass of slowly exploited or unexploited CVEs anchors the lower tail.

CISA's Known Exploited Vulnerabilities catalog provides the most authoritative public measure of confirmed exploitation, but it understates exploitation velocity in an important way: inclusion in the catalog requires CISA to confirm exploitation with sufficient confidence to mandate federal agency remediation, which introduces a verification delay. Exploitation researchers and threat intelligence firms often document active exploitation days or weeks before KEV inclusion. The KEV metric therefore represents a conservative lower bound on exploitation speed, and the compression it shows is almost certainly understating the true acceleration.

The Pre-Patch Exploitation Problem

One of the most significant developments captured in recent threat data is the emergence of exploitation that precedes patch availability. Rapid7's 2026 analysis documented that a significant fraction of Known Exploited Vulnerabilities showed exploitation preceding or coinciding with CVE publication—meaning some vulnerabilities were being actively exploited in the wild before the security community had formally assigned them an identifier, let alone before any patch existed. [1][10] This represents a material shift from the historical pattern in which zero-day exploitation capability was primarily concentrated among nation-state actors and the most sophisticated criminal organizations, rather than accessible to the broader exploitation ecosystem.

The explanation lies in the asymmetry of disclosure pipelines. Vendors frequently know about vulnerabilities for days or weeks before publishing their advisories and associated CVE identifiers; during this period, coordinated disclosure processes are underway, but threat actors with access to similar research, with contacts in underground markets, or with their own vulnerability research programs may already possess the technical knowledge required to exploit the flaw. When a patch ships, the act of shipping itself broadcasts the existence of the vulnerability to anyone capable of analyzing the binary difference between the patched and unpatched versions.

The Binary Diff Problem

The release of a security patch simultaneously discloses the vulnerability it corrects to any analyst with the patience and skill to compare the updated binary against its predecessor. This technique—patch diffing—has been practiced by security researchers and threat actors for years. What AI has changed is the speed and accessibility of the analysis. Where a skilled reverse engineer might require days to isolate a patched code path, understand its security implications, and develop a working exploit primitive, AI-augmented analysis pipelines can process the binary diff, extract potentially relevant changes, score them by security significance using large language models, and provide developers of exploit tooling with a focused starting point. [9] Security practitioners have observed that AI-augmented patch analysis can substantially compress the time required for triage and comprehension—phases that historically created the longest delays between patch release and exploit availability.

The practical effect is that the window between patch release and exploit availability has collapsed from weeks to a timeframe measured in hours for many vulnerability classes. As security researcher Chris Wysopal observed, once patches ship, attackers can "differentiate the patch, isolate the vulnerable code path, and use automation and AI to generate working exploit paths far faster than enterprises can test and deploy." [9] The structural implication is that shipping a patch is now closer to full public disclosure of a vulnerability than it was when manual analysis dominated the exploit development workflow.

The AI Exploitation Toolkit

AI-Assisted Exploit Development

The most direct evidence that AI has altered exploit development timelines comes from controlled research studies that have operationalized LLM-based exploitation agents against real vulnerability inventories. The landmark study by Fang, Bindu, Gupta, and Kang at the University of Illinois Urbana-Champaign demonstrated in April 2024 that GPT-4, when configured as an autonomous agent with tool-calling capabilities, could exploit 87 percent of a dataset of 15 real-world one-day vulnerabilities when provided with the CVE description—a success rate that no other tested model, including GPT-3.5 and leading open-source LLMs, came close to matching, and that neither commercial vulnerability scanners achieved in equivalent conditions. [5] The 87 percent figure should not be extrapolated carelessly to claim AI can exploit any published CVE with minimal human involvement—the evaluation was bounded by a specific dataset, and the success rate dropped dramatically to approximately 7 percent when CVE descriptions were withheld, highlighting that automated exploitation still struggles with vulnerability discovery relative to exploitation of known flaws. [5] But as a measure of what a capable AI agent can accomplish with the information that is publicly available on CVE publication day, the result is significant.

