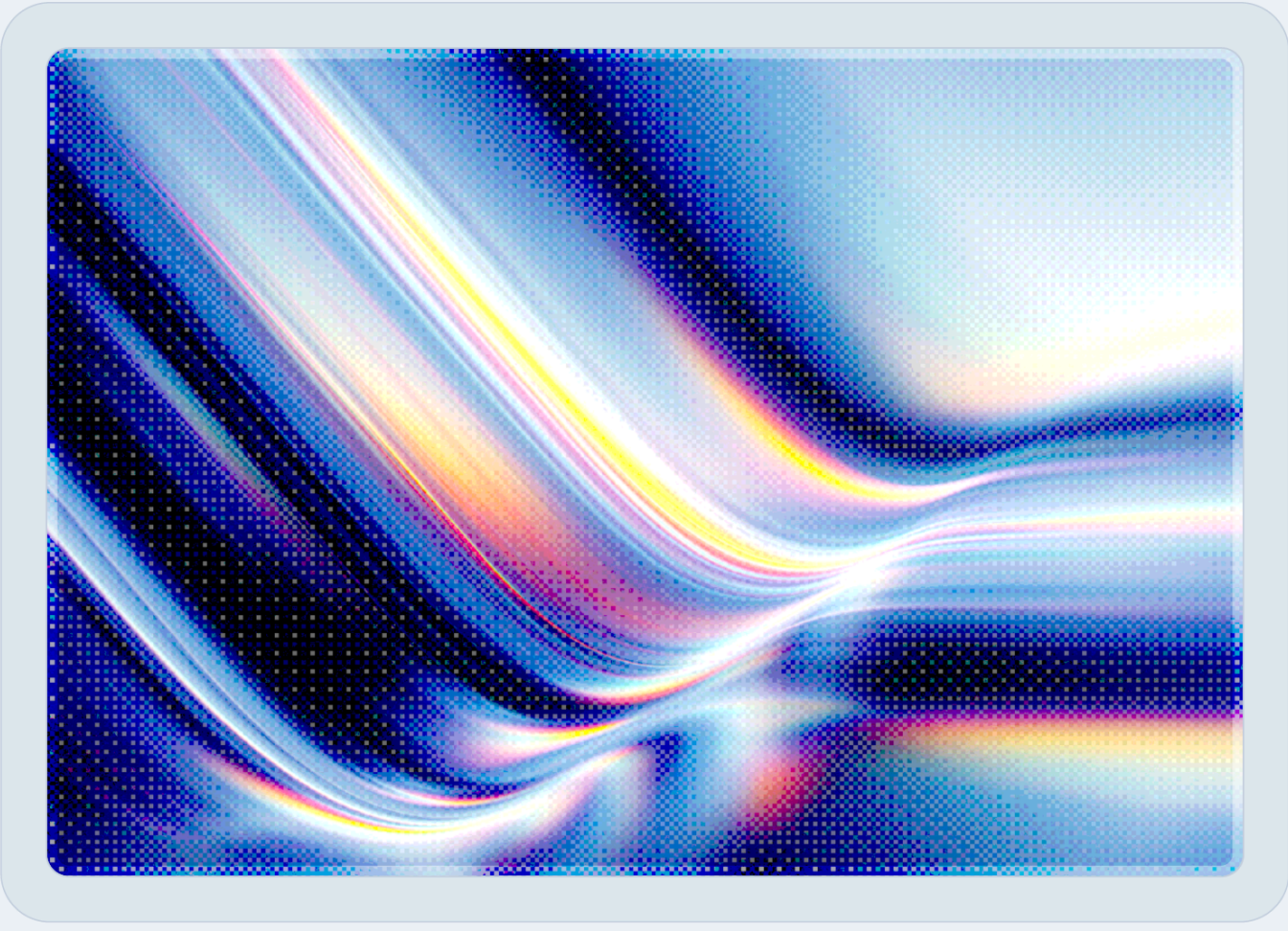
CSAI Foundation | Cloud Security Alliance

AI-Accelerated Exploitation and Asymmetric Vulnerability Velocity

DBIR 2026, Patch Debt, and Systemic Risk in the AI Threat Era

2026-05-30

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Table of Contents

Executive Summary

For nineteen consecutive years, the Verizon Data Breach Investigations Report (DBIR) documented stolen credentials as the most common way attackers gained their first foothold in a victim environment. The 2026 edition broke that pattern. Vulnerability exploitation overtook stolen credentials as the leading initial access vector for the first time in the report's history, accounting for 31% of known initial access vectors against 13% for credential abuse as an initial vector [1][2]. This inversion is not a statistical curiosity. It reflects a measurable change in attacker economics: exploiting an unpatched, internet-facing flaw has become faster, cheaper, and more reliable than acquiring and validating stolen credentials, and artificial intelligence is the principal force driving that shift.

The same DBIR, produced in collaboration with Anthropic, found that AI is compressing the attacker's time-to-exploit from months to hours [3]. Independent threat intelligence corroborates the trend from multiple vantage points. Mandiant's M-Trends 2026 places the mean time-to-exploit at negative seven days, meaning that on average, exploitation now begins before a patch is available [4]. Rapid7 reported that exploited high- and critical-severity vulnerabilities surged 105% year over year, while the median time from CVE publication to inclusion in CISA's Known Exploited Vulnerabilities (KEV) catalog fell from 8.5 days to 5.0 days [5]. VulnCheck found that nearly 29% of KEV vulnerabilities in 2025 were exploited on or before the day their CVE was published [6].

