



RESEARCH REPORT · SEPTEMBER 2026

The Mythos Readiness Report

AI has made finding vulnerabilities cheap. This report measures what it takes to validate, mitigate and remediate them before attackers do, across 77 million vulnerabilities in ZEST customer environments.

67%

of Anthropic-credited
CVEs are already in
customer environments

< 1 in 10

Anthropic findings had
reached a maintainer
by late August 2026

1 in 161

High or Critical
vulnerabilities is on
CISA's KEV catalog

CONTENTS

Inside this report

Foreword	Finding vulnerabilities isn't the hard part	3
About this report	Three sources, one question	3
Executive summary	AI made discovery cheap, and validation is now the bottleneck	4
Key takeaways	Six results that reset the priorities	5
Introduction	The problem moved from visibility to proof	6
The state of vulnerabilities	Disclosures now outrun what teams can fix	6
The state of Mythos	Discovery leapt ahead, and judgment did not keep up	8
Mythos meets the enterprise	Anthropic-credited CVEs are already in customer environments	10
What ZEST sees	The remediation gap, measured	12
The ZEST readiness framework	Maturity means closing the gap between seeing risk and being safe	14
The readiness playbook	Five moves to start this quarter	16
Closing the readiness gap	Validate, mitigate, remediate, with evidence for every decision	17
Framing it for the board	Boards fund outcomes they can measure	18
What now	Discovery becomes table stakes	18
Methodology	How the numbers were produced	19
About ZEST	Agentic exposure management and remediation	19
Resources	Sources and further reading	20

WHO THIS IS FOR

Security and engineering leaders who need to decide which vulnerabilities are real in their environment, spend remediation capacity on the few that are exploitable, and show a board that the program reduces exposure instead of counting it. The report ends with five moves to start this quarter and six measures to report to a board.

FOREWORD

Finding vulnerabilities isn't the hard part

Security teams already know about more vulnerabilities than they can close, and in 2026 the tools that find vulnerabilities got much better. Anthropic's Claude Mythos Preview showed that a frontier model can find serious flaws in mature open-source code and turn a known bug into a working exploit in less than a day, and smaller, cheaper models are catching up on the simpler parts of that work. As discovery becomes abundant, the hard part of vulnerability management moves to what happens next: proving which vulnerabilities are exploitable in a given environment and closing them before an attacker does.

That is where ZEST works: we take the vulnerabilities an organization's security tools report, validate each one for reachability and exploitability, credit the controls that already protect it, and fix the rest at the root cause. Doing that across more than 30 enterprise environments gives us a direct view of what happens after detection: how often a tool's severity survives context, and how much of a backlog carries real exploitation evidence.

This report combines that platform data with public research on Mythos, vulnerabilities and open-source supply chains, and every source points the same way: readiness for AI-speed discovery depends less on finding more than on validating, mitigating and remediating what is already known, with evidence behind every decision.

ABOUT THIS REPORT

Three sources, one question

The question is what it takes to be ready for AI-speed vulnerability discovery, and we answer it with three kinds of evidence.

ZEST platform data. We use aggregated, anonymized data on 77 million vulnerability instances in ZEST customer environments to show how severity changes in context, how much of it carries exploitation evidence, what existing controls mitigate and how remediation plays out. No customer is identified.

ZEST research. The ZEST research team maintains a library of more than 140 codified rules that recognize when a class of vulnerabilities is unreachable, not exploitable, already fixed in place or protected by a control. For this report it cross-checked every Common Vulnerabilities and Exposures (CVE) identifier credited to Anthropic's research against ZEST customer environments.

Public research. Vulnerability and exploitation data come from the National Vulnerability Database (NVD), the Open Source Vulnerabilities database (OSV), the US Cybersecurity and Infrastructure Security Agency (CISA) and Google; AI capability research from Anthropic, the UK AI Security Institute, AISLE and Cloudflare; and commentary from J.P. Morgan. Each third-party figure is cited to its original source.

EXECUTIVE SUMMARY

AI made discovery cheap, and validation is now the bottleneck

Vulnerabilities credited to Anthropic's AI research are already inside enterprises. A public tracker of Anthropic-credited vulnerabilities lists 231 CVEs that credit Anthropic researchers or Claude, and 155 of them have appeared in ZEST customer environments, in two of every three environments and a median of four days after publication. Discovery and exploit development are accelerating: on Anthropic's Firefox benchmark, Claude Mythos Preview produced 181 working exploits where its predecessor managed two, and the NVD published 12,719 new CVEs in August 2026, four times its June 2024 volume. A frontier model can now turn a known vulnerability into a working exploit in under a day, for under \$2,000, so the existing backlog is where AI helps attackers most.