Subsequent research scaled this capability substantially. A multi-agent framework evaluated against 841 CVEs published in 2024 and 2025 successfully reproduced exploits for approximately 51 percent of the dataset—428 verified exploits—at an average cost of approximately $2.77 per CVE. [22] The cost dimension matters: it describes the economics of large-scale exploitation, where the marginal cost of targeting an additional vulnerable organization approaches the cost of one LLM inference call. The general direction of these results—across different methodologies, vulnerability types, and research teams—suggests that AI-assisted exploit generation has moved from a demonstration capability to a practical tool for well-resourced threat actors.

Google's DeepMind and Project Zero provided a complementary data point in late 2024, when their Big Sleep agent autonomously discovered a previously unknown buffer underflow vulnerability in the SQLite database—demonstrating that AI-assisted exploitation is not limited to reproducing known flaws but extends to discovering novel attack surfaces in widely deployed software. [21]

The Economics of Exploitation at Scale

The cost compression that AI enables in exploit development is not merely a technical curiosity; it transforms the threat model for enterprise security programs. When exploit development required weeks of skilled labor, it was economically rational only for high-value targets. The resources required to develop and operationalize a working exploit against a specific enterprise application imposed natural selection pressure on which targets threat actors pursued.

AI eliminates much of that selection pressure. If a capable language model can produce a functional exploit for a published CVE at costs approaching one dollar, [20] the economic calculus that once constrained targeting no longer applies. Mid-market enterprises, government contractors, critical infrastructure operators, and healthcare organizations—previously less attractive to sophisticated actors because the cost of bespoke targeting exceeded the expected return—become viable targets when the variable cost of exploitation drops by orders of magnitude. This proliferation of exploitation capability does not mean that all organizations face the same threat actors; nation-states retain specialized capabilities and pursue strategically motivated targeting. But it does mean that the volume of opportunistic exploitation can scale with the total population of unpatched vulnerable systems rather than with the labor capacity of criminal organizations.

Agentic Exploitation Frameworks

The emergence of agentic AI architectures—systems where AI models plan and execute multi-step attack sequences with minimal human intervention—represents a qualitative escalation beyond AI-assisted exploit code generation. In agentic exploitation, the AI system does not merely produce code for a human operator

to deploy; it autonomously conducts reconnaissance, identifies exploitation opportunities, develops and executes exploit primitives, confirms successful access, and may proceed to post-exploitation activities including lateral movement and data staging.

Controlled research provides early evidence of the operational viability of such frameworks. In a 2025 evaluation conducted against a university network of approximately 8,000 hosts across 12 subnets, an agentic multi-agent framework called ARTEMIS—featuring dynamic prompt generation, specialized sub-agents, and automatic vulnerability triaging—identified 9 valid vulnerabilities with an 82 percent valid submission rate, outperforming 9 of 10 human security professionals in the same evaluation. [16] This is a research context, not an adversarial one, and the framework operated within defined ethical constraints; but it demonstrates the trajectory of agentic security tools that operate without per-task human approval across complex, realistic enterprise network topologies.

The implications for enterprise defenders are direct. If agentic AI systems can navigate a realistic enterprise network, enumerate attack surfaces, and identify exploitable vulnerabilities faster than human red team members, it is reasonable to expect that adversarially oriented versions of similar architectures are being developed by well-resourced threat actors. Unit 42's 2026 Global Incident Response Report documented AI being leveraged across reconnaissance, phishing, scripting, and operational execution phases of confirmed intrusions, enabling "machine-like speed at scale" across the attack lifecycle. [4] The transition from AI-assisted human attackers to AI-autonomous attack pipelines may already be underway in the most sophisticated threat actor programs.

The Widening Defender Gap

Enterprise Patch Cycles Under Structural Stress

The compression of attacker timelines would matter less if enterprise patch cycles had kept pace. They have not. The Verizon 2025 Data Breach Investigations Report, which analyzed 22,052 security incidents and 12,195 confirmed data breaches, found that organizations took a median of 32 days to patch known vulnerabilities, and that only 54 percent of vulnerable devices were fully remediated within a year of the vulnerability's disclosure. [3] Set against a median time-to-exploit of 5 days, the arithmetic describes a structural exposure: for the median enterprise, attackers have the ability to operationalize a critical vulnerability and achieve initial access in roughly one-seventh the time it takes the organization to deploy a patch.

