CSAI Foundation

Cloud Security Alliance AI Safety Initiative

The Collapsing Exploit Window: AI-Speed Vulnerability Weaponization

Systemic Enterprise Risk in the Age of Machine-Accelerated Exploitation

Unofficial AI-assisted Research

2026-04-25

Table of Contents

Executive Summary

The operational window between vulnerability disclosure and weaponized exploitation is collapsing. Once measured in months, this window has compressed to days and, with increasing frequency, to hours. The mean time to exploit a disclosed vulnerability fell from roughly 32 days in 2022 to approximately 5 days as measured for 2023 exploitation activity [1, 20], with 2025 data showing that 32.1% of newly tracked exploits appeared on or before the CVE's public disclosure date—an 8.5-percentage-point increase from 2024 [2].

Artificial intelligence has become the primary force driving this compression. AI systems can now generate working proof-of-concept exploit code for published CVEs in as little as 10 to 15 minutes at a cost of approximately one dollar per attempt [3]. The CVE-Genie multi-agent framework reproduced 51% of all CVEs published in 2024–2025—complete with verifiable exploits—at an average cost of $2.77 per CVE [4]. A single AI agent swarm identified more than 100 exploitable kernel vulnerabilities across major hardware vendors in 30 days at a total cost of $600 [5]. What previously required days of expert analysis has become a semi-automated process accessible to a dramatically wider population of threat actors.

Enterprises, meanwhile, are not remediating faster. The mean time to remediation for complex enterprise applications reached five months and ten days as measured in 2026 benchmark data [6]. Approximately 45% of enterprise vulnerabilities remain unpatched after twelve months [7]. The CISA Known Exploited Vulnerabilities catalog grew by 20% in 2025 alone, reaching 1,484 entries [8], and the CVE volume published in 2025 reached a record 48,185—a 263% increase from 2020 [9].

The result is a structural asymmetry that conventional patch management frameworks were not designed to absorb. Attackers operate at machine speed; defenders operate at organizational speed. Patch cycles measured in weeks or months bear no useful relationship to exploit windows measured in hours. This paper examines the mechanics of that asymmetry, analyzes its systemic dimensions for enterprise risk programs, and offers a prioritized response framework grounded in CSA's current AI risk governance work.

1. Introduction: A Window That Has Nearly Closed

The concept of the exploit window—the time between a vulnerability's public disclosure and its weaponized use against production systems—has always been a foundational parameter in enterprise patch management. Organizations built their vulnerability programs around it. The 30-day critical patch cycle, the 90-day remediation SLA for high-severity findings, the structured change management process for testing patches before deployment: all of these governance choices assume that some meaningful gap exists between when a flaw is known to defenders and when it is operational in the hands of attackers.

That assumption is no longer valid as a general operating principle. It remains applicable for a subset of vulnerabilities in a subset of environments, but it cannot serve as the default planning premise for enterprise vulnerability programs. The structural preconditions that sustained it—the time required for an attacker to understand a vulnerability, build or acquire an exploit, and deploy it at scale—have been systematically dismantled by the convergence of AI-powered tooling, underground exploit economies, and the proliferation of high-quality vulnerability disclosure data.

Understanding why requires separating the problem into its component parts. The exploit window has not merely shortened uniformly; it has bifurcated into a bimodal distribution in which some vulnerabilities are exploited within hours while others remain unweaponized indefinitely. The proportion falling into the former category is growing rapidly, and artificial intelligence is the primary mechanism of selection. Vulnerabilities that are well-documented, affect widely deployed software, and target commonly understood vulnerability classes—memory corruption, authentication bypass, deserialization—are increasingly candidate for same-day or pre-disclosure exploitation. Vulnerabilities that are obscure, poorly documented, or require deep target-specific knowledge remain in the longer tail.

For enterprise risk managers, this bifurcation is not a source of comfort. The vulnerabilities most likely to receive rapid AI-assisted weaponization are precisely those found in the software most widely deployed across enterprise environments: operating systems, browsers, edge network devices, virtualization platforms, and content management systems. These are not exotic targets. They are the backbone of every enterprise's attack surface.

This whitepaper proceeds in four analytical steps. First, it examines the quantitative evidence for exploit window compression, placing the current moment in historical context. Second, it analyzes the AI capabilities driving that compression and the economic dynamics making those capabilities accessible. Third, it examines the structural barriers on the defensive side that prevent patch deployment from keeping pace. Fourth, it synthesizes these dynamics into a systemic risk framework and offers practical guidance for enterprise security programs working to close the asymmetry.