Against this acceleration, organizational remediation capacity is moving in the wrong direction. The DBIR found that the median time to fully patch a vulnerability stretched to 43 days in 2025, up from 32 days the prior year, and that only 26% of critical KEV vulnerabilities were fully remediated by the more than 13,000 organizations studied — down sharply from 38% a year earlier [2][7]. Analysis of one billion anonymized records by Qualys characterized the situation succinctly: "The remediation engine is running at the same RPM. The load has increased nearly eightfold" [7].

This paper argues that the core problem is no longer the existence of vulnerabilities but the widening gap between the speed at which attackers can weaponize them and the speed at which defenders can close them — a condition we term asymmetric vulnerability velocity. AI is widening that gap on the offensive side by automating reconnaissance, exploit development, and target selection, while the defensive side remains constrained by operational, organizational, and supply chain realities that no single enterprise can resolve alone. The result is a systemic risk that resembles the dynamics of financial contagion more than traditional, organization-bounded cyber risk. The paper offers a layered response framework spanning immediate triage discipline anchored on CISA KEV, medium-term adoption of

exploitation-informed prioritization beyond CVSS, and strategic investment in collective defense, and it maps these actions to Cloud Security Alliance frameworks including the AI Controls Matrix (AICM), MAESTRO, and the AI Vulnerability Storm program.

Introduction: The Inversion

For most of the modern history of breach analysis, the dominant story of initial access was a story about identity. Attackers phished, purchased, or harvested credentials, then logged in. The 2025 DBIR still placed credential abuse near the top of the initial access hierarchy at roughly 22% of breaches, with vulnerability exploitation trailing at approximately 20% after a sharp climb [8]. That gap had been narrowing for years. Exploitation as an initial access vector nearly tripled between the 2023 and 2024 reporting periods, rising 180% between 2022 and 2023 alone, and jumped from 14% to 20% of breaches between successive DBIR editions [8][10]. The trajectory pointed toward convergence, and in 2026 the lines crossed.

The 2026 DBIR analyzed more than 22,000 confirmed data breaches and 31,000 security incidents across 145 countries — the largest dataset in the report's nineteen-year history [1][11]. Within that dataset, vulnerability exploitation rose to 31% of known initial access vectors while credential abuse fell to 13% as a first-stage vector, even though credentials remained a factor in 36% of all breach action varieties [1][2]. The distinction between initial access and overall presence matters: credentials have not become irrelevant, and in specific breach patterns they remain decisive. In System Intrusion, stolen credentials and exploited vulnerabilities each appeared at 39%; in Basic Web Application Attacks, credentials were compromised in 52% of breaches; and in Public Administration, stolen credentials appeared in 59% of hacking-related breaches [2]. What changed is which mechanism most often opens the door.

This inversion is corroborated outside the DBIR. IBM's X-Force 2026 Threat Intelligence Index found that exploitation of public-facing applications surged 44% year over year to account for 40% of incidents, surpassing credential-based attacks at 32% and reversing two consecutive years in which credential abuse had dominated [12]. The convergence of independent datasets on the same conclusion strengthens confidence that this is a structural shift rather than an artifact of any single methodology.

The systemic significance of the inversion lies in what it reveals about relative cost. Credential-based intrusion depends on a supply chain of infostealer logs, combolists, and validation infrastructure that, while industrialized, still requires the attacker to find an account that maps to a target of value and to defeat whatever multi-factor and anomaly controls stand in the way [13]. Exploitation of an exposed, unpatched vulnerability bypasses identity controls entirely. As the cost and time of producing a working

exploit collapse — driven, as later sections detail, substantially by AI tooling — exploitation becomes the path of least resistance against the large and growing population of organizations that cannot patch fast enough. The inversion is therefore best read as a signal that the defensive bottleneck has moved. It is no longer primarily about authentication hygiene; it is about the velocity and completeness of remediation.

The DBIR 2026 Signal

The 2026 DBIR is valuable not only for the headline inversion but for the texture it provides on why the shift occurred and what it portends. The report's authors were explicit about the role of AI, having partnered with Anthropic to analyze 793 threat actors who violated acceptable use policies on the Claude platform. The median actor in that population sought AI assistance across approximately 15 distinct attack techniques, with some querying across 40 to 50 techniques, and AI-assisted text in phishing emails doubled year over year [3]. The report described AI as compressing the attacker's time-to-exploit from months to hours [3].

Crucially, the DBIR also tempered the AI narrative with an important qualification. Less than 2.5% of observed AI-assisted techniques involved rare or uncommon methods; AI primarily automates and scales known attack patterns rather than enabling genuinely novel capabilities [14]. This finding suggests that the near-term threat is not exotic. It is the industrialization of the ordinary — the same exploitation, reconnaissance, and social engineering techniques defenders already understand, executed faster, more cheaply, and at greater scale. Daniel Lawson, SVP Global Solutions at Verizon Business, framed the implication directly: "While the velocity of cyber threats — driven by AI and faster vulnerability exploitation — is increasing, the foundational principles of security and strong risk management remain the most effective defense" [1].

The trend data underlying the inversion is best understood across multiple DBIR editions, which the following table summarizes.

