HADRIAN

OFFENSIVE SECURITY BENCHMARK REPORT

2026

TABLE OF CONTENT

INTRODUCTION

When I talk to CISOs and security leaders across Europe and the US, I hear the same frustration in different words: we have more security data than ever, but less clarity on what actually matters. The number of dashboards has exploded. The number of tickets has exploded. The number of “critical” alerts has exploded. And yet the practical question remains painfully hard to answer on a Tuesday morning: what is exploitable right now, and what do we do about it before an attacker does?

This report exists because the gap between visibility and outcomes has become the defining security problem of 2026.

Attackers are not waiting for our quarterly patch cycles, our prioritization meetings, or our tool consolidation projects. They are operating with industrial speed, using automation and AI to scale recon, weaponize edge exposures, and chain weaknesses faster than most teams can validate them. The uncomfortable truth is that “known vulnerabilities” are no longer a reliable boundary between safe and unsafe. Exploitation increasingly starts before disclosure. And the time between exposure and compromise is shrinking.

At the same time, defenders are being buried under noise. Most findings do not represent real-world exploitability. That is not a minor inefficiency, it is a strategic failure mode.

If the majority of your effort goes into chasing theoretical risk, you are implicitly accepting the real risk. This is why you can be “busy” and still be losing.

AI intensifies this dynamic in two directions. First, it lowers the cost of offense. The barrier to entry for exploitation keeps dropping, and the scale at which attackers can probe and iterate keeps rising. Second, it changes how software is built. AI assisted development is now mainstream, and it is accelerating delivery, but it is also accelerating the introduction of insecure patterns and new AI native attack surfaces.

We built this analysis from verified exposure data collected by Hadrian’s offensive security platform, combined with firsthand insight from security leaders who are living this reality inside large, complex organizations. The through line is consistent: the winning security teams are shifting from volume to verification.

The organizations that adapt to this shift will not just be “more secure.” They will be more resilient, more predictable, and better positioned to move quickly in a world where AI accelerates everything, including the downside. Let’s get specific.

ROGIER FISCHER

CEO & CO-FOUNDER, HADRIAN

EXECUTIVE SUMMARY

This report examines how modern attacks unfold and how security teams respond in practice. The analysis combines real-world, exploit-verified exposure data collected by Hadrian's offensive security platform with firsthand insights from security leaders.

01

99.5% of detected risks don't matter

Despite unprecedented visibility, security teams can't act decisively. Only 0.47% of legacy vulnerability scanner findings prove exploitable, leaving teams buried in noise, unable to distinguish real-world threats from theoretical risk.

02

50% of AI-generated code is insecure, even when it "works"

AI has become the norm in software and attack workflows. While accelerating delivery, it introduces insecure code, new AI-native attack surfaces like MCPs, and dramatically lowers the skill barrier for scalable exploitation.

03

23% of all verified exposures still originate in DNS

"It's always DNS" seems to still be an industry wisdom. DNS is still the largest source of verified exposures. Often ignored as background infrastructure, DNS increasingly exposes misconfigurations that directly enable attacks on modern, multi-cloud environments.

04

70% of intrusion chains now begin with edge exploitation

Attackers are shifting decisively to mass exploitation of exposed services, APIs, and edge infrastructure. Rapid weaponization and AI acceleration leave little reaction time, while undocumented APIs and logic flaws remain largely invisible to traditional scanners.

05

32% of zero-day vulnerabilities are exploited before disclosure

Organizations believe they respond quickly, but exploitation increasingly precedes CVE disclosure. Median remediation still takes over a month, while attackers weaponize edge vulnerabilities immediately, turning public disclosure into a lagging indicator of active compromise.

06

Only 33% of CTEM programs measure whether exploitable risk is actually reduced

Security programs keep adding tools and scope, yet outcomes stagnate. Without automation, standardization, and exploitability verification, SecOps teams generate more findings but fail to translate effort into faster remediation or meaningful risk reduction.