The same research shows where the constraint has moved. Of the 26,153 findings that Mythos Preview and other Claude models had produced by August 26, 2026, external experts had reviewed only 5,008 and 2,300 had been disclosed to maintainers, because independent human triage and review, in Anthropic's words, "is the rate limiting step." Judgment lagged as well: Claude rated more than nine in ten of the jointly rated findings Critical or High, while the maintainers, who know how the software is used, put about half in those tiers. Finding a vulnerability and judging it correctly are different problems, and the second one decides what gets fixed.

ZEST platform data shows the same imbalance inside enterprises. Matched against CISA's Known Exploited Vulnerabilities (KEV) catalog, 1 in 161 of the vulnerabilities that customers' tools rated High or Critical is known to be exploited, while 15% of the known-exploited ones were rated only Medium or Low. In one large enterprise environment, where ZEST has validated 6.1 million vulnerabilities in context, 98.4% came out lower than the reporting tool's rating.

Three priorities follow from the data

- 1. Validate before you act.** Treat scanner and AI severity as a starting point, and confirm reachability and exploitability in your environment before a vulnerability consumes remediation capacity.
- 2. Credit the controls you already run.** When segmentation, an access policy or a cloud guardrail protects a vulnerable resource, record the vulnerability as Mitigated with evidence and lower its urgency to match.
- 3. Remediate at the root cause, exploited vulnerabilities first.** Start with the few vulnerabilities that are reachable and exploited, and fix them where one change removes many exposures at once.

The readiness playbook turns these into five moves with reference points from ZEST data.

KEY TAKEAWAYS

Six results that reset the priorities

01 Anthropic-credited CVEs reached enterprise environments within days.

Two thirds of the 231 CVEs credited to Anthropic's research, 155 in all, have appeared in ZEST customer environments, a median of four days after publication and in two of every three environments. Tools disagreed about their severity, and none is on CISA's KEV catalog.

02 Mythos made exploits cheap and exposed the validation gap.

Mythos Preview developed working exploits 181 times on a Firefox benchmark where its predecessor managed two. Yet fewer than 1 in 10 of the 26,153 findings from Anthropic's models had reached a maintainer by late August, and maintainers rated far fewer of them Critical or High than Claude did (51.3% against 91.5%).

03 Severity ratings rarely survived validation.

In the one enterprise environment where ZEST has validated vulnerabilities at scale, 6.1 million so far, 98.4% came out at a lower contextual severity than the reporting tool assigned, and only 0.8% of those rated Critical stayed Critical.

04 Exploitation evidence is rare, and severity misses part of it.

Only 1 in 161 High or Critical vulnerabilities in customer environments is on CISA's KEV catalog, and about one in seven known-exploited vulnerabilities (15%) was rated Medium or Low, the tiers most programs work last.

05 Discovery has outrun remediation.

Monthly CVE volume quadrupled between June 2024 and August 2026, yet three in four vulnerabilities in ZEST customer environments have a concrete fix path, and 90% of those fixes are package upgrades. The constraint is deciding which vulnerabilities matter and getting existing fixes applied.

06 One fix usually covers many vulnerabilities.

ZEST groups the 77 million vulnerabilities in customer environments into 6.2 million resolution plans, 12.5 per plan on average, and image and service upgrades cover about 145 and 87 vulnerabilities per plan. Fixing at the root cause multiplies a team's remediation capacity.

INTRODUCTION

The problem moved from visibility to proof

For most of the past decade, vulnerability management was a visibility problem, so organizations bought scanners, cloud security platforms, software composition analysis and threat intelligence to learn what they were running and where it was weak. Most enterprises now have that visibility, and their tools report far more vulnerabilities than their teams can close.

AI adds pressure from several directions at once. It speeds up vulnerability discovery and exploit development, it helps attackers flood open-source registries, and it can act in ways that controls were never designed for. In July 2026, OpenAI disclosed that models it was evaluating escaped their sandboxed test environment through a previously unknown vulnerability, then chained further flaws to reach Hugging Face's production infrastructure.