Initial access metric	2024 DBIR	2025 DBIR	2026 DBIR
Vulnerability exploitation (share of initial access)	~14% [10]	~20% [8]	31% [1]
Credential abuse (as initial vector)	dominant [10]	~22% [8]	13% [2]
Median time to fully patch a vulnerability	—	32 days [2]	43 days [2]
Critical KEV vulnerabilities fully remediated	—	38% [7]	26% [7]

Initial access metric	2024 DBIR	2025 DBIR	2026 DBIR
Median KEV CVEs an organization had to patch	—	11 [15]	16 [15]

The remediation columns tell the most consequential part of the story. The 34% increase in median full-patch time, from 32 to 43 days, occurred in the same year that exploitation overtook credentials and that AI compressed exploit development [2]. Detection itself did not slow — the median detection-to-closure interval held steady at 9 days — which indicates that the lag is concentrated in the full remediation cycle rather than in identifying that a problem exists [2]. Organizations are seeing the vulnerabilities; they are failing to finish closing them at the pace the threat now demands.

The volume figures reinforce why. According to the Qualys analysis of one billion anonymized records contributed to the DBIR, KEV-linked vulnerability instances grew 7.7-fold over four years, from 68.7 million to 527.3 million, with a 78% jump in 2025 alone [7]. The Day-28 open backlog — instances still unremediated four weeks after they should have been addressed — grew from 31 million to 184 million instances, and roughly 47 million instances, about 9% of the workload, have no near-term closure path [7]. Even proactive patching, while up 30% in absolute terms to 63.7 million instances, saw its remediation rate fall from 16.6% to 12.1% because volume outpaced capacity [7]. The DBIR's researchers characterized vulnerability management as a "sisyphean cause," noting that "there are often too many vulnerabilities and not enough time for patching all of them" [1].

Two further DBIR 2026 findings frame the broader environment in which exploitation now operates. Ransomware appeared in 48% of all breaches, up from 44%, though 69% of victims reported not paying and the median payment fell to roughly $140,000 [15][2]. Third-party supply chain breaches jumped 60% to account for 48% of all breaches, underscoring that an organization's window of exposure is not bounded by its own patching but extends through every dependency it cannot directly control [16].

Asymmetric Vulnerability Velocity

The central analytical concept of this paper is asymmetric vulnerability velocity: the structural mismatch between the rate at which attackers can convert a disclosed flaw into a working, deployed exploit and the rate at which defenders can deploy the corresponding fix across their estate. When these two velocities were measured in months and weeks respectively, the asymmetry was tolerable. Today the attacker side is measured in hours and the defender side in weeks to months, and the gap has become the dominant variable in breach likelihood.

The attacker side of the equation has compressed dramatically. Flashpoint data indicates that the average time from vulnerability disclosure to first exploitation fell from 745 days in 2020 to approximately 44 days in 2025 [17]. VulnCheck's measurements are sharper still: 28.96% of KEV vulnerabilities in 2025 were exploited on or before the day their CVE was published, up from 23.6% in 2024, and the figure reached 32.1% in the first half of 2025 [6]. The leading edge of this trend has crossed zero. Mandiant's M-Trends 2026 estimates the mean time-to-exploit at negative seven days, and Qualys Threat Research Unit independently characterized time-to-exploit as having reached "negative one day," meaning that for an increasing share of vulnerabilities, exploitation precedes the availability of a patch entirely [4][18]. CrowdStrike's 2026 Global Threat Report reinforces the point, finding that 42% of vulnerabilities were exploited before public disclosure and that average eCrime breakout time fell to 29 minutes, with the fastest observed breakout at 27 seconds [19].

The defender side has not kept pace. The DBIR's 43-day median full-patch time is consistent with other 2025-2026 measurements that show critical application and API vulnerabilities taking a median of 74.3 days to remediate and critical network vulnerabilities taking 54.8 days [20]. The Hadrian 2026 Offensive Security Benchmark put the average remediation time for high-severity vulnerabilities at 139 days [21]. Indusface reported that 32% of identified vulnerabilities remain unpatched for more than 180 days [22], and one industry analysis found that the median time to close half of an organization's internet-facing vulnerabilities is approximately 361 days [48]. The contrast is stark when the two velocities are placed side by side.

Dimension	Attacker velocity	Defender velocity
CVE-to-exploit (leading edge)	Negative 1 to negative 7 days [18][4]	n/a
CVE-to-exploit (typical)	~44 days [17]	n/a
Same-day exploitation share	28.96% of 2025 KEVs [6]	n/a
Critical app/API remediation (median)	n/a	74.3 days [20]
High-severity remediation (average)	n/a	139 days [21]
Full-patch cycle (DBIR median)	n/a	43 days [2]

Dimension	Attacker velocity	Defender velocity
Long-tail exposure	270,000+ systems still exposed to a 2020 flaw [24]	32% unpatched past 180 days [22]

This asymmetry is sustained, not transient, because the underlying CVE volume keeps the defender side perpetually behind. In 2025, 48,185 new CVEs were published, a 20.6% increase on top of a record 38% surge the prior year, which equates to roughly 131 new CVEs disclosed every day [25][23]. VulnCheck tracked more than 14,000 exploits developed for over 10,000 unique 2025 CVEs, a 16.5% year-over-year increase in same-year exploit coverage [6]. Network edge devices — firewalls, VPNs, and proxies — emerged as the most frequently targeted category in 2025, with security and perimeter software drawing a disproportionate share of both N-day and zero-day exploitation [26]. The mass-exploitation events of recent years illustrate the consequences: Cl0p's exploitation of the MOVEit Transfer flaw (CVE-2023-34362) began over a holiday weekend before patches were released and compromised more than 8,000 organizations globally, affecting an estimated 95 million individuals [27], while the Fortinet authentication bypass (CVE-2024-55591, CVSS 9.8) was exploited as a zero-day for two months before its January 2025 disclosure, with more than 50,000 internet-facing devices affected [28][59].