The verification crisis

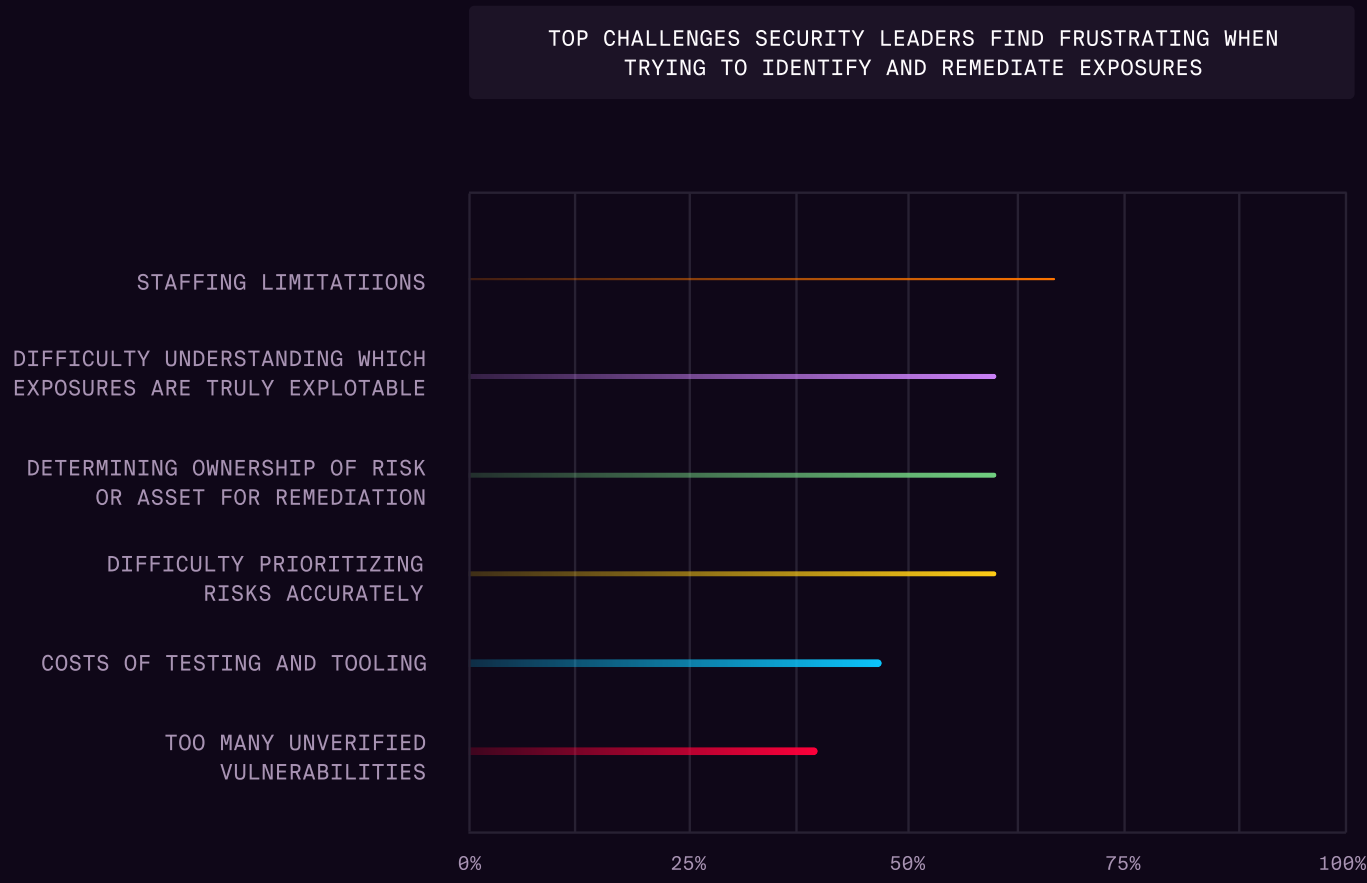
01

More visibility was supposed to make security decisions easier. Instead, it has exposed a deeper problem: teams can see more exposures than ever, but understand less of what actually matters.

A visibility paradox

Security teams have significantly expanded visibility across their environments, yet confidence in decision making has declined. In our study, 95% of security leaders report dissatisfaction with their ability to prioritize remediation based on real world exposure. Over 70% said they struggle to determine which exposures are actually exploitable, and nearly two thirds cite unverified vulnerabilities as their most frustrating operational challenge.

At the same time, satisfaction with asset discovery and exposure identification is relatively high. The gap emerges after discovery. As the volume of findings increases, teams lack reliable ways to determine which issues warrant immediate action, creating hesitation rather than faster response.