2. Measuring the Collapse: From Months to Minutes

2.1 The Historical Trajectory

The modern exploit window has a measurable history. In 2018, the median time between vulnerability disclosure and confirmed exploitation was 756 days—more than two years [5, 20]. By 2021, that figure had declined significantly but remained above 100 days for most vulnerability classes. In 2022, the average sat at roughly 32 days [1]. By late 2023, that global average had collapsed to approximately 5 days [1, 20], and by 2025 a third of all exploitation events were occurring on or before the day of public CVE disclosure [2].

The Mandiant and Google Threat Intelligence Group analysis captured this trajectory in stark terms: "what once unfolded over weeks now materializes in days, and in some cases, minutes." [10]. VulnCheck's State of Exploitation report for the first half of 2025 put specific numbers to the acceleration: 432 CVEs with confirmed exploitation evidence were tracked in those six months alone, with 32.1% showing exploitation on or before disclosure day—up from 23.6% in 2024 [2].

Exploitation patterns have also become more concentrated by target class. Content management systems led with 86 vulnerabilities actively exploited in the first half of 2025, followed by network edge devices (77), server software (61), open source software (55), and operating systems (38) [2]. This is not a random distribution. It reflects attacker prioritization of targets that are high value, widely deployed, and—critically—well-understood enough for automated tooling to operationalize quickly.

2.2 The CVE Backlog Problem

Exploitability acceleration is occurring against a backdrop of record CVE volume. In 2025, a record 48,185 CVEs were published—a 263% increase from 2020—and submissions in the first quarter of 2026 were running approximately one-third higher than the same period in 2025 [9]. NIST, responding to the volume pressure, announced in April 2026 that it would triage NVD enrichment by prioritizing KEV catalog entries, federal software, and critical infrastructure applications, explicitly acknowledging that comprehensive coverage is no longer sustainable at current submission rates [9].

This creates a compounding problem. The CVE ecosystem already exhibits a structural lag: analysis by CyberMindr and others found an average 23-day delay between an exploit's publication and its formal CVE assignment, and approximately 80% of exploits are published before their corresponding CVEs receive official designation [1]. Defenders relying on CVE publication as a trigger for remediation action are therefore operating on a timeline that is already behind the threat curve before they start.

The CISA Known Exploited Vulnerabilities catalog, which was designed to cut through this noise by identifying confirmed in-the-wild exploitation, grew 20% in 2025 to reach 1,484 entries [8]. Of those entries, 304—roughly 20.5%—have been exploited by ransomware groups, and CISA identified 24 of the 245 vulnerabilities added in 2025 as known ransomware vectors [8]. The catalog provides a useful floor for prioritization, but by the time a vulnerability appears in KEV, exploitation is already underway.

2.3 Enterprise and Zero-Day Exploitation Trends

Google's annual zero-day exploitation review tracked 90 zero-day vulnerabilities exploited in the wild in 2025—higher than 2024's 78, though below the record 100 in 2023 [11]. The more significant finding was the composition: enterprise technologies accounted for 48% of tracked zero-days, an all-time high. Security and networking appliances—Cisco, Fortinet, Ivanti, and VMware products—comprised half of enterprise-targeting zero-days, reflecting the strategic value adversaries place on gaining footholds in infrastructure that is difficult to monitor [11].

Nation-state and financially motivated actors alike are sharpening their exploitation tempo. PRC-nexus groups demonstrated increased exploit sharing among affiliated threat clusters, reducing the time from initial exploit development to operational deployment. Ransomware operators exploited 9 zero-days in 2025, nearly matching the all-time high set in 2023, and commercial surveillance vendors—for the first time—were attributed more zero-days than traditional state actors, indicating that zero-day capability is proliferating to a broader customer base through commercial market mechanisms [11].

3. The AI Acceleration Engine

3.1 From Expert Craft to Automated Pipeline

Traditional exploit development required a specific sequence of expert labor: reverse engineering or source code analysis to identify the precise vulnerability mechanism, understanding of the target environment's memory layout and control flow, construction of a payload that reliably produces the desired effect, and iterative testing against representative systems. Each stage demanded specialist knowledge and meaningful time. A well-resourced threat actor might develop a working exploit for a complex vulnerability in days to weeks. An average attacker could not do it at all without acquiring an existing exploit.

AI has disrupted every stage of that process simultaneously. Language models trained on vulnerability research, CVE databases, security advisories, and publicly available exploit code can analyze a new CVE disclosure and produce candidate exploit code in minutes, without the manual reverse engineering phase. Multi-agent frameworks can automate the test-debug-iterate cycle that previously required a human expert continuously monitoring results. The net effect is a reduction in the skill and time required to move from disclosure to working exploit by an order of magnitude.