This is not an argument that enterprises are careless or negligent. Enterprise patch management involves genuine operational constraints. Patches must be tested against production configurations to avoid introducing system failures, a process that cannot be safely compressed below some minimum timeline for

complex infrastructure. Organizations with large estates of heterogeneous systems—the modal enterprise—cannot maintain the inventory visibility necessary to identify all affected assets and prioritize them consistently. Many vulnerabilities affect products for which no patch has been released by the vendor, placing organizations in the position of relying on compensating controls or accepting residual risk while awaiting remediation options. And security teams routinely operate with insufficient staff to execute even their defined patch processes against the volume of critical and high vulnerabilities disclosed each month.

The volume challenge alone is severe. CISA added 245 vulnerabilities to its Known Exploited Vulnerabilities catalog in 2025—more than 30 percent above the trend from prior years—bringing the catalog's total to 1,484 confirmed actively exploited vulnerabilities. [7][14] Each of those entries represents a vulnerability that is known to be exploited in production environments. Against a backdrop where organizations are remediating critical vulnerabilities in a median of 32 days, and where roughly 46 percent of vulnerable devices remain unpatched after a full year, the scale of the exposure is systemic.

Edge Infrastructure as Ground Zero

The concentration of exploitation activity is not uniform across vulnerability types. The Verizon 2025 DBIR found that 22 percent of all vulnerability exploitation breaches targeted edge infrastructure—firewalls, VPN concentrators, and remote access gateways—an eightfold increase over the previous year's proportion. [3] This pattern reflects attacker logic: edge devices offer network access without requiring initial compromise of endpoint systems, they often run firmware update cycles that lag behind endpoint products, and they sit at a network boundary where traffic volumes and protocol diversity can complicate behavioral anomaly detection. For organizations that have invested heavily in endpoint detection and response capabilities, the edge often remains a relative blind spot.

Rapid7 documented that newly exploited vulnerabilities in Microsoft SharePoint and Cisco ASA and FTD products experienced mass exploitation in 2025 that demonstrated the "narrowing window between patch disclosure and in-the-wild attacks." [2] The pattern in these campaigns is consistent: a patch ships, threat actors analyze the binary change, a functional exploit is developed and weaponized, and organizations that have not yet completed their patch testing cycles find themselves facing active exploitation of a vulnerability they knew was critical before their remediation process could conclude. The Rapid7 Q3 2025 report quoted Christiaan Beek, senior director of threat intelligence, as observing: "Attackers are no longer waiting. Instead, they're weaponizing vulnerabilities in real time and turning every disclosure into an opportunity for exploitation." [2]

The edge infrastructure concentration also reflects a change in how attackers approach enterprise networks. Where intrusions once predominantly began with endpoint compromise—through malware delivery via phishing or malicious downloads—the shift toward edge device exploitation provides a different kind of initial access: one that begins at the network perimeter with device-level control, without requiring

any user interaction or endpoint security product circumvention. This makes the exploitation of perimeter devices especially high-value: a compromised firewall or VPN gateway provides both network access and a persistent, difficult-to-detect foothold from which lateral movement can proceed.

The Identity Collapse Multiplier

The compressed exploit window interacts with a parallel vulnerability class that does not depend on patching at all: identity-based compromise. Unit 42's 2026 Global Incident Response Report, drawing on more than 750 major incident response engagements across more than 50 countries, found that identity weaknesses played a material role in nearly 90 percent of investigated intrusions, and that 65 percent of initial access was identity-driven—relying on stolen credentials, MFA bypass, and IAM misconfigurations rather than software vulnerability exploitation. [4] The fastest 25 percent of intrusions in the Unit 42 dataset reached data exfiltration in just 72 minutes from initial access, a fourfold acceleration over the prior year. [4]