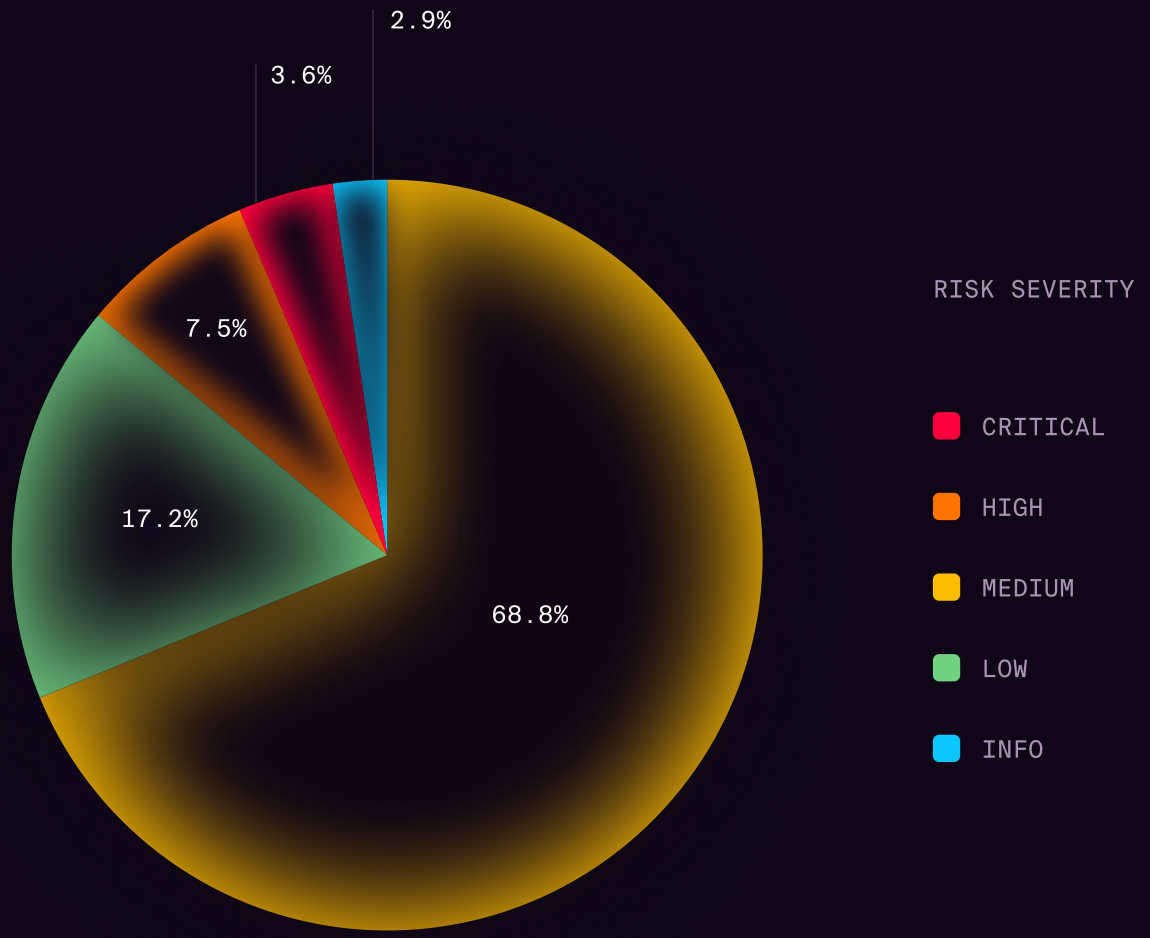


The noise problem

Security teams are not overwhelmed because they miss exposures. They are overwhelmed because almost none of what they see actually matters. On average, only 0.47% of risks detected by vulnerability scanners turn out to be real and require action. Teams spend their time triaging thousands of issues that appear urgent but lack real-world impact. What remains is a flood of findings that look serious on paper but collapse under scrutiny.

Nearly 90% of verified risks are classified as medium or low severity. High severity issues account for just 7%, and critical risks make up only 3% of total findings. As a result, teams face thousands of findings that appear significant but lack context. The volume of medium and low severity findings obscures the small set of exposures that pose immediate, real world danger.

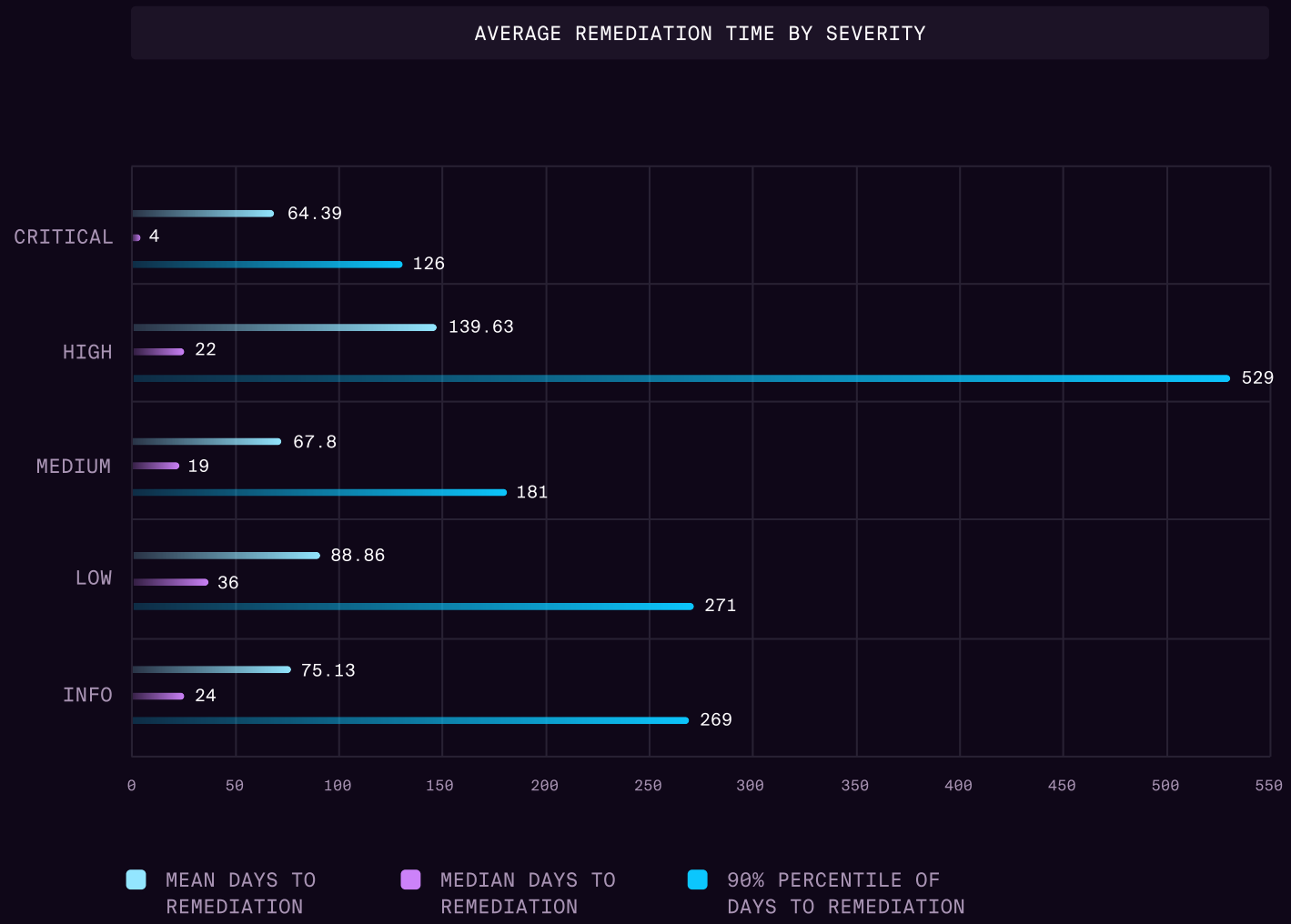
SEVERITY DISTRIBUTION OF VERIFIED EXPOSURES DISCOVERED IN THE ATTACK SURFACE



Remediation paralysis

Remediation data shows that speed is possible when the risk is clear. Critical exposures have a median remediation time of just four days. However, the average stretches to 64 days, and the slowest 10% take more than 120 days to resolve.