The economics are remarkable. Research and reporting from 2025 and early 2026 documented AI systems generating working CVE exploits in 10 to 15 minutes at approximately one dollar per attempt [3]. The CVE-Genie framework—a multi-agent LLM system designed to reproduce CVE vulnerabilities—successfully reproduced 51% of all CVEs published in 2024 and 2025, complete with verifiable exploits, at an average cost of $2.77 per CVE [4]. At that cost structure, an attacker willing to spend $100 could theoretically work through dozens of recent CVEs in a single automated overnight run.

3.2 The AI Offensive Tooling Ecosystem

These capabilities have not remained confined to academic research settings. Before GPT-4's public release in April 2023, fewer than five open-source AI-assisted penetration testing tools were in widespread use. By March 2026, that number had grown to more than 70 documented tools [5]. The ecosystem spans every phase of the attack chain, with particularly mature capabilities in reconnaissance—where AI-assisted scanning and enumeration approaches near-ceiling performance—and early exploitation.

The economic compression these tools introduce is extraordinary. Benchmarks published in early 2026 documented an AI framework compromising four of five hosts in a realistic Active Directory environment at a total cost of $28.50, compared to $15,000 to $50,000 for a manual penetration test of comparable scope

[5]. The CAI framework demonstrated a 156-times cost reduction relative to human-performed equivalents, executing the work 3,600 times faster [5]. RapidPen, another tool in this cohort, produced IP-to-shell access at $0.30 to $0.60 per IP address.

The operational model shift is as significant as the cost change. Traditional penetration testing proceeds sequentially: a human tester moves through one target at a time, applying judgment at each step. AI-assisted frameworks operate in parallel across entire attack surfaces simultaneously, testing every subdomain, port, and service at once [5]. The economic and speed properties that make traditional exploitation costly and slow—the need for human attention at each step—do not apply to automated AI pipelines.

CyberStrikeAI, an AI-powered attack framework published to GitHub in November 2025, had confirmed attacks against more than 600 devices across 55 countries by January 2026—a two-month deployment cycle [5]. The Model Context Protocol (MCP), which standardizes how LLMs interface with external tools, has become a force multiplier in this ecosystem, enabling attack tools to be composed and deployed with minimal integration work [13].

3.3 AI-Discovered Zero-Days: A New Threat Frontier

The most significant long-term risk may be less about AI-accelerated exploitation of known vulnerabilities and more about AI-autonomous discovery of unknown ones. Google's Big Sleep project—a collaboration between Project Zero and DeepMind—reported 20 previously unknown vulnerabilities in widely used open source software as of August 2025, including an exploitable stack buffer underflow in SQLite and a critical zero-day (CVE-2025-6965) that was at risk of active exploitation before Big Sleep's discovery and disclosure [12]. In January 2026, an AI agent system discovered all 12 zero-day vulnerabilities in a new OpenSSL release independently [5]. The same month, a separate AI agent swarm identified more than 100 exploitable kernel vulnerabilities across AMD, Intel, NVIDIA, Dell, Lenovo, and IBM driver codebases over a 30-day run costing $600 in compute [5].

These capabilities currently sit predominantly in the hands of well-resourced defenders, but that asymmetry is unstable. The same open-source model weights, the same multi-agent scaffolding, and the same vulnerability research corpora that enable defensive AI programs are accessible to offensive actors. There is no technical moat separating "defensive AI vulnerability hunting" from "offensive AI zero-day discovery" when the underlying methodology is identical. The defensive advantage today is primarily organizational—trusted access to vendor patch pipelines, coordinated disclosure relationships, and the ability to act on findings before publicizing them—rather than technical. Each new zero-day that enters the exploitation economy before a patch exists represents a period of structural defender disadvantage as AI capabilities continue to lower the barrier to novel vulnerability discovery.

4. The Defense Side: Why Patching Has Not Kept Pace

4.1 The Remediation Gap in Numbers

If attacker speed has accelerated by roughly a factor of six since 2022—from 32 days to 5 days to weaponize a disclosed vulnerability—enterprise remediation speed has, if anything, degraded. The April 2026 Qualys enterprise patch benchmark found that the mean time to remediation for complex enterprise applications reached 5 months and 10 days—a figure that would have been considered unacceptably slow a decade ago and is catastrophically misaligned with current exploitation timelines [6].

The Edgescan 2025 Vulnerability Statistics Report put a different dimension on the problem: 45.4% of discovered enterprise vulnerabilities remain unpatched after twelve months, with 17.4% of unpatched findings classified as high or critical severity [7].