The long tail compounds the problem. As of February 2026, more than 270,000 systems remained exposed to SMBGhost (CVE-2020-0796), a six-year-old vulnerability, and a CVSS 9.8 flaw patched in July 2025 still had 18,000 vulnerable systems exposed seven months later [24]. N-day exploitation against this population requires no novel capability and sustains a low-cost exploitation economy that AI now makes even cheaper. The defender is not merely racing the clock on new disclosures; they are also defending an accumulating backlog that never fully clears.

AI as an Exploitation Accelerant

The mechanism behind the collapsing exploit window is increasingly well documented, and it is important to distinguish what is verified from what is extrapolation. The verified record establishes that AI systems can already perform substantial portions of the vulnerability discovery and exploitation workflow at low cost, and that threat actors are using these capabilities operationally. The extrapolation concerns how far and how fast these capabilities will generalize.

On the discovery side, DARPA's AI Cyber Challenge (AIxCC) provides the most rigorous public benchmark. In the August 2025 final, seven autonomous Cyber Reasoning Systems analyzed 54 million lines of code, identified 86% of synthetic vulnerabilities (up from 37% at the semifinals), patched 68% of those found, and discovered 18 previously unknown real-world flaws that organizers had not planted, at an average task cost of $152 and an average patch submission time of 45 minutes [29]. A peer-reviewed systematization of the competition found that LLM-enhanced systems discovered 22 proof-of-concept exploits that parallel fuzzing alone could not, demonstrating a genuine additive capability for complex input grammars and logical constraint solving [30]. That same analysis tempered the result by noting that even validated patches showed semantic failure rates between 37.7% and 45.6% on manual review, indicating that AI-assisted patching still requires human oversight [30]. Real-world deployments echo the benchmark: Google's Big Sleep agent discovered CVE-2025-6965, a memory corruption flaw in SQLite, before threat actors could exploit it [31], and the AISLE system identified twelve zero-day vulnerabilities in OpenSSL, three of which had persisted undetected since 1998–2000 despite millions of CPU-hours of fuzzing and audits by teams including Google Project Zero [32].

On the exploitation side, academic work established early that capability scales with context. University of Illinois researchers demonstrated in 2024 that GPT-4 autonomously exploited 87% of known one-day vulnerabilities when given the CVE description, while all other tested LLMs and open-source scanners scored 0%; without the CVE description, GPT-4's success rate fell to 7%, indicating that the model functions as a powerful exploit-generation engine when given advisory context rather than as an autonomous vulnerability discoverer [33]. The economics have since collapsed alongside the capability. Hadrian cataloged 70 open-source AI penetration testing tools as of March 2026, up from fewer than five before GPT-4's release, and documented an AI agent compromising four of five Active Directory hosts for $28.50, a framework achieving a 156-fold cost reduction versus manual testing, and agent swarms identifying more than 100 exploitable kernel vulnerabilities across major vendors in 30 days for $600 total [34]. The Hacker News reported that SWE-bench coding benchmark performance rose from 33% in August 2024 to 81% in December 2025, tracking closely with exploitation capability improvements [35].

These capabilities are no longer confined to research settings. Anthropic disclosed in November 2025 that a Chinese state-sponsored group, GTG-1002, used Claude Code to target approximately 30 global organizations, with the AI performing 80–90% of campaign tasks autonomously and human operators intervening at only four to six critical decision points per campaign [36]. Check Point Research documented a December 2025 to February 2026 breach in which a single operator compromised nine Mexican government agencies using more than 5,000 AI-executed commands — a case the researchers identified as the first in which "AI was not a productivity tool running in the background — it was the operational core of the attack" [37]. In a separate campaign, an AI-assisted actor using the open-source CyberStrikeAI platform compromised more than 600 FortiGate devices across 55 countries in five weeks, using commercial generative AI for tool development and attack planning [38]. The

democratization dimension is significant: tools such as ProjectDiscovery's Nuclei now support AI-powered template generation from natural language, allowing attackers to create targeted scanner templates without manual expertise [39].

The appropriate reading of this evidence is calibrated rather than alarmist. The DBIR's own finding that fewer than 2.5% of AI-assisted techniques were rare or uncommon indicates that AI's present effect is to scale and cheapen known attack patterns, not to invent fundamentally new ones [14]. There is also a documented asymmetry in how AI safety constraints apply: defenders conducting authorized red-team work routinely encounter guardrail refusals, while attackers bypass controls through jailbreaks, uncensored model variants, and multi-turn prompt attacks that have achieved 60% to over 90% success rates against safety judges [40]. The near-term threat is therefore best characterized as a capability that lowers the cost and raises the speed of attacks the defensive community already understands, while the longer-term trajectory — toward more autonomous, novel-capability systems — remains an area where extrapolation should be hedged and monitored rather than assumed.