This dynamic does not reduce the significance of MTTE compression; it compounds it. When an attacker achieves initial access through a rapidly exploited edge vulnerability, the subsequent movement across the environment is accelerated by the prevalence of identity weaknesses. Compromised credentials obtained from prior breaches, widely available in criminal marketplaces, are used to move laterally across systems without triggering the vulnerability-scanning detection that might identify traditional exploitation techniques. The result is that the time from initial compromise to material impact—the window during which defenders have the opportunity to detect and contain the intrusion—is now measured in hours rather than days for the fastest-moving threats. Google's M-Trends 2026 data documents exploitation preceding patch availability in some vulnerability categories—a pattern in which confirmed exploitation precedes widespread patch deployment, confirming that the window between disclosure and exploitation has inverted for a growing share of high-severity vulnerabilities. [18] Detection arriving after the first privileged lateral movement is structurally late, and material impact in ransomware and data exfiltration scenarios can follow within the same operational window.

The Industrialization of Exploitation

Criminal Ecosystem Maturation

The acceleration of exploitation timelines is not solely a function of AI capabilities; it reflects the maturation of a criminal ecosystem in which exploit development, initial access, and monetization have been separated into specialized functions linked by market mechanisms. Initial access brokers—criminal actors who specialize in compromising enterprise environments and selling access rather than monetizing the access themselves—have strong economic incentives to weaponize vulnerabilities as rapidly as possible after disclosure. Access

to a newly vulnerable enterprise is a perishable commodity: its value declines sharply once the organization patches or detects the intrusion. This economic structure creates market pressure to exploit as rapidly as possible and to scale across as many vulnerable targets as the attack infrastructure can reach.

Rapid7's research found 88 active ransomware groups operating in Q3 2025, up from 65 in Q2 and 76 in Q1, with ransomware involved in 42 percent of Rapid7 MDR investigations across the year. [2] The ecosystem supporting these groups has developed the characteristics of a mature market: initial access brokers selling directly to ransomware affiliates, exploit code available through underground marketplaces, AI-assisted tooling for reconnaissance and post-exploitation, and leaked or publicly available toolsets that continuously lower the capability threshold for entry-level actors.

The adaptive malware LAMEHUG, documented by Rapid7 in 2025, illustrates the integration of AI into criminal tools at the production level. LAMEHUG uses AI components to dynamically generate new commands in response to the specific environment it encounters [2]—a design that presents challenges for signature-based detection methods and shifts the detection burden toward behavioral and anomaly-based approaches. This type of AI integration—not AI as an external research assistant, but AI as an embedded component of deployed malware—represents the direction of development for the criminal ecosystem's most capable actors.

The Volume and Velocity Challenge

The combination of faster weaponization and a higher volume of disclosed vulnerabilities creates a compound challenge for security operations centers. The approximately 130 to 140 new CVE identifiers published daily represent a stream of potential targets that no human team can individually assess and prioritize in real time. [6] When the median time to KEV inclusion drops to 5 days and the fastest exploited vulnerabilities are weaponized within hours, the human review cycles that anchor traditional threat intelligence operations are structurally inadequate for the most time-sensitive decisions.

The ransomware leak site data quantifies the downstream consequences: Rapid7 documented 8,835 ransomware leak posts in 2025, a 46.4 percent year-over-year increase, representing the organizational victims whose data was published following a ransomware intrusion. [1] Not all of these intrusions exploited rapidly weaponized vulnerabilities, but the volume indicates the scale at which the exploitation ecosystem is operating and the breadth of organizational exposure. The IBM 2025 Cost of a Data Breach Report found that organizations using AI and automation in their security operations identified and contained breaches roughly 80 days faster than those relying on manual processes, and saved an average of $1.9 million per incident. [8] The implication is not that AI-assisted defense eliminates the attacker's advantage, but that the organizations that fail to adopt automated detection and response capabilities will face materially worse outcomes as the gap widens.

Organizational Risk Implications

The Structural Exposure

Organizations must update their threat models to reflect that the assumption of a meaningful patch window is no longer empirically supported. A vulnerability rated CVSS 9.0 that is disclosed on a Tuesday cannot be treated as a thirty-day remediation target when threat intelligence shows that the median time to exploitation for comparably rated vulnerabilities is five days and that the fastest campaigns achieve exploitation within hours. The thirty-day target may remain appropriate as a *maximum* for lower-priority vulnerabilities, but for critical vulnerabilities affecting externally facing infrastructure, it describes exposure to a window of exploitation that is nearly seven times the window attackers require.