The Rapid7 2026 Global Threat Landscape Report documented that confirmed exploitation of CVSS 7–10 vulnerabilities increased 105% year over year between 2024 and 2025, rising from 71 to 146 documented cases, and noted that "threat actors are weaponizing known, preventable conditions faster—not necessarily developing new attack sophistication" [10]. The implication is pointed: much of the increase in exploitation damage is not attributable to novel attacker techniques but to the widening window between known vulnerability and completed remediation.

4.2 Why Organizations Patch Slowly

The remediation gap is not primarily a technical problem. Organizations understand how to apply patches. The impediment is structural and organizational: complex enterprise environments require patch testing for compatibility with business-critical applications before deployment; change management processes require approvals and maintenance windows; legacy systems may lack vendor support or may break when updates are applied; and operational teams are under pressure from multiple competing priorities simultaneously.

The Qualys benchmark observed that complex runtime components—Java, .NET, Citrix Workspace App, Visual C++ Redistributable—consistently showed the longest remediation windows because they require extensive compatibility testing before deployment in production [6]. Even with automation handling 27% of patch deployments in the benchmark study, the remaining 73% requiring human involvement created a ceiling on remediation throughput that no amount of tooling alone can overcome.

These structural constraints are compounded by the volume challenge. A record 48,185 CVEs were published in 2025 [9], and with NIST explicitly acknowledging that it can no longer enrich all of them for risk context, organizations relying on CVSS scores and NVD metadata for prioritization are operating with incomplete information. The average 23-day lag between exploit publication and formal CVE assignment means that the period when exploitation is already active is also the period when organizations have the least authoritative risk context to act on [1].

The result is a triage problem with no clean solution. An organization cannot patch every vulnerability within 5 days—the operational cost would be paralyzing, and not every vulnerability demands that urgency. But the proportion of vulnerabilities that do demand near-immediate action is growing faster than organizations' capacity to identify and respond to them.

4.3 The CISA KEV Window: Narrowing Fast

The CISA Known Exploited Vulnerabilities catalog was designed to give organizations a prioritized, authoritative list of vulnerabilities with confirmed in-the-wild exploitation. For federal agencies, remediation deadlines apply to KEV entries. For all organizations, the catalog provides a practical floor: if a vulnerability appears in KEV, it has confirmed active exploitation and should be treated as a genuine emergency.

But the value of KEV as a planning tool depends critically on the lead time between catalog entry and organizational action. If adversaries are weaponizing vulnerabilities within 5 days, and the KEV listing process itself takes several days after confirmed exploitation is observed, the window available for remediation after a KEV entry may be shorter than the minimum time required to test and deploy a patch in most enterprise environments. For the eight vulnerabilities CISA added to KEV in April 2026 with federal patching deadlines of two to three weeks, an enterprise operating with a typical 30- to 60-day patch cycle is already structurally behind before it starts [15].

The February 2026 KEV update, which added four critical vulnerabilities requiring immediate action, and the April 2026 update, which added eight more with deadlines of April 23 and May 4, 2026, illustrate the operational cadence organizations are now expected to sustain [15]. Meeting that cadence while also maintaining the compatibility testing, change management, and operational oversight that enterprise governance requires is not an incremental improvement to existing programs—it is a different kind of program entirely.

5. Systemic Risk Dimensions

5.1 The Ransomware-Exploitation Nexus

The compression of exploit windows is not a generic threat; it has a specific and well-documented relationship with ransomware economics. Of the 1,484 vulnerabilities in the CISA KEV catalog as of 2025, 304—roughly 20.5%—have been exploited by ransomware groups [8]. Ransomware operators exploited 9 zero-days in 2025, and named threat actors including FIN11/CL0P targeted Oracle EBS vulnerabilities with zero-days, while UNC2165 (Evil Corp) used a WinRAR zero-day for its first confirmed zero-day initial access operation [11].

The ransomware ecosystem has always functioned on exploit arbitrage: finding and weaponizing vulnerabilities before the defender population can patch them, achieving initial access, and monetizing before detection. As AI tools compress the time required for that weaponization step, the arbitrage window widens for attackers even as the defender window narrows. Organizations that can close the gap—through rapid detection, aggressive patch prioritization, or compensating controls—remove themselves from the pool of viable targets. Those that cannot absorb the remediation speed increase become structurally more exposed.

5.2 Supply Chain Amplification

Vulnerability exploitation does not only threaten the organization that is directly targeted; it propagates through supply chains. The Black Kite 2025 Supply Chain Vulnerability Report documented how ransomware operators leverage vulnerabilities in widely deployed enterprise products—Microsoft Exchange, Cisco ASA, Fortinet FortiOS—as supply chain access vectors, moving from initial exploitation of one vendor's product to lateral access across that vendor's customer base [16]. Microsoft has consistently led CISA KEV additions by vendor, with 39 vulnerabilities added in 2025 and over 100 confirmed ransomware-related flaws in the catalog overall [8].