As observed last year, high severity risks take longer to remediate than less severe issues, with an average remediation time of 139 days. This is not because they matter less, but because they often require significantly more effort to fix than Medium or Low severity findings. At the same time, they lack the organizational urgency that a Critical rating creates, leaving teams caught between complexity and insufficient mandate.



AI on the rise

02

AI is accelerating both software delivery and attacker capability. The same systems that improve productivity are expanding the attack surface and lowering the barrier to exploitation.

AI coding introduces security risks

As AI becomes embedded in everyday software development, it is increasingly shaping the security posture of modern applications. A developer survey by GitHub found that more than 97% of developers have used AI coding tools at work, making AI-generated code a default input into production environments.

Benchmarking shows that functional correctness does not equate to security. Even when AI-generated code passes functional tests, roughly half of those outputs still contain security flaws. This holds true even for recent models that are explicitly instructed to avoid known vulnerability classes. The result is a growing volume of insecure code entering production despite appearing correct during development and testing.

MCPs create a new attack surface

The Model Context Protocol (MCP) is emerging as a foundational layer for agentic AI systems. It standardizes how large language models connect to external tools, APIs, and services, allowing AI agents to read files, query databases, invoke cloud services, and take real actions on behalf of users. This capability significantly expands what AI systems can do, but it also extends the attack surface beyond the model itself into the infrastructure those tools access.

In 2025, approximately 36.7% of Microsoft MarkItDown MCP servers, one of the most widely deployed MCP implementations, were found to be vulnerable to a severe server-side request forgery issue. As organizations adopt MCP at scale, weaknesses in these integrations risk becoming a direct and high-speed path from AI interaction to enterprise compromise.

Unrestricted AI enables new techniques

In its August 2025 threat intelligence report, Anthropic identified “vibe hacking” as an emerging threat pattern. The report documents how attackers used Claude to automate reconnaissance, network penetration, and data exfiltration. In one case, a single actor used AI-assisted workflows to target at least 17 organizations across healthcare, government, and emergency services.

At the same time, the AI model ecosystem continues to expand rapidly. Hugging Face now hosts more than 2.2 million models with more than 3,000 models explicitly labeled as uncensored and released without alignment constraints or safety controls.

These uncensored models impose no restrictions on output. They will generate exploit code, malware logic, phishing content, and procedural guidance for abuse without refusal or filtering. Unlike commercial models, they do not attempt to limit harmful use, making them effective tools for scaling offensive activity rather than isolated experimentation.

The percentage of security leaders concerned about AI-driven threats

67%

DNS: The forgotten attack surface

03

DNS is one of the oldest and most fundamental layers of internet infrastructure. It is also one of the most consistently mishandled, leading to severe consequences.

An early indicator of active threats

DNS issues should not be treated as background noise. Recent threat intelligence shows how heavily attackers rely on DNS as part of active campaigns. Between August and November 2025, researchers identified 7.6 million new threat-related domains.

This activity is not limited to unused or speculative domains. DNS-based malware discovered in October 2025 had compromised more than 30,000 legitimate websites, using domain infrastructure to pivot attacks by redirecting traffic, capturing credentials, and distributing malware. Furthermore, Global Cyber Alliance (GCA) found that securing DNS servers, can prevent more than 33% of cybersecurity data breaches from occurring.

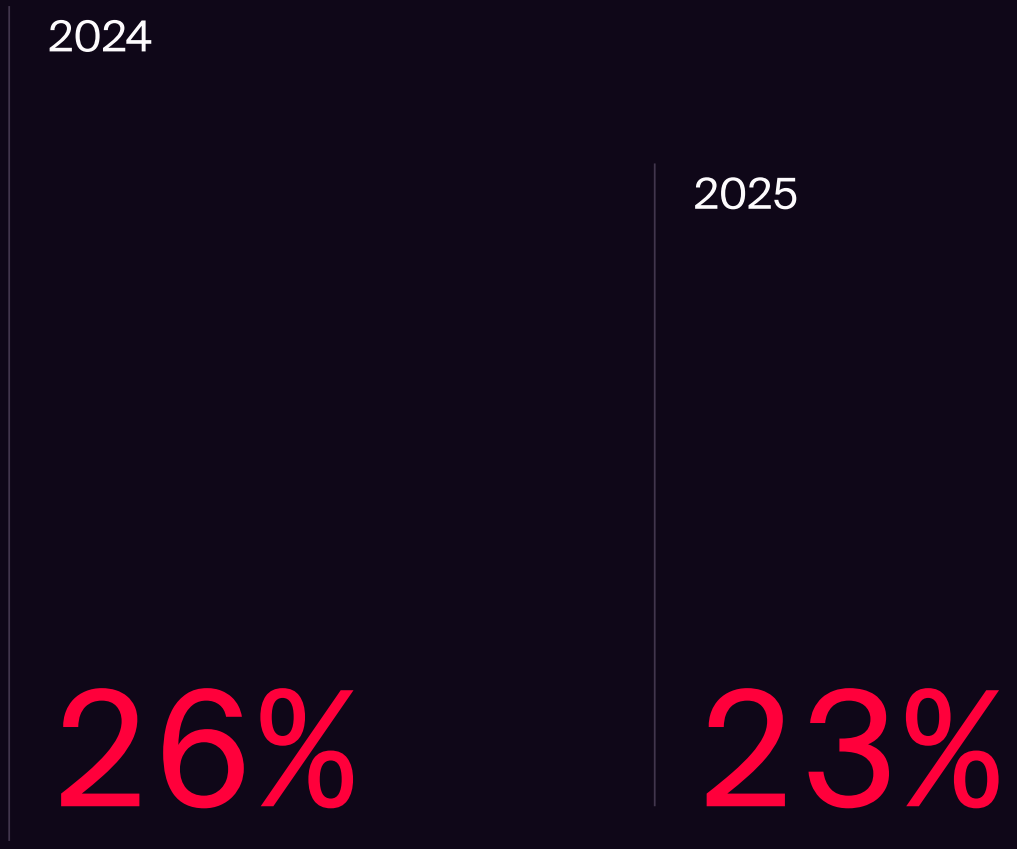
The percentage of data breaches can be prevented by securing DNS servers

33%

The largest source of verified exposure

DNS-related issues represent the single largest category of verified risk in the dataset. Across more than 300 organizations analyzed, DNS accounts for 23% of all verified findings, similar to last years 26%. This growth tracks closely with the adoption of multi-cloud infrastructure, increased reliance on third-party SaaS services, and the frequent provisioning and decommissioning of internet-facing resources.