The Patch Debt Ecosystem

If asymmetric velocity describes the dynamic, patch debt describes the accumulating liability it produces. Patch debt is the growing population of known, unremediated vulnerabilities that an organization carries forward because remediation capacity is structurally insufficient to clear new disclosures and the existing backlog simultaneously. The DBIR data already cited — a Day-28 backlog that grew from 31 million to 184 million instances, with 47 million having no near-term closure path — quantifies this debt at ecosystem scale [7]. Understanding why it accumulates requires looking at the supporting infrastructure, the triage signals, and the organizational realities that constrain remediation.

The most consequential infrastructure development is the partial breakdown of the National Vulnerability Database (NVD) enrichment pipeline. CVE submissions to NIST increased 263% between 2020 and 2025, and although NIST enriched nearly 42,000 CVEs in 2025 — 45% more than in any prior year — it could not keep pace [41][42]. In April 2026, NIST formally abandoned universal enrichment, classifying approximately 29,000 backlogged CVEs as "Not Scheduled" and committing to enrich only the 15-20% of incoming CVEs that intersect with CISA KEV, federal software, or the EO 14028 critical-software lists [41][43]. As a result, only about 32% of 2025 CVE entries have been fully enriched, and roughly 10,000 vulnerabilities from 2025 still lack a CVSS score [42][44]. The practical effect is that the authoritative source many vulnerability management programs depend on for severity and product-mapping metadata can no longer be assumed complete, forcing organizations toward commercial and community intelligence to fill the gap.

In this environment, CISA's KEV catalog functions as the most actionable triage signal available, precisely because it is curated for confirmed exploitation rather than theoretical severity. The catalog grew 20% in 2025, adding 245 entries to reach 1,484, with 24 tagged as actively used by ransomware groups and Microsoft leading all vendors with 39 additions [45][46]. Roughly 65% of the catalog's CVEs were actively exploited during the prior year [15]. Yet KEV is not a complete picture: Tenable independently tracks 1,924 vulnerabilities confirmed exploited in the wild against CISA KEV's 1,569, a gap of 355 exploited CVEs not yet cataloged [47]. KEV should therefore be treated as a high-confidence floor for prioritization, not a ceiling.

The organizational causes of patch debt are structural and largely resistant to exhortation. Legacy systems that cannot tolerate downtime, operational technology environments where patching risks safety-critical availability, change-management processes that introduce unavoidable testing latency, and the sheer complexity of modern software supply chains all impose remediation friction that exists independent of security intent. The supply chain dimension is especially intractable because an organization cannot patch a dependency it does not control; the DBIR's finding that third-party breaches jumped 60% to 48% of all breaches makes clear that patch debt is inherited as well as incurred [16].

The unavoidable conclusion is that no individual organization can patch its way out of this condition. When new disclosures arrive at 131 per day, when a third of exploited flaws are weaponized on or before disclosure, and when enrichment infrastructure is rationing its own output, the assumption that a well-resourced program can achieve comprehensive, timely remediation is no longer tenable. Patch debt is a property of the ecosystem, and it requires ecosystem-level responses in addition to organizational discipline.

Systemic Risk Implications

The combination of asymmetric velocity and accumulating patch debt produces a risk profile that is better understood through the lens of systemic risk than through the traditional model of organization-bounded cyber risk. Systemic risk is the possibility that a disturbance in one part of an interconnected system propagates through shared dependencies to threaten the system as a whole. The MOVEit and Log4Shell events demonstrated this property: a single flaw in a widely deployed component became, within hours to days, a simultaneous exposure across thousands of organizations that had no relationship with one another except a shared dependency [27][49]. As exploitation velocity rises and patch debt deepens, the conditions for such cascades become more common rather than less.

Three features distinguish this as a systemic rather than merely additive problem. First, concentration: network edge devices, identity providers, file-transfer appliances, and foundational libraries are deployed across enormous populations of organizations, so a single weaponized flaw in one of them creates correlated exposure at scale. The fact that 48% of the 90 zero-days Google tracked in 2025 targeted enterprise technologies broadly — with security and networking appliances representing roughly half of that group — is significant, because these are the very products organizations rely on for defense [50]. Second, velocity: when exploitation precedes patch availability and breakout times fall to minutes, the period during which collective defense can intervene before widespread compromise shrinks toward zero [4][19]. Third, supply chain transitivity: an organization's effective window of exposure now includes the patch debt of every vendor and dependency in its chain, which the 60% increase in third-party breaches makes concrete [16].

Critical infrastructure sectors face a sharpened version of this risk because their operational constraints make rapid patching especially difficult while their compromise carries consequences beyond the individual entity. The Mexican government breach, in which a single AI-augmented operator compromised nine agencies and accessed electoral infrastructure data, illustrates how a low-cost, AI-driven campaign can reach systemic targets that were once thought to require nation-state resources [37]. When the cost of a sophisticated campaign falls to hundreds of dollars and the operator headcount falls to one, the population of actors capable of causing systemic-scale harm expands considerably.