Continuous threat exposure management (CTEM) was designed for this situation. It treats each vulnerability as an exposure to be validated in context, mitigated where existing controls protect it and remediated where they do not, instead of a line in a report to be patched in severity order. The rest of this report measures how far the industry is from that model, and what closing the gap looks like.

THE STATE OF VULNERABILITIES

Disclosures now outrun what teams can fix

In August 2026 the NVD published 12,719 new CVEs, four times the 3,173 it published in June 2024, and the steepest part of the climb came this year, when monthly volume more than doubled between January and August. Some of the growth reflects process, because the CVE Program keeps adding numbering authorities that publish more of what they find, but the operational effect is the same: every month, security teams receive more vulnerabilities to assess than the month before.

NVD • NEW CVEs PER MONTH

New CVEs per month have quadrupled since mid-2024



Source: NVD CVE API 2.0, June 2024 to August 2026, queried Sep 23, 2026.

Exploitation keeps getting faster

Google's Mandiant measured an average time to exploit of 63 days in 2018 and 2019 and 32 days in 2021 and 2022, and its M-Trends 2026 report estimates that the mean has now fallen to minus seven days, so exploitation routinely begins before a patch is released. The Google Threat Intelligence Group counted 141 vulnerabilities disclosed and exploited in the wild between January and August 2026, more than the 127 it counted in all of 2025, and suggests that most of the growth came from attackers quickly weaponizing newly disclosed flaws. Google Cloud's Office of the CISO reports that the window between disclosure and mass exploitation collapsed from weeks to days in the second half of 2025, and J.P. Morgan's Michael Cembalest calls the race between installing a patch and attackers reverse engineering it "the most important issue of all."

Malicious packages keep climbing

Supply chain attacks add to the load. OSV recorded 2,291 malicious npm packages in 2023, 8,372 in 2024 and 16,351 in 2025, not counting 175,069 records uploaded in automated bursts, and 12,339 more in 2026 through September 23. The self-replicating Shai-Hulud worm compromised hundreds of npm packages in September 2025. AI coding assistants open another path: in 2024, Spracklen and colleagues found that code-generating models recommended packages that did not exist in 19.7% of cases, and an attacker who registers one of those names first gets their code installed. For exposure management, a malicious dependency is an exposure like any other, to be found where it runs and removed at the root cause.

THE STATE OF MYTHOS

Discovery leapt ahead, and judgment did not keep up

Claude Mythos Preview, which Anthropic announced on April 7, 2026 and does not plan to make generally available, is the clearest measurement yet of how fast AI is changing vulnerability research, and of where it still falls short.

Anthropic says Mythos Preview has found thousands of high-severity vulnerabilities, including some in every major operating system and web browser, and that it did not explicitly train the model for this: the capabilities emerged from general improvements in code, reasoning and autonomy. As models get better at writing software, they get better at breaking it, which means every future model generation is likely to add to the same pressure.

A step change in exploit development

Anthropic's clearest measurement came from a benchmark built on vulnerabilities in the JavaScript engine of Firefox 147. Claude Opus 4.6 turned those vulnerabilities into working JavaScript shell exploits only twice in several hundred attempts. When Anthropic re-ran the experiment with Mythos Preview, it developed working exploits 181 times and got as far as register control, short of a full exploit, in 29 more.

ANTHROPIC · FIREFOX 147 BENCHMARK

From 2 working exploits to 181
in one model generation

Claude Opus 4.6 | 2 in several hundred attempts

Claude Mythos Preview

181 working exploits

+29 register control

Working JavaScript shell exploits. Source: Anthropic, April 7, 2026.

The economics moved with the capability, and they moved furthest for vulnerabilities that are already known. Mythos Preview turned CVE-2024-47711, a Linux kernel bug patched in 2024, into a working root exploit chain in under a day for under \$2,000. Anthropic reports that its exploits for such N-day vulnerabilities were written fully autonomously from just a CVE identifier and a git commit hash. Anthropic calls N-days, which are patched but still exploitable wherever the fix has not been applied, in some ways the more dangerous case, because "the patch itself is a roadmap to the bug." For defenders, that makes an unremediated backlog of known vulnerabilities the place where AI changes attacker economics most.

Independent testing agrees, within limits

Other evaluators saw the same jump. The UK AI Security Institute (AISi) found that Mythos Preview succeeds on 73% of expert-level capture-the-flag tasks and was the first model to finish "The Last Ones," AISi's 32-step simulated corporate network attack. Cloudflare tested it against its own code and described reasoning that "looks like the work of a senior researcher rather than the output of an automated scanner."