For most organizations, DNS also represents one of the fastest opportunities for risk reduction. When identified and correctly triaged, DNS issues are typically quick to remediate and have recorded the lowest median remediation time observed by Hadrian for two consecutive years.



PERCENTAGE OF VERIFIED EXPOSURES DISCOVERED STEMMING FROM DNS ISSUES

Mass exploitation campaigns

04

Modern intrusions increasingly begin at the edge. Exposed services, APIs, and rapidly weaponized vulnerabilities now define how attackers gain initial access.

Targeting the edge

Exploitation of vulnerabilities emerged as the fastest-growing initial access vector, increasing 34% year over year. Attackers are no longer relying primarily on phishing or credential reuse. Instead, they are moving directly against exposed systems. Exploit Public-Facing Application (T1190) appeared in more than 70% of intrusion chains where vulnerability exploitation was observed, making it the dominant technique in these campaigns.

The shift is especially pronounced at the network edge. Exploitation of edge devices and VPNs rose from 3 percent to 22%, as attackers increasingly target externally exposed infrastructure and weaponize vulnerabilities before patches are applied, leaving defenders little time to react.

APIs are a blindspot

Most API risk originates from design and authorization logic rather than traditional code defects. Vulnerability scanners are effective at identifying injection flaws and known weaknesses, but they are poorly suited to detecting logic abuse that depends on how an application is intended to function.

Across surveyed organizations, only 30% of APIs are fully documented, which further limits testing coverage and expands blind spots. Undocumented endpoints are rarely tested and often excluded from security reviews. As a result, many high impact API exposures remain invisible to automated scanning. Business logic flaws are typically uncovered only through adversarial testing that exercises real user behavior across roles and permission boundaries.

The percentage of increase in OWASP API Security Top 10 related incidents – Akamai

32%

The percentage of companies report that regular issues third-party APIs – Lunar.dev