This recalibration has direct implications for risk quantification. Risk formulas that incorporate probability of exploitation over a given remediation period must be updated with exploitation timeline data that reflects current attacker capabilities rather than historical norms. Organizations that report vulnerability risk metrics to boards and executives using outdated MTTE assumptions are materially misrepresenting their exposure.

Prioritization Under Pressure

The volume of critical vulnerabilities disclosed each month creates prioritization pressure that AI can help alleviate but cannot fully resolve. Not all CVSS 9.0 vulnerabilities are equal in terms of enterprise risk: the risk profile depends on whether the vulnerable component is present in the environment, whether it is network-accessible, whether exploit code is publicly available, whether threat intelligence shows active exploitation targeting the organization's industry, and whether compensating controls reduce the likelihood of successful exploitation even if the patch is not yet applied. Risk-based vulnerability management—prioritization that incorporates these factors rather than relying solely on CVSS scores—is not new, but its adoption has been inconsistent, and the compression of exploitation timelines makes ad hoc or CVSS-only prioritization increasingly inadequate.

CISA's Known Exploited Vulnerabilities catalog represents the minimum threshold of urgency: vulnerabilities on that list have been confirmed as actively exploited and must be treated as immediate remediation priorities, not queue items. In 2025, 245 new entries were added to the catalog, meaning organizations that prioritize remediation based on KEV inclusion were already operating reactively in many cases—responding to confirmed active exploitation rather than preventing it. [7] The faster-moving organizations in the industry have shifted toward prospective prioritization informed by exploit code availability, threat actor tradecraft, and vulnerability reachability assessment, attempting to remediate before KEV inclusion rather than in response to it.

Behavioral Detection as a Compensating Control

When the patch window cannot be compressed to match attacker timelines, behavioral detection provides a critical compensating control. An unpatched vulnerability can still be detected at the exploitation or post-exploitation stage if the organization has deployed endpoint detection and response (EDR), network detection and response (NDR), and identity threat detection capabilities tuned to the behavioral signatures of post-exploitation activity. This is not a substitute for patching—behavioral detection has its own failure modes, including zero-dwell attacks where the attacker completes their objective before detection fires—but it provides the organization with opportunities to interrupt the attack chain between initial access and material impact.

The Unit 42 finding that the substantial majority—87 percent—of intrusions involved activity across multiple attack surfaces—endpoints, networks, cloud, SaaS, and identity systems simultaneously—underscores both the challenge and the opportunity. [4] Attacks that span multiple surfaces create more detection signals but also require integrated analysis to identify the intrusion pattern across telemetry sources that are typically managed by different tools and teams. Security operations capabilities that can correlate cross-surface signals and automate initial triage are positioned to detect complex intrusions faster than those operating in siloed tool environments.

Recommendations

Immediate Actions

Organizations that have not already responded to the compression of exploitation timelines should treat the following as time-sensitive priorities.

Externally facing infrastructure—firewalls, VPN concentrators, remote access gateways, web application servers, and any other component directly reachable from the internet—must be subjected to an emergency patch cadence that is decoupled from the standard monthly patch cycle. When a critical vulnerability affecting edge infrastructure is disclosed, the question of whether to apply the patch should be resolved the same day; the question of when should be resolved within 24 to 48 hours, not 30 days. Organizations that lack the ability to apply emergency patches to edge infrastructure within 48 hours of a critical disclosure should treat that capability gap as a top-tier security program investment. For organizations where 48-hour emergency patch capability is not yet achievable, compensating controls—virtual patching through web application firewall or intrusion prevention rules, temporary network access restrictions for affected assets, and enhanced monitoring—should be activated within the same window while the patching process proceeds. This may require changes to change management processes that have historically prioritized operational stability over security velocity.

The CISA Known Exploited Vulnerabilities catalog must be integrated into patch management workflows as an escalation trigger. Any vulnerability that appears in the KEV catalog should immediately move to the front of the remediation queue with a target remediation window of 48 hours for externally accessible instances, independent of the organization's standard patch cycle schedule.