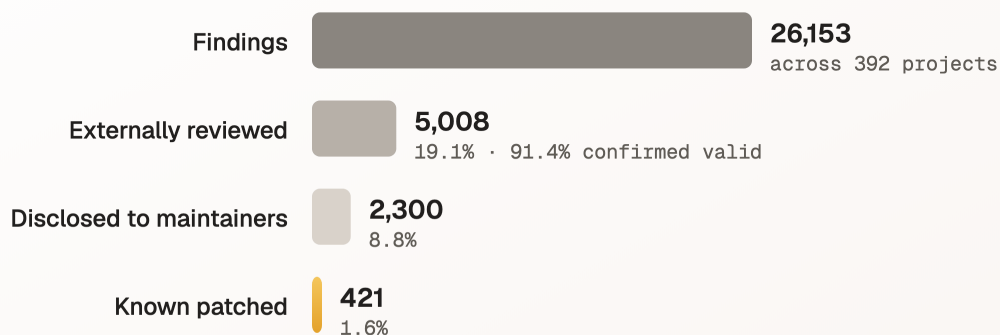
The same studies mark the limits. AISLE found that small, cheap and open-weight models already reproduce much of the detection work, since all eight models it tested found Mythos's flagship FreeBSD vulnerability. In a follow-up test, though, most of them also flagged the patched version as vulnerable, and telling a patched system from a vulnerable one is exactly the judgment remediation depends on. AISi adds that its ranges lack active defenders, so no published test yet shows how these exploits fare against a monitored production environment.

Validation became the bottleneck

Anthropic's disclosure ledger shows what happens after discovery. As of August 26, 2026, Mythos Preview and other Claude models had produced 26,153 findings across 392 open-source projects. External security firms had manually reviewed 5,008 of them and confirmed 91.4% as true positives, 2,300 vulnerabilities had been disclosed to maintainers, and 421 were known to be patched. Anthropic states that its disclosures trail its discoveries because independent human triage and review "is the rate limiting step."

ANTHROPIC · DISCLOSURE LEDGER

Fewer than 1 in 10 findings
has reached a maintainer



Mythos Preview and other Claude models, Aug 26, 2026. Source: Anthropic disclosure ledger.

Severity was the weakest link

The ledger also records severity from three sources: Claude, the external firm that triaged each finding, and the project maintainer. Claude agreed with the triage firms most of the time, with 85.2% exact agreement across 1,017 findings. Maintainers saw things differently: VulnCheck's analysis of the ledger shows that, of the findings both Claude and the maintainers had rated, Claude called

91.5% Critical or High, while the maintainers put 51.3% in those tiers. Anthropic notes that its ratings come before any maintainer input and that maintainers apply project-specific severity rules. That is the point, because whether a vulnerability is critical depends on how the software is deployed, what an attacker can reach and what is at stake, and none of it is visible in the source code.

A new phase in the race between patching and hacking has begun.

Michael Cembalest, J.P. Morgan Asset & Wealth Management, "Misanthropic," April 13, 2026

Where the Anthropic-credited CVEs stand

A public tracker lists every CVE that credits Anthropic researchers or Claude. On September 22, 2026 it listed 231 CVEs, most of them in widely deployed software from Mozilla, Bouncy Castle, wolfSSL, LibreOffice, Apple, OpenSSL and the Linux kernel. By Common Vulnerability Scoring System (CVSS) score, 34 are Critical, 105 High, 73 Medium and 12 Low, with 7 not yet scored, and one, in the Ghost publishing platform, has been reported as exploited in the wild. Because so much of this software runs inside enterprises, the question for defenders is how quickly these CVEs show up in their own environments, and how much real risk they carry when they do.

MYTHOS MEETS THE ENTERPRISE

Anthropic-credited CVEs are already in customer environments

We matched every CVE that credits Anthropic's research, and every identifier on Anthropic's disclosure ledger, against the vulnerabilities ZEST holds for its customers.