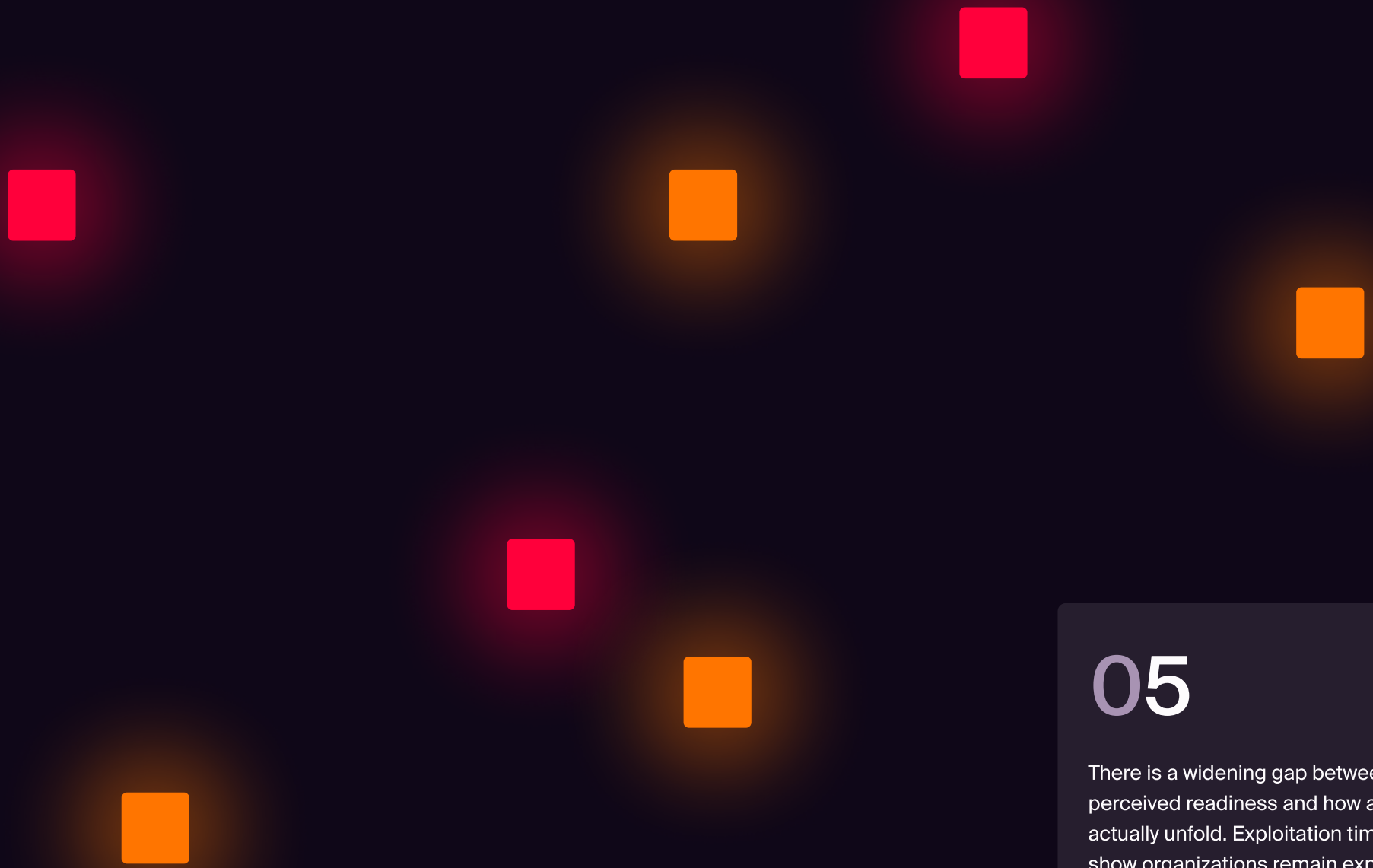
88%

AI will drive more exploitation

Responding to AI-driven threats has become the top concern for enterprise security leaders, cited by approximately 67% of respondents. At Hadrian, advances such as Subwiz demonstrate how automation can continuously surface previously unknown and short-lived internet-facing assets that attackers increasingly target.

This shift is reflected in recent research from Anthropic. Claude Sonnet 4.5 more than doubled its success rate on Cybench CTF challenges compared with earlier versions and completed complex, multi-step tasks such as traffic analysis and malware decompilation significantly faster than a skilled human. As these capabilities mature and proliferate, AI is expected to play an increasing role in accelerating and scaling data breaches.

Rapid weaponization



05

There is a widening gap between perceived readiness and how attacks actually unfold. Exploitation timelines show organizations remain exposed far longer than they believe.

The lies we tell ourselves

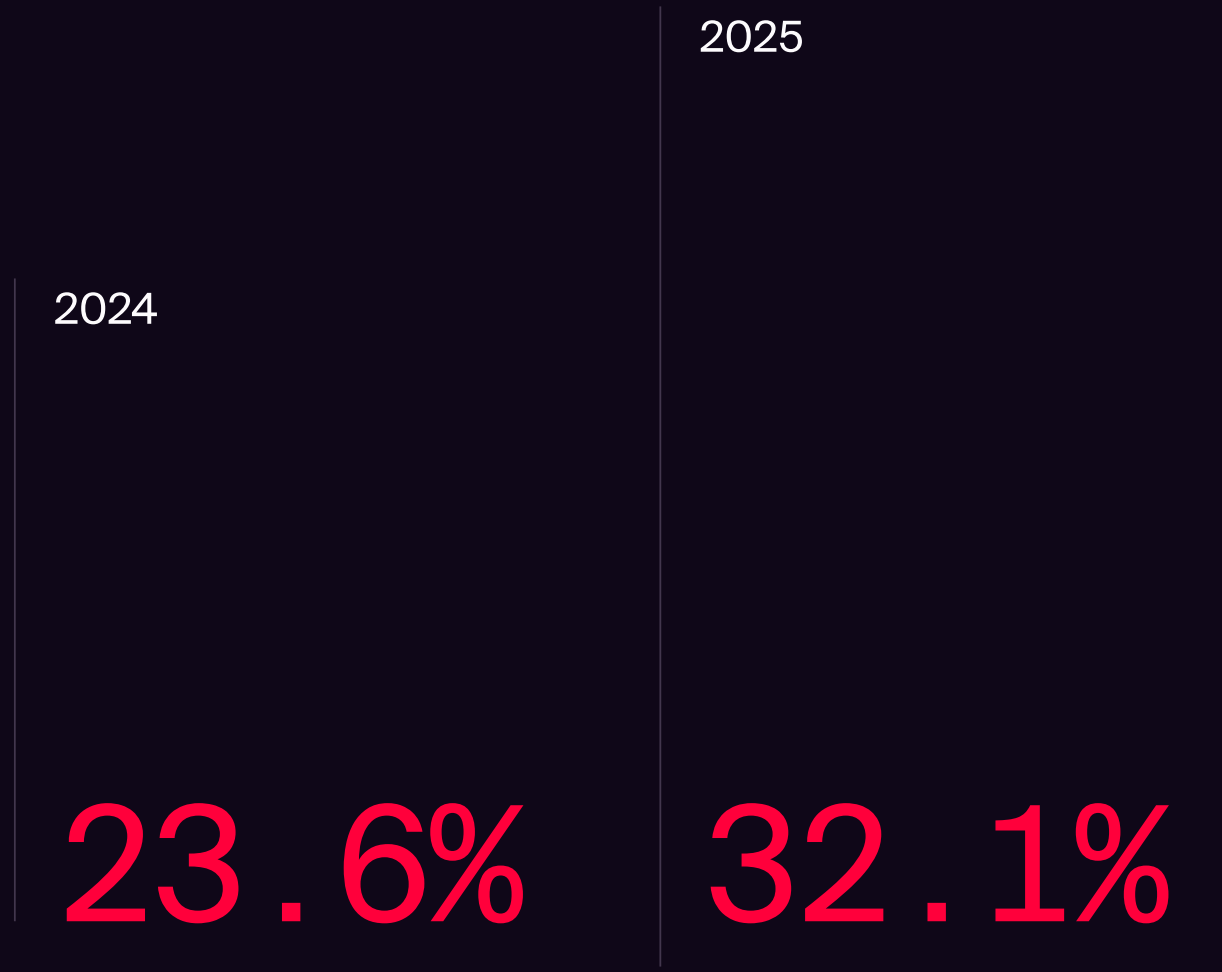
Zero day response timelines highlight a gap between perceived and actual performance. Ninety four percent of security teams report remediating zero day vulnerabilities within five days, with 47% claiming resolution in one to two days and 7% reporting same day remediation.