The supply chain dimension amplifies the systemic character of the exploit window problem. A single widely deployed software component with a zero-day vulnerability is not one vulnerability—it is a vulnerability replicated across every enterprise that runs that component, all of which are simultaneously exposed. The attacker does not need to target each organization individually; a single campaign exploiting a shared dependency reaches the entire customer base. For organizations whose security posture is predicated on patching at organizational speed, the supply chain exposure model breaks the assumption that defenders can respond faster than attackers can scale.

5.3 AI Systems as Novel Exploit Targets

The emergence of AI components in enterprise environments introduces a new class of vulnerability exposure that sits partially outside the traditional CVE ecosystem. The OWASP GenAI Exploit Round-up Report for Q1 2026 documented eight significant AI-related security incidents between January and April 2026, including a Flowise remote code execution vulnerability (CVE-2025-59528) that affected 12,000 to 15,000 exposed instances and a breach affecting approximately 150 GB of sensitive government data [13]. The report noted that most AI-related security events are not yet mapped to traditional CVE identifiers, meaning the existing vulnerability management infrastructure—already under strain—is not capturing the full exposure surface.

Attacks on AI systems are shifting from theoretical vectors to operational ones: prompt injection has evolved into a documented data exfiltration mechanism, agent orchestration layers are being targeted directly, and supply chain vulnerabilities in third-party AI tools have produced confirmed compromises [13]. The IBM X-Force research concurs, observing that agentic AI systems introduce new vulnerability classes—including excessive agency, insecure tool invocation, and trust-chain compromises—that do not map cleanly onto the CVE system or existing exploit detection methodologies [18].

For enterprise risk managers, the AI vulnerability dimension is not a separate problem from the exploit window problem—it is an extension of it. Organizations deploying AI components without applying the same vulnerability governance rigor applied to traditional software are expanding their attack surface into territory where the existing detection and remediation infrastructure does not yet operate effectively.

5.4 The Asymmetric Cost Structure

Traditional security economics assumed rough parity between the cost of attack and the cost of defense. An attacker investing significant time and resources in exploit development faced a defender who, with a reasonable patch program, could neutralize the investment. That cost parity is eroding. AI-assisted exploitation costs $1 per CVE tested at scale [3]; enterprise patch testing and deployment for complex applications costs months of team capacity. The defender's marginal cost of responding to each exploitation attempt is rising while the attacker's is falling.

This asymmetry has direct implications for how enterprise risk programs should allocate resources. Treating all CVSS-scored vulnerabilities as equivalent and working through them in scored order is a strategy calibrated for a world in which attacker capability was more uniformly distributed. In a world where AI tooling concentrates exploitation activity on well-documented, widely deployed vulnerabilities—the ones most likely to be amenable to automated weaponization—risk prioritization needs to incorporate exploitation likelihood as a primary factor, not merely severity.

The following table summarizes the key asymmetric parameters:

Metric	Attacker (Current)	Defender (Typical Enterprise)
Time to weaponize a CVE	10–15 minutes (AI-assisted) [3]	Not applicable
Cost to test a CVE	~$1 [3]	Not applicable
Mean time to remediate (complex apps)	N/A	5 months 10 days [6]
Vulnerabilities unpatched at 12 months	N/A	~45% [7]
CVEs exploited ≤ day of disclosure (1H 2025)	32.1% [2]	Baseline assumption often 30+ days
New CVEs published in 2025	48,185 [9]	Enrichment coverage declining [9]
Zero-days exploiting enterprise tech (2025)	48% of tracked zero-days [11]	Limited pre-patch mitigation options

6. The Dual-Use Paradox: AI as Both Threat and Defense

6.1 Defensive AI Capabilities

The same AI capabilities accelerating exploitation are producing genuine defensive advances that would not be possible at human speed. Google's Big Sleep agent identified 20 previously unknown vulnerabilities in widely used open source software, including a confirmed zero-day in SQLite that was preemptively patched before exploitation [12]. Google DeepMind's CodeMender applies Gemini Deep Think's reasoning capabilities to automated vulnerability repair, not merely detection [19]. The National Vulnerability Database and several commercial vendors have begun using AI-assisted analysis to accelerate CVE enrichment and prioritization.