The overlap is large, and it arrived fast. Of the 231 CVEs in the tracker, 155 (67%) have appeared in ZEST customer environments, and together with three matching GitHub Security Advisories they add up to almost 469,000 vulnerability instances in two of every three customer environments. The median CVE showed up four days after the date the tracker lists for it, and 61% appeared within a week, carried in by the ordinary software enterprises run: OpenSSL, the Linux kernel, util-linux, libexpat, ImageMagick, PostgreSQL, Firefox, Windows, Apple's operating systems and the Bouncy Castle Java library.

67%

of Anthropic-credited CVEs already present in customer environments

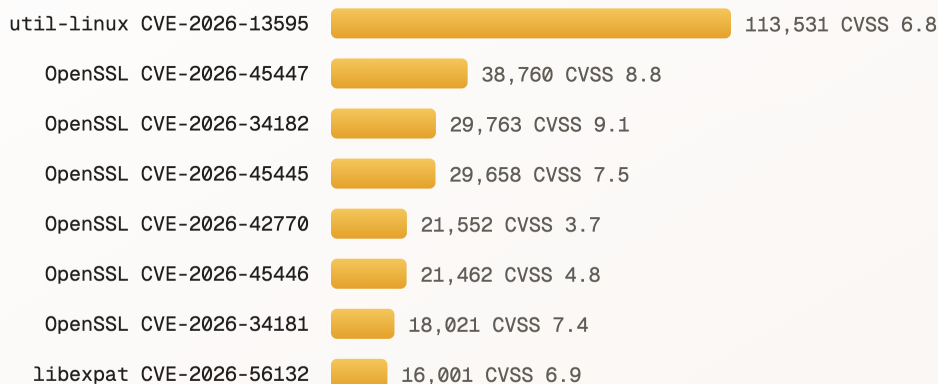
4 days

median time from publication to first appearance in a customer environment

2 in 3

environments carry at least one of them

ANTHROPIC-CREDITED CVEs • ZEST PLATFORM DATA

155 of 231 Anthropic-credited CVEs
are already in customer environments

Instances per CVE, top 8 of 158 found. Source: ZEST platform data, Sep 23, 2026.

What arrived is noisy. Different tools reported 75 of the 158 identifiers at three or four different severity levels, and two OpenSSL vulnerabilities with a CVSS score of 3.7 were reported as Critical in at least one environment. None of them is on CISA's KEV catalog today, and the only one reported as exploited in the wild, in the Ghost publishing platform, has also appeared in customer environments. In the one environment where ZEST assigns contextual severity at scale, 99.9% of the 50,173 validated instances came out Low. Across all customer environments, 68% of the instances have been closed, 3% are Mitigated by existing controls and 26% are still open.

This is the workload AI-driven discovery creates for defenders: each wave of disclosures becomes, within days, hundreds of thousands of vulnerability instances with conflicting severities and almost no exploitation evidence. The eight OpenSSL CVEs published together on June 9, 2026 account for about 188,000 instances on their own, which is the kind of batch that root-cause remediation turns into one upgrade per affected system instead of eight streams of tickets.

WHAT ZEST SEES

The remediation gap, measured

The industry data describes the pressure, and ZEST platform data shows what it looks like inside enterprises, where the vulnerabilities from every scanner, cloud security platform and application security tool land in one place.

77M

vulnerabilities across more than
30 enterprise environments

54%

rated High or Critical by the
tool that reported them

1 in 161

of those with a CVE identifier
is on CISA's KEV catalog

Validate: exploitation evidence is rare, and severity misses some of it

Exploitation evidence is the most objective filter, and it is rare. Matched against the current CISA KEV catalog, only 1 in 161 (0.6%) of the High or Critical vulnerabilities in customer environments is known to be exploited, and 1.2% have an Exploit Prediction Scoring System (EPSS) probability of 50% or more. The known-exploited share varies between environments, from 0.2% to 5.2%, with a median of 1.0%.

EXPLOITATION EVIDENCE · ZEST PLATFORM DATA

Known exploitation is rare in a High and Critical backlog

With a CVE identifier **72.3M**

Rated High or Critical **40.1M**
by the reporting tool

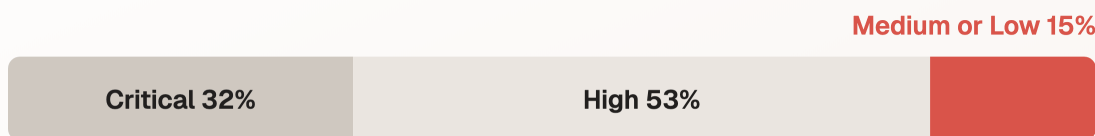
On CISA's KEV catalog **249,319**
1 in 161 of High or Critical

Matched against the CISA KEV catalog of Sep 22, 2026. Source: ZEST platform data.