Telemetry tells a different story. According to Verizon, vulnerabilities listed in the Known Exploited Vulnerabilities catalog have a median time to full remediation of approximately 38 days after disclosure. This contrast suggests that confidence in response speed often outpaces what occurs in practice, particularly outside tightly defined zero day scenarios.

Exploitation leads disclosure

For the first time, zero-day exploits targeting edge devices were, on average, exploited before they were publicly reported. In 2025, 32.1 percent of Known Exploited Vulnerabilities showed evidence of exploitation on or before the day the CVE was issued, up from 23.6% in 2024.

This shift is driven primarily by internet-facing technologies such as VPNs, firewalls, and gateways, which are continuously scanned and attacked at scale. For these edge vulnerabilities, CVE publication no longer marks the beginning of exploitation. It increasingly reflects activity that is already underway, making disclosure a lagging indicator rather than an early warning signal.



VULNERABILITIES WITH ACTIVE EXPLOITATION OBSERVED BEFORE KEV CATALOG INCLUSION

— Vulncheck

CISA is resetting remediation expectations

CISA has increasingly used its federal authority to impose remediation timelines that are far more aggressive than traditional industry practices.

This shift became especially clear in September 2025, when CISA issued Emergency Directive 25-03 in response to actively exploited Cisco zero-day vulnerabilities. The directive required U.S. federal civilian agencies to identify, mitigate, and patch affected systems within 24 hours. This marked a sharp acceleration from prior federal expectations, where agencies were generally required to remediate critical exposures within 15 days and high-severity issues within 30 days of detection.

The structural challenge

06

Security programs continue to expand in scope and tooling, yet operational outcomes remain flat. The data shows that without automation, standardization, and verification, SecOps cannot translate effort into progress.

Why SecOps can’t break the cycle

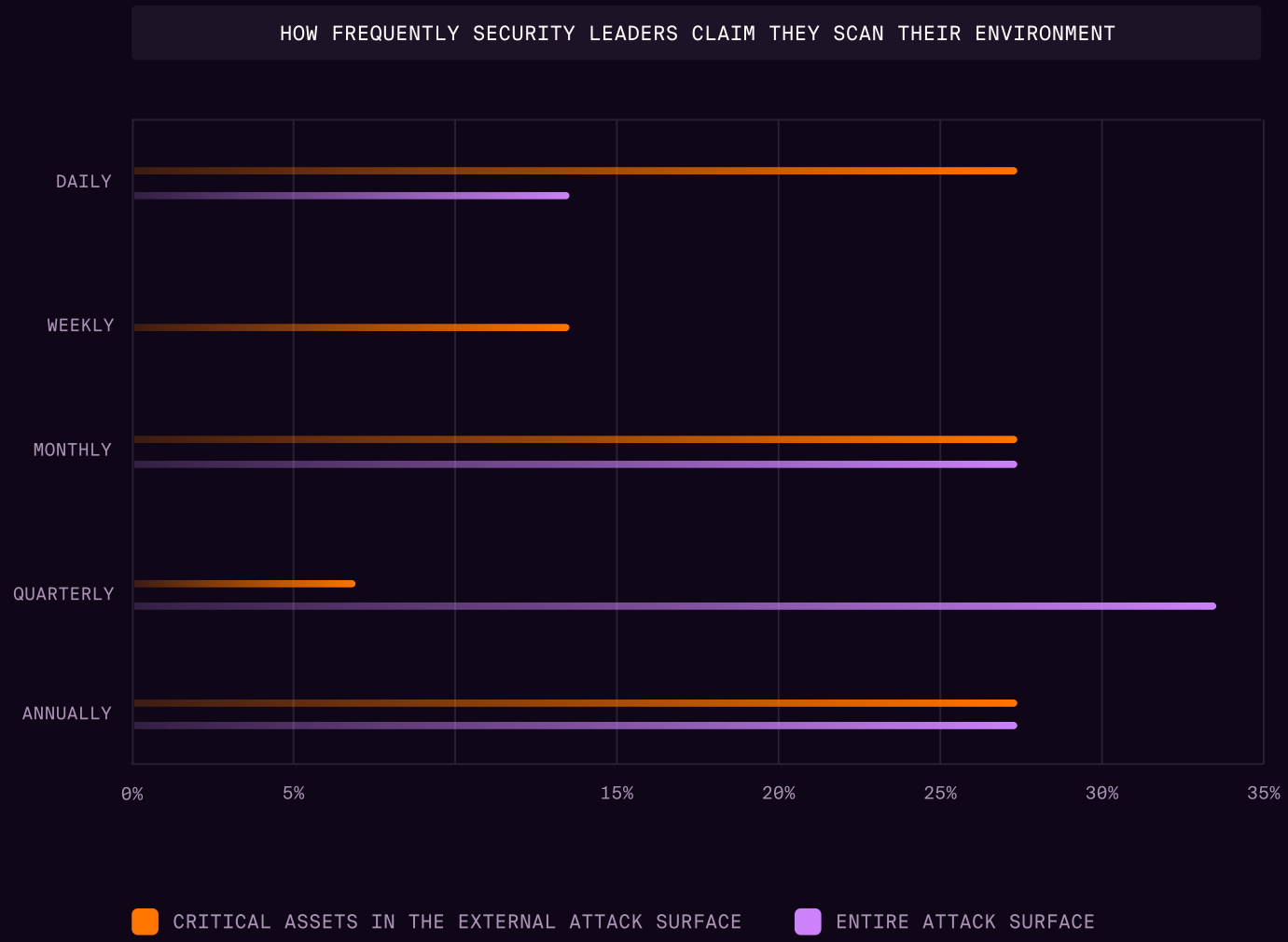
Most organizations have invested heavily in security tooling, yet outcomes have not improved proportionally. Ninety three percent of organizations use vulnerability scanners, 87% conduct manual penetration testing, and 73% operate internal red teams. Despite this, only 40% have adopted automated penetration testing. High tool adoption has not translated into faster remediation or clearer prioritization.