Inventory completeness for externally facing assets is a prerequisite to either of the above. Organizations that lack a continuously updated inventory of externally accessible infrastructure cannot operationalize either emergency patching or KEV-based escalation. Attack surface management tooling, combined with continuous scanning of publicly addressable infrastructure, provides the operational foundation required to execute on this guidance.

Short-Term Mitigations

Within a timeframe of 30 to 90 days, organizations should pursue a set of capability improvements that address structural weaknesses in their detection and response posture.

Risk-based vulnerability prioritization should incorporate exploit availability, threat actor targeting patterns, and asset criticality rather than relying on CVSS scores alone. Security vendors offering exploit prediction scoring systems, threat intelligence feeds that surface active exploitation campaigns, and vulnerability reachability tools that assess whether vulnerable components are network-accessible all provide inputs to a prioritization model that is better calibrated to actual exploitation risk. The goal is not to replace urgency-based escalation for the highest-severity vulnerabilities but to improve the quality of prioritization decisions for the long tail of high and critical vulnerabilities that cannot all receive immediate attention.

Network segmentation and application of zero trust principles to the most sensitive asset categories reduces the blast radius of successful exploitation. An attacker who achieves initial access through a perimeter vulnerability should encounter additional authentication and authorization boundaries before reaching the crown jewel assets that represent the highest-value targets. Organizations that have concentrated privileged access to sensitive systems in network zones adjacent to edge infrastructure should treat the segmentation of that access as a compensating control for the period during which their patch cadence cannot match attacker timelines.

Behavioral detection coverage for post-exploitation activity—lateral movement, privilege escalation, credential harvesting, data staging, and command-and-control communications—provides detection opportunities that do not depend on identifying the initial exploitation event. This coverage should be validated through purple team exercises that simulate the specific techniques documented in current threat intelligence, including techniques associated with AI-enhanced operational execution.

Strategic Considerations

The compression of mean-time-to-exploit is a trend, not an event; it will continue as AI capabilities improve and as the criminal ecosystem matures. Organizations that respond with purely tactical measures—faster patching, better detection—will find themselves perpetually responding to an adversary advantage that is growing. The strategic imperative is to invest in capabilities that are structurally adaptive rather than statically configured.

Automated patch testing and deployment capabilities represent the most direct response to the attacker's time advantage. Organizations that currently require weeks of manual testing before deploying critical patches can invest in automated regression testing, canary deployment frameworks, and virtual patching capabilities that provide temporary protection for unpatched vulnerabilities while full patch qualification proceeds. Vendor virtual patching capabilities—available through web application firewalls, intrusion prevention systems, and some endpoint protection platforms—can block exploitation of specific vulnerability patterns without requiring the underlying software to be patched, buying time for the testing cycle to complete.

AI-assisted security operations represent a second structural investment that addresses the volume problem. Security operations centers that continue to rely primarily on human analysts to review the daily stream of vulnerability disclosures, threat intelligence, and alert data will face an escalating workload as the volume of both disclosures and exploitation events increases. AI-assisted triage, automated initial enrichment of security alerts, and LLM-powered threat intelligence summarization can help human analysts concentrate their attention where it matters most. The IBM 2025 data documenting an 80-day improvement in breach identification and containment timelines for organizations using AI and automation extensively in their security programs suggests the operational impact is material. [8]

CSA Resource Alignment

MAESTRO: Agentic AI Threat Modeling

CSA's MAESTRO framework—Multi-Agent Environment, Security, Threat, Risk, and Outcome—provides organizations with a structured approach to understanding and defending against AI-enabled attack chains. [11] The framework's seven-layer architecture, from Foundation Models through Agent Ecosystem Integration, maps the attack surfaces that emerge when AI systems are deployed as either enterprise tools or adversarial weapons. For organizations grappling with AI-assisted exploitation, MAESTRO's layered threat model is directly applicable to understanding how agentic exploitation frameworks navigate enterprise environments: the framework's treatment of cross-layer exploitation, where a compromise at the infrastructure layer enables manipulation of higher-level agent behaviors, mirrors the lateral movement

patterns that AI-enhanced attackers are documented to employ. Security architects using MAESTRO to threat model their own agentic AI deployments will also be building the conceptual vocabulary needed to understand AI-assisted attacks against their conventional infrastructure.