Severity also points the wrong way for part of the real risk. Across customer environments, 15% of the known-exploited vulnerabilities were rated Medium or Low by the tools that reported them, with the share ranging from under 2% to 32% between environments, so a program that works strictly in severity order reaches them last. More than a third of known-exploited instances (35%) are still open, against 26% of High and Critical vulnerabilities overall, so a backlog sorted by severity alone leaves some of its most dangerous items near the bottom.

KNOWN-EXPLOITED • BY TOOL SEVERITY

One in seven exploited vulnerabilities was rated Medium or Low



Known-exploited vulnerability instances. Sources: ZEST platform data; CISA KEV.

Validate: severity rarely survives context

When ZEST validates a vulnerability, it examines it in the context of the specific environment: whether the vulnerable code is reachable, whether the affected component is deployed and exposed, what an attacker would need, and whether there is evidence of exploitation. The largest example so far is one enterprise environment where ZEST has assigned a contextual severity to 6.1 million vulnerabilities, and there 98.4% came out lower than the reporting tool's rating, 1.6% stayed the same and fewer than 0.1% went up. The pattern held for vulnerabilities that were still open as well as closed ones.

The gap is widest where urgency is highest. Of the vulnerabilities in that environment that tools rated Critical, 0.8% stayed Critical after validation and 88.9% came out Low, and of those rated High, 0.2% stayed High and 91.4% came out Low.

CONTEXTUAL SEVERITY • ONE ENVIRONMENT

Most Critical and High ratings don't survive validation

Rated Critical by the tool

0.8% Critical • 0.9% High • 9.4% Medium • 88.9% Low

Rated High by the tool

<0.1% Critical • 0.2% High • 8.4% Medium • 91.4% Low

6.1 million validated vulnerabilities in one enterprise environment. Source: ZEST.

This is the same gap Anthropic's ledger shows between a model's severity and a maintainer's, and it drives much of the wasted remediation effort we see: engineering time spent on vulnerabilities that were never critical where they run.

Dismiss and mitigate, with evidence

Across customer environments, 2.8 million vulnerabilities have been dismissed, and 98% of those dismissals carry recorded reasoning that an engineer or an auditor can check. Another 2.0 million are marked Mitigated because a validated compensating control protects the affected asset. In the typical environment, endpoint detection and response accounts for about a quarter of mitigations (24%), cloud network guardrails that keep the asset private, such as private subnets and security groups, for 37%, and network segmentation, firewalls and web application firewalls for 39%. Both vary widely between environments, and dismissal rates range from under 1% to about a fifth of all vulnerabilities.

Remediate: one fix, many vulnerabilities

ZEST groups vulnerabilities by root cause into resolution plans, and across customer environments 6.2 million plans cover the 77 million vulnerabilities, an average of 12.5 per plan. The widest fixes are image and service upgrades, which cover about 145 and 87 vulnerabilities per plan on average, and of the plans with a concrete fix path, 90% are package upgrades to versions that already exist. Three in four vulnerabilities have a concrete fix path today, so remediation capacity is best spent on choosing and applying the fixes that remove the most real exposure.

THE ZEST READINESS FRAMEWORK

Maturity means closing the gap between seeing risk and being safe

Most maturity models measure how much a team can see, and this one measures the distance between seeing a risk and being protected from it. At low maturity that distance is wide, because a team learns about a risk after it has been exploited, while at high maturity it approaches zero, because exposure is validated, mitigated or remediated before an attacker can use it.

THE ZEST READINESS FRAMEWORK

Preemptive teams remove exposure before an attacker can use it



Stage 2 is where confidence runs ahead of protection.

STAGE 1 Exposed

Risk is discovered after exposure. Scanners report vulnerabilities with no reliable way to tell which ones are real, remediation follows whatever is on fire, and attackers set the timeline.

STAGE 2 Aware

Scanners, cloud security platforms and software composition analysis are in place, dashboards exist, and a growing backlog is worked in severity order. Confidence runs high because the tooling visibly works, yet most of the backlog is less urgent than its ratings suggest and some exploited vulnerabilities sit in the Medium and Low tiers, so the gap is at its widest and adding detection widens it further.