This imbalance contributes directly to alert overload. Scanners generate large volumes of findings, while validation remains manual and slow. Teams spend more time managing outputs than reducing risk. The data shows that tooling expansion without automation and verification increases operational friction rather than resilience.

No standard for “continuous”

Testing frequency varies widely across organizations, even for critical assets. Twenty seven percent test critical assets daily, while an equal percentage test them annually. Monthly testing accounts for another 27%, with weekly and quarterly testing representing smaller segments. This variation reflects inconsistent definitions of acceptable risk rather than intentional strategy.

For entire external attack surfaces, quarterly testing is most common at 33%, followed by monthly and annual testing at 27% each. These gaps create long windows of unvalidated exposure. The data indicates that most organizations test on schedules driven by resource constraints, not by attacker behavior or asset criticality.



Programs measure the wrong thing

Continuous Threat Exposure Management (CTEM) programs are widely discussed, but measurement remains skewed toward discovery. Sixty seven percent of organizations measure CTEM success based on coverage gaps identified, while only 33% track reductions in exploitable exposures over time. Mean time to remediation is measured by 53%, but rarely tied to validation quality.

As a result, CTEM programs often surface more findings without improving decision making. Teams become better at knowing what exists, but not at determining what matters most. Without exploitability validation and ownership alignment, CTEM reinforces discovery without driving action. The data shows that reactivity persists because verification remains incomplete.

RECOMMENDATIONS

Modern attacks move faster than traditional security processes. Closing the gap requires security programs to move beyond legacy approaches and continuously validate real-world exposure.

01

Shift from visibility-first to verification-first security. Prioritize controls that validate exploitability in real-world conditions, not just presence of vulnerabilities. Integrate attack path analysis, exploitation testing, and adversary-context signals to reduce noise, create urgency, and ensure remediation effort aligns with actual business risk exposure.

02

Treat AI as a production dependency, not a developer convenience. Enforce security validation on AI-generated code, assess MCP integrations as high-risk infrastructure, and continuously test AI-exposed attack paths. Security programs must assume attackers use unrestricted AI at scale and validate exposure accordingly.

03

Elevate DNS to a first-class security control. Continuously monitor for malicious domains, dangling records, and misconfigurations tied to cloud and SaaS assets. Prioritize rapid triage and cleanup of DNS findings, as they offer one of the fastest, highest-impact opportunities for reducing real-world attack paths.

04

Continuously inventory and test all internet-facing assets, including ephemeral edge services and APIs. Prioritize exposure-based testing that validates exploitability, not just patch status. Treat APIs as adversarial interfaces by documenting endpoints, enforcing authorization testing, and regularly simulating real attacker behavior at the edge.

05

Prioritize continuous exposure validation on internet-facing systems, especially edge devices, rather than relying on CVE alerts. Align remediation processes with attacker timelines and emerging regulatory expectations by pre-authorizing rapid response, temporary mitigations, and automated verification of exploitability.

06

Refocus SecOps on measurable exposure reduction. Standardize continuous testing based on asset criticality and attacker behavior, automate exploitability validation, and track outcomes tied to verified risk removal. Programs should be judged by how much exploitable exposure is eliminated, not how much activity or coverage they generate.

ABOUT HADRIAN

Hadrian is an offensive security platform built for teams that need proof, not predictions. We reveal exactly how an adversary could break in today by executing real attacker techniques against your external attack surface. By validating exploitability in production-safe ways, Hadrian eliminates false urgency, sharpens prioritization, and gives teams confidence to act.

Deployed in minutes and engineered for scale, Hadrian seamlessly fits into existing security workflows. Security leaders use Hadrian to move from periodic testing to continuous, evidence-backed exposure reduction.

RECOGNISED BY LEADING ANALYSTS

Gartner

GIGAOM

FROST & SULLIVAN

TRUSTED BY MARKET LEADERS

NBC

amadeus

McKESSON

BLINQX

CRÉDIT AGRICOLE

SHV ENERGY

ABN·AMRO

London Business School

RITUALS...

SIEMENS energy

LOTTOMATICA

LEROY MERLIN

WeatherTech

BIOLANDES

BLINQX

DAMEN

=EXACT

nedap

METHODOLOGY

This report is based on a combination of verified risk data collected throughout the 2025 calendar year and quantitative survey research conducted with senior security leaders.

VERIFIED RISK ANALYSIS

The quantitative risk data analyzed in this report was generated by Hadrian's agentic AI offensive security platform across more than 300 organizations. These organizations span multiple regions, including the United States, United Kingdom, the Netherlands, Germany, France, and Italy, and represent a broad range of industries such as manufacturing, financial services, healthcare, government, retail, and software.

Only verified, exploit-tested risks were included in the analysis. Findings were validated in real environments to confirm exploitability, exposure, and impact. Severity ratings reflect contextual factors including internet exposure, asset criticality, ease of exploitation, and potential for lateral movement, rather than CVSS scores alone.

SECURITY LEADER SURVEY

Survey data was collected in autumn 2025 from 34 CISOs and senior SecOps leaders. Respondents hold roles including CISO, Head of Information Security, Senior Director of Security, and SOC leadership. The survey focused on operational challenges, tooling adoption, testing cadence, remediation timelines, and preparedness for emerging threats in 2026.

HADRIAN.IO