The February 2026 CSA guidance on applying MAESTRO to real-world agentic AI threat models extends the framework from abstract threat taxonomy to practical threat modeling for CI/CD pipelines and production deployment architectures—a resource directly applicable to organizations seeking to evaluate their exposure to the agentic exploitation techniques described in this paper. [11]

AICM: AI Controls Matrix

The CSA AI Controls Matrix provides the most comprehensive available control framework for AI-related security risks, with 243 control objectives distributed across 18 security and governance domains. [12] For the exploitation acceleration threat described in this paper, several AICM domain clusters are particularly relevant. The Supply Chain Transparency domain addresses the risks introduced when AI-powered tools are integrated into development and security workflows, including the possibility that AI systems used for defensive purposes introduce their own vulnerabilities or supply chain risks. The Governance Risk and Compliance domain provides controls for integrating AI-specific threat intelligence into existing risk management processes. The Application and Interface Security domain addresses the security of AI models and interfaces deployed in enterprise environments, which are increasingly targeted by the same exploitation frameworks used against conventional software.

Organizations implementing the AICM can use self-assessment results to identify priority control gaps, particularly in areas where AI systems interface with vulnerability management workflows, patch automation pipelines, or security operations centers. The AICM's mappings to ISO 27001, NIST AI RMF 1.0, and EU AI Act provide integration paths for organizations already operating within established control frameworks.

CAVEaT: Cloud Adversarial Vectors, Exploits, and Threats

The CSA-MITRE CAVEaT collaboration provides a cloud-centric threat matrix that addresses the attack vectors most relevant to the exploitation patterns documented in this paper. [13] Edge infrastructure exploitation—the attack vector that accounted for 22 percent of vulnerability-initiated breaches in the Verizon 2025 data—falls within CAVEaT's coverage of cloud perimeter and network boundary attacks. Organizations using CAVEaT for threat-informed analysis can apply its framework to evaluate their detection and response coverage against the specific techniques associated with rapid n-day exploitation of cloud-adjacent infrastructure.

AI Organizational Responsibilities

CSA's AI Organizational Responsibilities framework, published across 2024, provides governance-level guidance for organizations deploying AI in security operations roles. [17] As organizations invest in AI-assisted vulnerability management, automated patch testing, and AI-augmented SOC capabilities, the accountability and risk management structures defined in the AI Organizational Responsibilities framework become operational requirements. Who is accountable when an AI-assisted prioritization system fails to flag an exploited vulnerability? How should organizations govern the use of AI for patch qualification testing in ways that do not introduce new risks? The framework provides the organizational scaffolding within which the technical recommendations in this paper can be safely implemented.

STAR: Security Trust Assurance and Risk

The CSA Security Trust Assurance and Risk program provides the assessment methodology through which organizations can evaluate their security posture, including their vulnerability management and AI security capabilities, against a structured set of requirements. For supply chains—the context in which the Verizon 2025 data showed third-party involvement in 30 percent of breaches—STAR provides a mechanism for organizations to assess their vendors' security posture rather than extending trust based solely on contractual attestation. As AI-assisted exploitation increasingly targets supply chain relationships as a path into enterprise environments, STAR-based third-party assessments become a relevant tool for managing the extended attack surface that partner dependencies create.

Conclusions

Across methodologically distinct sources, a consistent pattern emerges: the exploit window that enterprise security programs have historically relied upon as a buffer between vulnerability disclosure and active exploitation has not merely shrunk—it has collapsed for a significant fraction of critical vulnerabilities, and the trend has moved consistently in one direction. The combination of AI-assisted exploit development, automated patch diffing, agentic exploitation frameworks, and a mature criminal ecosystem that monetizes rapid exploitation has produced a threat environment in which a critical vulnerability disclosed on any given morning may be operationally exploited against unpatched enterprise systems within hours.