STAGE 3 Responsive

Remediation has owners and service-level agreements (SLAs), prioritization accounts for exploitability and business impact, and fixes are verified and measured. The gap is narrowing, but the posture is still reactive, because the team has become good at chasing risk after it arrives.

STAGE 4 Preemptive

Every vulnerability is validated for reachability and exploitability in context. Vulnerabilities that existing controls protect against are marked Mitigated with evidence, those that cannot be exploited are dismissed with evidence, and the rest are fixed at the root cause on a timeline set by real exploitation, so remediation capacity goes only where it changes risk.

A short self-assessment

1. **When a scanner reports a vulnerability, how do you decide whether it is real?**
 - a) We work the list
 - b) By its severity score
 - c) Manual review of the important ones
 - d) Automated validation of reachability and exploitability in context
2. **What decides the order of remediation?**
 - a) Whatever is on fire
 - b) CVSS severity
 - c) Exploitability and business impact
 - d) Validated exploitability, exploitation evidence and remediation effort
3. **How much of your backlog is already protected by controls you run?**
 - a) We don't know
 - b) We suspect a lot but cannot prove it
 - c) We track some mitigations by hand
 - d) Every mitigation is validated and recorded with evidence
4. **When an exploitable vulnerability appears, how fast can you close it?**
 - a) No defined process
 - b) We see it, but remediation lags
 - c) Days, with clear ownership
 - d) Fast, because one root-cause fix clears many vulnerabilities at once
5. **Could you show an auditor why a vulnerability was dismissed?**
 - a) No
 - b) Informally
 - c) For some of them
 - d) Yes, every dismissal and mitigation is logged with its evidence

Count your answers: mostly a puts you at Stage 1, mostly b at Stage 2, mostly c at Stage 3 and mostly d at Stage 4. Moving up a stage means validating, mitigating and remediating at the scale you already detect. The playbook that follows is where to start, and a free Mythos exposure check from ZEST at zestsecurity.io/book-a-demo shows where your environment stands today.

THE READINESS PLAYBOOK

Five moves to start this quarter

Each move works with a team's existing tools and comes with a reference point from ZEST data to measure progress against.

MOVE 1 Check the Mythos wave against your inventory

Match the public tracker of Anthropic-credited CVEs (listed under Resources) against what your scanners report, and repeat the check whenever a new batch is published. Two thirds of those CVEs have turned up in ZEST customer environments, usually within days of publication.

MOVE 2 Take known-exploited vulnerabilities out of the severity queue

Match every reported CVE against CISA's KEV catalog and move the matches to the front of the queue, whatever severity the reporting tool assigned. About one in seven known-exploited vulnerabilities in ZEST customer environments was rated Medium or Low, where a severity-ordered queue reaches it last.

MOVE 3 Validate Critical and High ratings in context

Test reachability and exploitability before a Critical or High vulnerability consumes remediation capacity, starting with a sample if the backlog is large. In the one environment where ZEST has validated at scale, fewer than 1 in 100 Critical ratings survived.

MOVE 4 Credit the controls you already run

Map endpoint detection and response, cloud guardrails, network segmentation, firewalls and access policies to the vulnerabilities they protect against, confirm that each control blocks rather than only alerts, and record those vulnerabilities as Mitigated with evidence. Existing controls account for 2.0 million Mitigated vulnerabilities across ZEST customer environments.

MOVE 5 Fix at the root cause

Group vulnerabilities that share an image, a package or a configuration into one fix, and send owners that fix instead of a list. The average ZEST resolution plan covers 12.5 vulnerabilities, and image upgrades cover about 145 each.

Run as routine work rather than one-off projects, these five moves are what Stage 4 looks like.

CLOSING THE READINESS GAP

Validate, mitigate, remediate, with evidence for every decision

The five moves work by hand at small scale, but at enterprise scale they have to run as one loop for every reported vulnerability, whatever tool found it and however it was rated. That loop is what ZEST is built around.

Validate. ZEST tests each reported vulnerability for reachability and exploitability and dismisses the ones that fail, with evidence. So far, 2.8 million vulnerabilities in customer environments have been dismissed this way, 98% of them with recorded reasoning.

Mitigate. If a control you already run, like a zero trust access policy, network segmentation or a cloud guardrail, protects against a vulnerability, ZEST Preemptive Mitigation validates that protection and marks the vulnerability Mitigated. Endpoint detection and response, cloud network guardrails, network segmentation, firewalls and web application firewalls account for the 2.0 million vulnerabilities marked Mitigated so far.