The defensive proposition is sound in principle: if attackers use AI to find and exploit vulnerabilities faster, defenders should use AI to find and patch them faster. At sufficient scale and speed, an AI-powered defensive program could in theory neutralize the AI exploitation advantage by closing vulnerabilities before they can be weaponized. Google's framing of the Big Sleep work explicitly targets this possibility: autonomous discovery and coordinated disclosure ahead of attacker exploitation [12].

The practical challenge is organizational. AI-powered vulnerability discovery, even when successful, feeds into the same testing, change management, and deployment processes that are already the primary bottleneck in enterprise remediation. Discovering a vulnerability in an internally deployed application faster is useful only if the organization can also remediate it faster—a constraint that involves legal, operational, and governance dimensions that AI does not automatically resolve.

6.2 The Defensive Implementation Gap

The Rapid7 data makes the implementation gap concrete: the most common initial access vector in 2025 enterprise incidents was not sophisticated AI-discovered zero-days but valid account compromise with no multi-factor authentication, which accounted for 44% of initial access incidents [10]. Vulnerability exploitation accounted for 25%—significant, but competing with a much simpler attack pathway that is entirely unrelated to the exploit window problem.

This suggests that the exploit window compression phenomenon, while genuine and serious, does not uniformly dominate enterprise risk profiles. Organizations with severe gaps in foundational controls—identity hygiene, asset inventory, network segmentation—face a compound risk in which AI-accelerated exploitation is one of several concurrent exposure vectors, not an isolated problem amenable to a dedicated mitigation.

Remediation programs that focus narrowly on patching speed while neglecting identity and access management, network monitoring, or asset visibility will improve performance on one vector while leaving others unaddressed.

The dual-use paradox, then, is not merely technical: it is also strategic. Defenders have genuine AI tools available. The question is whether organizations can reorganize their security programs to take advantage of those tools at a pace consistent with the threat's evolution.

7. Enterprise Response Framework

7.1 Reframe Risk Prioritization Around Exploitability

The first and most consequential change enterprise security programs should make is to reframe vulnerability prioritization around exploitation likelihood rather than severity score alone. CVSS scoring measures the theoretical impact of successful exploitation; it does not measure how likely exploitation is to occur, or how quickly. In the current threat environment, a CVSS 8.1 authentication bypass in a widely deployed edge device that is actively being scanned by AI-assisted reconnaissance tools demands a materially different response than a CVSS 9.8 memory corruption vulnerability in an obscure application with no public exploit and no known threat actor interest.

Exploitation likelihood can be assessed through several practical proxies: CISA KEV catalog inclusion, active exploit evidence in threat intelligence feeds such as VulnCheck, EPSS (Exploit Prediction Scoring System) scores, presence in underground exploit markets, and vendor advisories indicating active exploitation. Organizations should treat KEV inclusion as an automatic escalation trigger with remediation timelines measured in days, not weeks—while reserving standard patch cycle timelines for vulnerabilities with no exploitation evidence.

Critically, this triage logic should be applied continuously, not at fixed intervals. A vulnerability that is a low-exploitation-likelihood finding on Tuesday may become a KEV entry by Thursday. Programs operating on monthly or quarterly review cycles will systematically miss the escalation trigger.

7.2 Build Compensating Control Layers for the Patch Gap

No enterprise of meaningful scale will close the gap between a five-day exploitation window and a five-month remediation timeline for complex applications entirely through patching speed. The structural constraints are real. The appropriate response is to build layered compensating controls that reduce the probability and impact of successful exploitation for vulnerabilities that cannot be patched immediately.

Network segmentation limits the lateral movement available to an attacker who achieves initial access through a vulnerable service. Strict egress filtering prevents compromised systems from establishing command-and-control connections. Web application firewalls and virtual patching can reduce the exploitability of known vulnerabilities while formal patches are prepared and tested. Endpoint detection and response tools can identify exploitation attempts in progress before they complete. Each of these layers adds friction to the exploitation chain without requiring the vulnerable software to be updated.

The specific compensating controls appropriate to each vulnerability class vary. For network edge devices—the most heavily targeted enterprise category in 2025, accounting for 77 actively exploited CVEs in the first half of the year alone [2]—the most impactful compensating controls include restricting management plane access to known administrative sources, enabling anomalous authentication alerting, and segmenting managed device traffic from general enterprise networks. For web-facing applications, runtime application self-protection and WAF rule updates tied to new CVE disclosures provide meaningful interim coverage.

7.3 Accelerate Patching for Highest-Risk Asset Classes

Even with triage logic and compensating controls in place, some asset classes should be subject to patching timelines that are structurally faster than current enterprise norms. Internet-facing systems, authentication infrastructure, network security appliances, and systems that process sensitive data represent a relatively small fraction of most enterprise asset inventories but carry disproportionate exploitation risk. Building dedicated fast-track remediation processes for these asset classes—with pre-approved change management, maintained testing environments, and on-call deployment capacity—is operationally feasible without redesigning the entire patch program.