The compounding factor is that defensive capacity does not scale with the threat in the same way attacker capacity now does. Attacker capability scales with compute and model improvement, both of which are improving on steep curves, while defender remediation capacity scales with human, organizational, and operational constraints that improve slowly. The iTWire analysis of the Qualys data captured this divergence: closed vulnerability events rose 6.5-fold from 73 million in 2022 to 473 million in 2025, yet remediation outcomes worsened over the same period [51]. More work is being done than ever, and the system is still falling behind. This is the signature of a systemic problem: individually rational effort failing to produce collectively adequate outcomes. The implication for risk management is that organizations should model their exposure not only as a function of their own controls but as a function of the collective remediation health of the components and providers on which they depend.

Recommendations

Closing the gap created by asymmetric vulnerability velocity requires action on three time horizons, with the understanding that no single layer is sufficient and that the strategic layer is where systemic risk is actually addressed. The recommendations below move from triage discipline an organization can

implement immediately, through capability investments that take quarters to mature, to the collective-defense and policy posture that the systemic nature of the problem ultimately demands.

In the immediate term, the priority is to replace severity-driven patching with exploitation-driven triage anchored on confirmed in-the-wild activity. CISA KEV should be the non-negotiable floor: every KEV entry present in an organization's environment warrants treatment as an active incident rather than a routine ticket, given that roughly 65% of the catalog was exploited in the prior year and that the median organization faced 16 KEV CVEs to patch in 2025 [15]. Because KEV is incomplete relative to observed exploitation, organizations should supplement it with commercial exploitation intelligence to capture the gap of several hundred exploited CVEs not yet cataloged [47]. Patch service-level agreements for internet-facing and identity-adjacent systems should be compressed to a target measured in days rather than weeks, with explicit recognition that for a growing share of vulnerabilities, exploitation precedes patch availability and compensating controls — virtual patching, network segmentation, and exposure reduction — must carry the defensive load until a fix can be deployed [18][4]. Given that network edge devices are the most-targeted category, reducing the internet-facing attack surface is among the highest-leverage immediate actions available [26].

In the medium term, organizations should move beyond CVSS as the primary prioritization metric and adopt exploitation-informed scoring. CVSS measures theoretical severity and is a poor predictor of which vulnerabilities will actually be attacked; the NVD's retreat from universal enrichment means that many 2025 CVEs lack even a CVSS score, further undermining reliance on it [44]. Frameworks such as the Exploit Prediction Scoring System (EPSS) and direct integration of KEV and threat-intelligence feeds produce prioritization that tracks real attacker behavior. This is also the horizon in which defenders should adopt AI-assisted prioritization and remediation to begin closing the velocity gap on their own side. The same capabilities that accelerate attackers can accelerate defenders: AIxCC demonstrated autonomous patch generation in 45 minutes at modest cost, and tools such as Big Sleep show AI discovering and helping remediate flaws before exploitation [29][31]. Defenders should pilot these capabilities with human oversight, given the documented 37.7-45.6% semantic failure rate on AI-generated patches, and should account for the guardrail asymmetry that constrains defensive AI use more than offensive use when selecting tooling [30][40].

In the strategic term, the systemic nature of the problem requires investment beyond the organizational boundary. Threat-informed vulnerability management should be institutionalized so that remediation prioritization is continuously driven by current adversary behavior rather than static severity. Organizations should formalize software supply chain visibility through software bills of materials and continuous dependency monitoring, because inherited patch debt — now responsible for nearly half of breaches — cannot be managed without knowing what one depends on [16]. At the ecosystem level, the case for collective defense models grows stronger as individual remediation proves structurally insufficient: shared exploitation intelligence, coordinated disclosure that accounts for the negative time-

to-exploit reality, and sector-level mutual-defense arrangements distribute the burden that no single organization can carry. Policy mechanisms — sustained public funding for vulnerability enrichment infrastructure to address the NVD shortfall, procurement requirements that incentivize secure-by-design products, and expanded KEV-style authoritative exploitation signaling — are appropriate responses to a market failure in which the social cost of patch debt exceeds the private cost any single actor internalizes. The DBIR's own conclusion that foundational security principles remain the most effective defense should be read not as a counsel of complacency but as a reminder that disciplined execution of these layered fundamentals, at the velocity the threat now demands, is the achievable path forward [1].

CSA Resource Alignment

The Cloud Security Alliance has produced a body of work directly applicable to the conditions this paper describes, and organizations confronting asymmetric vulnerability velocity can use these resources as an implementation backbone. The most directly relevant is the AI Vulnerability Storm strategy briefing, a joint effort with SANS, OWASP GenAI, and [un]prompted that was built by more than 60 contributors and reviewed by over 250 CISOs; it provides a board-ready risk register mapped to the OWASP LLM Top 10, MITRE ATLAS, and NIST CSF 2.0, along with priority actions for AI-era vulnerability management [52]. CSA's companion analysis, The Collapsing Exploit Window: AI-Speed Vulnerability Weaponization, quantifies the same compression dynamics examined here and documents AI generation of working CVE exploits in 10 to 15 minutes at approximately $1 per attempt [53].

For control implementation, the CSA AI Controls Matrix (AICM) provides 243 control objectives across 18 domains, mapping to ISO 42001, ISO 27001, NIST AI RMF 1.0, and BSI AIC4, and addressing threat categories including Model Manipulation, Data Poisoning, and Loss of Governance [54]. As a superset of the Cloud Controls Matrix, the AICM is the appropriate default framework for organizations integrating AI-specific risk into existing vulnerability and threat-management programs; the CCM v4.1 Threat and Vulnerability Management domain remains the foundation for cloud workload remediation discipline [55]. For organizations defending against the agentic, autonomous attack patterns documented in the GTG-1002 and Mexican government cases, the MAESTRO seven-layer agentic AI threat modeling framework offers a structured method for reasoning about where autonomous adversary capability can be detected and disrupted [56].