This is not an argument for despair. The recommendations in this paper describe concrete, executable responses that organizations can implement across immediate, near-term, and strategic timeframes. Emergency patch cadences for edge infrastructure, KEV-based escalation triggers, behavioral detection for post-exploitation activity, and AI-assisted security operations capabilities all represent actionable improvements that reduce organizational exposure. The IBM 2025 data showing that organizations with

extensive AI automation in their security programs identified and contained breaches 80 days faster than those without it demonstrates that the same technological developments that have accelerated attacker capabilities can be directed toward defense.

What the data does not support is the continuation of security program designs that assume a meaningful patch window. The 32-day median remediation time documented by Verizon, set against a 5-day median time to KEV inclusion, describes not a temporary operational gap but a structural condition that must be addressed through capability investment and process redesign. Organizations that achieve 48-hour emergency patch capability for edge infrastructure, that adopt prospective vulnerability prioritization, and that deploy behavioral detection across the attack lifecycle will be materially better positioned than those that do not—regardless of the continued evolution of attacker AI capabilities.

The era in which disclosure without immediate exploitation was the norm is over for critical, externally facing infrastructure, and is narrowing for enterprise environments broadly. [19] The security industry, and the enterprises it serves, must adapt to a threat landscape that operates at machine speed.

References

[1] Rapid7. "2026 Global Threat Landscape Report." Rapid7, 2026.

[2] Rapid7. "Q3 2025 Threat Landscape Report: Ransomware Alliances, AI Weaponization, and the Obsolescence of 'Time to Patch'." GlobeNewswire, November 2025.

[3] Verizon. "2025 Data Breach Investigations Report." Verizon Business, 2025.

[4] Palo Alto Networks Unit 42. "2026 Unit 42 Global Incident Response Report: Attacks Now 4x Faster." Palo Alto Networks Blog, February 2026.

[5] Fang, R., Bindu, R., Gupta, A., and Kang, D. "LLM Agents can Autonomously Exploit One-day Vulnerabilities." arXiv:2404.08144, University of Illinois Urbana-Champaign, April 2024.

[6] CISA. "Known Exploited Vulnerabilities Catalog." Cybersecurity and Infrastructure Security Agency, continuously updated.

[7] Cyble. "2025 CISA KEV Catalog Hits 1,484 Exploited Vulnerabilities." Cyble Blog, 2025.

[8] IBM. "Cost of a Data Breach Report 2025." IBM Security, 2025.

[9] CSO Online. "Patch Windows Collapse as Time-to-Exploit Accelerates." CSO Online, 2025.

[10] CyberMindr. "Average Time-to-Exploit in 2025." CyberMindr Insights, August 2025.

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

[12] Cloud Security Alliance. "AI Controls Matrix (AICM)." Cloud Security Alliance, July 2025.

[13] Cloud Security Alliance. "Cloud Adversarial Vectors, Exploits, and Threats (CAVEaT)." Cloud Security Alliance, November 2023.

[14] SecurityWeek. "CISA KEV Catalog Expanded 20% in 2025, Topping 1,480 Entries." SecurityWeek, 2025.

[15] Flashpoint. "Key Trends in Vulnerability Exploitation and Ransomware: Insights from the 2025 Verizon DBIR." Flashpoint Blog, 2025.

[16] "Comparing AI Agents to Cybersecurity Professionals in Real-World Penetration Testing." arXiv:2512.09882, December 2025.

[17] Cloud Security Alliance. "AI Organizational Responsibilities." Cloud Security Alliance, 2024.

[18] Google Cloud / Mandiant. "M-Trends 2026: Data, Insights, and Strategies From the Frontlines." Google Cloud Blog, April 2026.

[19] Qualys. "Zero-Day Zero: The AI Attack That Just Ended the Era of the Forgiving Internet." Qualys Security Blog, November 2025.

[20] CyberPress. "AI Systems Can Craft Exploits for Known CVEs in Minutes." CyberPress, 2025.

[21] Google Project Zero. "From Naptime to Big Sleep: Using Large Language Models To Catch Vulnerabilities In Real-World Code." Google Project Zero Blog, October 2024.

[22] "From CVE Entries to Verifiable Exploits: An Automated Multi-Agent Framework for Reproducing CVEs." arXiv:2509.01835, September 2025.