The Qualys benchmark data offers a useful reference point: 27% of patches deployed in the most recent 12-month period were deployed autonomously with no human involvement [6]. Expanding autonomous deployment for well-understood, low-disruption patches on high-risk assets—browser updates, OS security patches for internet-facing systems, authentication platform updates—would materially improve patching speed for the asset classes most relevant to exploitation risk without requiring human bandwidth that does not exist.

7.4 Monitor Exploitation Intelligence in Near-Real Time

The 23-day average lag between exploit publication and CVE assignment [1] means that organizations relying on NVD as their primary intelligence source are operating with a systematic blind spot in the earliest and most dangerous phase of a vulnerability's exploitation lifecycle. Bridging that gap requires integrating threat intelligence sources that track exploitation evidence directly—CISA KEV update feeds, VulnCheck, Shadowserver, GreyNoise, and vendor PSIRT advisories—alongside traditional NVD-based scanning.

Automated ingestion of these sources into vulnerability management platforms, with alerting configured to flag new entries in known-exploited evidence sources, can reduce the organizational response lag even when the underlying patch timeline cannot be compressed. Knowing that a vulnerability affecting your environment has confirmed in-the-wild exploitation evidence before the CVE is formally assigned does not provide a patch, but it does enable faster compensating control deployment and executive escalation.

7.5 Apply AI Governance to AI-Introduced Vulnerability Surface

As organizations deploy AI components—language model APIs, agentic workflows, AI-assisted tools for security operations—they introduce a new category of exploitable software that does not map cleanly onto the traditional CVE ecosystem. Prompt injection, excessive agency, insecure tool invocation, and supply chain vulnerabilities in AI libraries are not currently tracked in CISA KEV or NVD at meaningful scale [13]. Organizations whose vulnerability programs are scoped only to CVE-assigned findings will miss this exposure surface entirely.

Extending vulnerability governance to cover AI components requires inventory visibility into every AI model and agent deployment, clear ownership for monitoring vendor AI security advisories, and integration of AI-specific vulnerability classes—drawn from the OWASP LLM Top 10 and agentic application security frameworks—into risk assessment workflows. This is not a separate security program but an extension of existing vulnerability management infrastructure to cover a new and growing asset class.

8. CSA Resource Alignment

The threat dynamics documented in this paper connect directly to multiple active CSA frameworks and programs, each of which addresses a distinct dimension of the exploit window problem.

The **AI Controls Matrix (AICM)** provides the most comprehensive governance coverage for organizations deploying AI components in their environments. Its treatment of AI supply chain security, model security, and application provider responsibilities maps directly onto the AI-introduced vulnerability surface described in Section 5.3. Organizations using AICM to assess their AI deployments should ensure that their control implementation includes provisions for AI-specific vulnerability monitoring and incident response, not solely the traditional CVE-scoped processes that existing controls often assume.

The **MAESTRO** threat modeling framework—CSA's methodology for agentic AI threat analysis—addresses the emerging exposure created by AI agents operating with access to security-sensitive tools and data. As AI-assisted exploitation frameworks increasingly target orchestration layers and agent identities rather than just model outputs [13], MAESTRO-based threat models should explicitly include the exploitation of AI management planes as an in-scope threat scenario. The proliferation of MCP-connected tools and agent frameworks makes this threat surface more concrete and more accessible to automated exploitation with each passing month.

CSA's Zero Trust guidance provides the compensating control architecture most relevant to the patch gap problem. Zero Trust's core principle—that no implicit trust is extended based on network location or device identity—is directly aligned with the threat model in which an attacker achieves initial access through a vulnerable, unpatched system and then moves laterally. Microsegmentation, continuous verification, and least-privilege access collectively limit the blast radius of successful exploitation even when patching cannot keep pace with the exploit window.

The **STAR for AI** program, including the Catastrophic Risk Annex currently in development, addresses the organizational accountability dimension of AI-accelerated threats. As AI tools become both the attack mechanism and the defense mechanism for vulnerability exploitation, the governance questions of who is accountable for AI-powered attack detection, AI-assisted remediation decisions, and autonomous patch deployment will require clear answers. STAR for AI provides the assessment framework within which those accountability questions can be structured and audited.

The **Agentic AI Red Teaming Guide** (2025) provides operational methodology for testing AI-integrated environments against the threat vectors most relevant to AI-accelerated exploitation, including MCP-based tool chain attacks, excessive agency exploitation, and supply chain compromise of AI dependencies.