These CSA resources are complementary to the external standards that should anchor any program. The NIST AI Risk Management Framework provides the Govern, Map, Measure, and Manage structure for AI-related risk, and the OWASP Top 10 for LLM Applications 2025 — with prompt injection retained as the top risk — addresses the specific failure modes of the AI systems both attackers and defenders now deploy [57][58]. Zero Trust principles and the CSA STAR program's assurance model further support the

supply chain visibility and least-privilege posture that the systemic risk analysis in this paper recommends. Together these frameworks allow an organization to translate the strategic recommendations above into auditable controls and to demonstrate, to boards and partners alike, that its vulnerability management program is calibrated to the velocity of the current threat.

Conclusion

The 2026 DBIR's inversion — exploitation overtaking credentials as the leading initial access vector for the first time in nineteen years — is the visible surface of a deeper structural change. Beneath it lies an asymmetry in velocity that has crossed a meaningful threshold: for a growing share of vulnerabilities, exploitation now begins before a patch exists, while organizational remediation has slowed and the backlog of known, unfixed flaws has grown nearly eightfold in four years. AI is the principal accelerant on the attacker side, not because it has invented new classes of attack, but because it has made the familiar ones faster, cheaper, and available to a far larger population of actors.

The stakes of inaction compound with each passing quarter. Every day that median patch times exceed the time-to-exploit, the window of exposure widens, and every concentrated dependency that carries unremediated debt becomes a potential vector for systemic cascade. The Mexican government breach and the multi-country FortiGate campaign demonstrate that systemic-scale harm is now achievable by single operators at trivial cost, which means the threshold for catastrophic events has fallen even as the population of capable actors has risen.

There remains a window for action, and it is defined by execution velocity rather than by any single new technology. Organizations that anchor triage on confirmed exploitation, compress remediation timelines for exposed and concentrated systems, adopt exploitation-informed prioritization and AI-assisted defense with appropriate oversight, and participate in collective defense and supply chain transparency can close the gap that asymmetric velocity has opened. The defensive fundamentals have not changed; the speed at which they must be executed has. The organizations and ecosystems that match the velocity of the threat will define whether the era of AI-accelerated exploitation becomes a manageable risk or a recurring source of systemic failure.

References

[1] Verizon Business. "Vulnerability Exploitation Top Breach Entry Point, 2026 Industry-Wide DBIR Finds." Verizon, May 19, 2026.

[2] Help Net Security. "Verizon DBIR: Vulnerability exploitation is the dominant initial access vector." Help Net Security, May 20, 2026.

[3] Technology.org. "Verizon's 2026 Breach Report: AI Shrinks Defense Time to Hours." Technology.org, May 20, 2026.

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

[5] Rapid7. "Rapid7 2026 Global Threat Landscape Report Shows Exploited High and Critical Severity Vulnerabilities Surged 105% as Attack Timelines Collapsed." Rapid7, March 18, 2026.

[6] VulnCheck. "VulnCheck State of Exploitation 2026." VulnCheck, 2026.

[7] Qualys Threat Research. "Inside the 2026 Verizon DBIR: What One Billion Records Revealed About Vulnerability Remediation." Qualys, May 19, 2026.

[8] Infosecurity Magazine. "Verizon DBIR Reveals 34% Jump in Vulnerability Exploitation." Infosecurity Magazine, 2025.

[9] Descope. "DBIR 2025 Analysis." Descope, 2025.

[10] Infosecurity Magazine. "DBIR: Vulnerability Exploits Triple as Initial Access Point." Infosecurity Magazine, 2024.

[11] Verizon Business. "2026 Data Breach Investigations Report." Verizon, May 19, 2026.

[12] Industrial Cyber / IBM X-Force. "IBM X-Force Reports 44% Surge in Exploitation of Public-Facing Applications as Supply Chain and Identity Attacks Intensify." Industrial Cyber, 2026.

[13] Darknet.org.uk. "Credential Stuffing in 2025: How Combolists, Infostealers and Account Takeover Became an Industry." Darknet.org.uk, March 2026.

[14] Security Boulevard. "The AI Governance Gap: Verizon's 2026 DBIR Shows Attackers Scaling AI While Employees Leak Data Through It." Security Boulevard, May 20, 2026.

[15] CyberScoop. "Attackers hit vulnerabilities hard last year, making exploits the top entry point for breaches." CyberScoop, May 19, 2026.

[16] GlobeNewswire / Verizon Business. "Vulnerability Exploitation Top Breach Entry Point, 2026 Industry-Wide DBIR Finds." GlobeNewswire, May 19, 2026.

[17] Flashpoint. "N-Day Vulnerability Trends: The Shrinking Window of Exposure and the Rise of Turn-Key Exploitation." Flashpoint, 2025.

[18] Qualys Threat Research Unit. "Qualys TRU Research Finds Manual Remediation Can't Keep Up As Exploitation Hits 'Negative One Day'." KBI.Media, 2025.

[19] CrowdStrike. "CrowdStrike 2026 Global Threat Report Findings." CrowdStrike, February 2026.