Organizations moving AI components into production security tooling should treat the red teaming guide as a mandatory pre-deployment assessment framework.

9. Conclusions: Living in the Gap

The collapsing exploit window is not a vulnerability to be patched. It is a structural feature of the current threat environment, produced by the convergence of AI capability, economic incentives in the exploit economy, and the irreducible organizational complexity of enterprise patch management. No technical fix eliminates it. The question for security leaders is not how to close the gap entirely but how to manage within it—prioritizing effectively, building layered controls that buy time, and expanding the defensive AI capabilities that offer the only realistic path to attacker-speed response.

Several things are clear from the evidence reviewed in this paper. First, the traditional patch window assumption that justified 30-day critical remediation cycles is no longer defensible as a default planning premise. Organizations that have not revisited their patch SLAs in light of current exploitation timeline data are operating on a risk model that is structurally misaligned with the threat. Second, CVSS-based prioritization without exploitation likelihood context is increasingly insufficient as a triage mechanism. Third, the AI-introduced vulnerability surface—inadequately covered by existing CVE infrastructure—requires active extension of vulnerability governance rather than reliance on existing tooling to detect novel threats.

What is equally clear is that the path forward is not paralysis. Organizations that invest in real-time exploitation intelligence, fast-track patching for highest-risk assets, layered compensating controls, and AI-specific vulnerability governance are meaningfully better positioned than those that do not, even if they cannot match the exploitation speed of AI-powered adversaries in absolute terms. The goal is not to win a speed race that organizational constraints make unwinnable. It is to build a risk architecture resilient enough that the exploit window, however narrow, does not translate automatically into organizational compromise.

The AI Safety Initiative's ongoing work on AICM, MAESTRO, Zero Trust, and STAR for AI provides the framework foundations for that architecture. The translation of those frameworks into operational programs —with specific controls, metrics, and accountability structures calibrated to the current exploitation environment—is the work that remains. The urgency of that work is real. The compression from 756 days to 5 days happened over six years. The next compression, driven by increasingly capable AI discovery and exploitation systems, may not take nearly as long.

References

[1] CyberMindr. "The Race Against Exploitation: Average Time-to-Exploit in 2025." CyberMindr Blog, 2025.

[2] VulnCheck. "State of Exploitation – A Look Into the 1H-2025 Vulnerability Exploitation & Threat Activity." VulnCheck Blog, 2025.

[3] CybersecurityNews. "AI Systems Can Generate Working Exploits for Published CVEs in 10-15 Minutes." CybersecurityNews, 2025.

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

[5] Hadrian. "The AI Hacking Boom: What 70 New Offensive Security Tools Mean for Defenders." Hadrian Security Blog, 2026.

[6] Qualys. "Enterprise Patch and Remediation Benchmark 2026." Qualys Blog, April 2026.

[7] Edgescan. "The Vulnerability Backlog Crisis: Why 45% of Enterprise Vulnerabilities Never Get Fixed." Edgescan Research, 2025.

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

[9] NIST. "NIST Updates NVD Operations to Address Record CVE Growth." NIST, April 2026.

[10] Infosecurity Magazine. "AI-Enabled Adversaries Compress Time-to-Exploit." Infosecurity Magazine, 2025. (Citing Rapid7 2026 Global Threat Landscape Report.)

[11] Google Cloud Threat Intelligence. "Look What You Made Us Patch: 2025 Zero-Days in Review." Google Cloud Blog, 2026.

[12] TechCrunch. "Google Says Its AI-Based Bug Hunter Found 20 Security Vulnerabilities." TechCrunch, August 2025.

[13] OWASP Gen AI Security Project. "OWASP GenAI Exploit Round-up Report Q1 2026." OWASP, April 2026.

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

[15] The Hacker News. "CISA Adds 8 Exploited Flaws to KEV, Sets April–May 2026 Federal Deadlines." The Hacker News, April 2026.

[16] Black Kite. "2025 Supply Chain Vulnerability Report." Black Kite, 2025.

[17] Cyble. "2025 CISA KEV Catalog Hits 1,484 Exploited Vulnerabilities." Cyble Research, 2025.

[18] IBM Security. "Agentic AI Is Growing Fast, as Are the Vulnerabilities." IBM X-Force, 2025.

[19] Google DeepMind. "Introducing CodeMender: An AI Agent for Code Security." Google DeepMind Blog, 2025.

[20] Google Cloud Threat Intelligence. "Think Fast: Time Between Disclosure, Patch Release and Vulnerability Exploitation." Google Cloud Blog, 2024.