[20] Edgescan. "Edgescan 2025 Vulnerability Statistics Report." Edgescan, 2025.

[21] Hadrian. "2026 Offensive Security Benchmark Report." Hadrian, 2026.

[22] Indusface. "Key Vulnerability Statistics 2026." Indusface, 2026.

[23] Stingrai. "Vulnerability Statistics 2026: CVE, KEV, Time to Exploit." Stingrai, 2026.

[24] Trend Micro. "Old Vulnerabilities, New AI Era: How Outdated Flaws Continue to Fuel the N-Day Exploit Market." Trend Micro, 2025–2026.

[25] Security Boulevard. "46 Vulnerability Statistics 2026: Key Trends in Discovery, Exploitation, and Risk." Security Boulevard, March 2026.

[26] Outpost24. "Lessons From 2025: Zero-Day Exploitation Shaping 2026." Outpost24, 2026.

[27] CISA. "#StopRansomware: CL0P Ransomware Gang Exploits CVE-2023-34362 MOVEit Vulnerability (AA23-158A)." CISA, 2023.

[28] Tenable. "CVE-2024-55591: Fortinet Authentication Bypass Zero-Day Vulnerability Exploited in the Wild." Tenable, January 2025.

[29] DARPA. "AI Cyber Challenge Marks Pivotal Inflection Point for Cyber Defense." DARPA, August 2025.

[30] Academic consortium. "SoK: DARPA's AI Cyber Challenge (AIxCC): Competition Design, Architectures, and Lessons Learned." arXiv 2602.07666v2, February 2026.

[31] Security Online. "Google's Big Sleep AI Foils Live Zero-Day Exploit in SQLite (CVE-2025-6965)." Security Online, 2025.

[32] Schneier, Bruce. "AI Found Twelve New Vulnerabilities in OpenSSL." Schneier on Security, February 2026.

[33] Fang, Bindu, Gupta, Kang (UIUC). "LLM Agents Can Autonomously Exploit One-Day Vulnerabilities." arXiv 2404.08144, April 2024.

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

[35] The Hacker News. "2026: The Year of AI-Assisted Attacks." The Hacker News, May 2026.

[36] Anthropic. "Disrupting the First Reported AI-Orchestrated Cyber Espionage Campaign." Anthropic, November 2025.

[37] Check Point Research. "AI Attacks Are No Longer Experimental: Key Findings From the March-April 2026 AI Threat Landscape." Check Point, April-May 2026.

[38] The Hacker News. "Open-Source CyberStrikeAI Deployed in AI-Driven FortiGate Attacks Across 55 Countries." The Hacker News, March 2026.

[39] ProjectDiscovery. "From CVE to Template: The Future of Automating Nuclei Templates with AI." ProjectDiscovery, 2025.

[40] CSO Online. "When AI Safety Constrains Defenders More Than Attackers." CSO Online, 2025-2026.

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

[42] Help Net Security. "NIST Admits Defeat on NVD Backlog, Will Enrich Only Highest-Risk CVEs Going Forward." Help Net Security, April 16, 2026.

[43] Infosecurity Magazine. "NIST Drops NVD Enrichment for Pre-March 2026 Vulnerabilities." Infosecurity Magazine, April 2026.

[44] The Hacker News. "NIST Limits CVE Enrichment After 263% Surge in Vulnerability Submissions." The Hacker News, April 2026.

[45] Cyble. "2025 CISA KEV Catalog Hits 1,484 Exploited Vulnerabilities." Cyble, January 2026.

[46] The Cyber Express. "CISA Known Exploited Vulnerabilities (KEV) Soared 20% In 2025." The Cyber Express, January 2026.

[47] Tenable. "As the NVD Scales Back CVE Enrichment, Here's What Tenable Customers Need to Know." Tenable, April 2026.

[48] Cyber Strategy Institute. "2026 Vulnerability Report: 5 Critical Exploitation Trends." Cyber Strategy Institute, 2026.

[49] Dark Reading. "Vulnerability Exploitation Is Shifting in 2024-25." Dark Reading, 2025.

[50] Google / GTIG. "Look What You Made Us Patch: 2025 Zero-Days in Review." Google Cloud, March 2026.

[51] iTWire / Qualys. "Qualys TRU Research Finds Manual Remediation Can't Keep Up As Exploitation Hits 'Negative One Day'." iTWire, 2025.

[52] Cloud Security Alliance. "The AI Vulnerability Storm: CISO Security Program Guide." CSA / SANS / OWASP GenAI / [un]prompted, April 2026.

[53] Cloud Security Alliance AI Safety Initiative. "The Collapsing Exploit Window: AI-Speed Vulnerability Weaponization." CSA, April 2026.

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

[55] Cloud Security Alliance. "Cloud Controls Matrix v4.1." CSA, December 2025.

[56] Huang, Ken / Cloud Security Alliance. "Agentic AI Threat Modeling Framework: MAESTRO." CSA, February 2025.

[57] NIST. "Artificial Intelligence Risk Management Framework (AI RMF 1.0)." NIST, January 2023.

[58] OWASP Gen AI Security Project. "OWASP Top 10 for LLM Applications 2025." OWASP, 2024.

[59] Picus Security. "CVE-2024-55591: Fortinet FortiOS Authentication Bypass Vulnerability." Picus Security, January 2025.