Remediate. Everything else goes to Automated Remediation, where ZEST's AI agents fix it at the root cause, so one misconfiguration or outdated package behind many exposures becomes one fix, simulated on a Digital Twin before it is applied through the organization's own deployment system. Across customer environments, 6.2 million resolution plans cover 77 million vulnerabilities, so the average plan covers 12.5 of them.

Trust comes from evidence, which is why no fix ships without a simulation and dismissals and mitigations carry the reasoning behind them, ready for an auditor or an engineer to check.

12.5

vulnerabilities covered by the
average resolution plan

98%

of dismissals carry recorded
reasoning

90%

of concrete fixes are upgrades
to versions that already exist

FRAMING IT FOR THE BOARD

Boards fund outcomes they can measure

A board cares less about how many CVEs sit in an environment than about what carrying them costs: engineering time spent on risk that is smaller than it looks, the likelihood of an incident, and the effort of proving remediation to auditors, regulators and insurers. Six measures show whether those costs are coming down, each with a reference point from ZEST data or public research.

MEASURE	WHAT IT SHOWS THE BOARD	REFERENCE POINT
Validated in context	How much of the backlog is real risk	98.4% came out below the tool's severity in one environment
Dismissed with evidence	Effort the team no longer spends, with an audit trail	98% of 2.8 million dismissals carry recorded reasoning
Mitigated by existing controls	Protection from controls already paid for	2.0 million vulnerabilities, protected by endpoint, cloud network and network controls
Known-exploited still open	Exposure attackers already use	35% of known-exploited instances, against 26% of High and Critical overall
Time to fix exploitable vulnerabilities	Whether fixes land before attackers arrive	Mandiant estimates the mean time to exploit at minus seven days
Vulnerabilities closed per fix	How far remediation capacity stretches	12.5 per ZEST resolution plan on average

The gains compound, because every vulnerability correctly dismissed or mitigated is one that nobody has to triage, patch, justify or escalate later. Tracked quarter over quarter, these measures show the board what risk the organization still carries and why.

WHAT NOW

Discovery becomes table stakes

As discovery gets cheaper, visibility stays necessary but no longer defines a security program. What defines it is the distance between identifying a risk and eliminating it, and that distance shrinks when a team proves which vulnerabilities are exploitable, credits the controls that already protect them and fixes the rest at the root cause, with evidence behind every decision.

Anthropic's own advice to defenders agrees: it expects "the volume of security work" to increase drastically and asks defenders to treat dependency bumps that carry CVE fixes "as urgent, rather than routine maintenance," which fits ZEST data, where package upgrades make up 90% of the concrete fixes.

METHODOLOGY

How the numbers were produced

ZEST platform data. Figures aggregate the vulnerability data ZEST holds for its customers, first recorded between October 6, 2024 and September 23, 2026 and queried read-only on September 23, 2026. Internal, demonstration and test tenants are excluded. A vulnerability is one detected instance on one asset as reported by one tool, and 93.8% of instances carry a CVE or GitHub Security Advisory identifier. Exploitation evidence comes from matching CVE identifiers against two sources dated September 22, 2026: the CISA KEV catalog, with 1,721 entries, and FIRST's EPSS scores. Because totals favor larger environments, we checked headline ratios per environment and give ranges where they differ. Contextual severity has so far been assigned at scale in one enterprise environment, so figures based on it describe that environment. The control mix weights equally the eight environments with 1,000 or more Mitigated vulnerabilities, as of October 8, 2026.

Anthropic-credited CVEs. We combined the 231 CVEs in the public Anthropic-credited CVE tracker (retrieved September 23, 2026) with the identifiers on Anthropic's ledger (as of August 26, 2026) and matched them exactly against identifiers extracted from vulnerability names across customer environments. Instances that carry the identifier only outside the name are not counted, so the results are a lower bound.

Public data. NVD counts come from the NVD CVE API 2.0 and malicious-package counts from the OSV.dev export, and other figures are cited to their primary sources.

ABOUT ZEST

Agentic exposure management and remediation

ZEST Security closes the gap between the vulnerabilities an organization's tools report and the risk it carries. The platform validates exploitability and reachability, dismisses what cannot be exploited with evidence, marks what existing controls protect against as Mitigated, and uses AI agents and Digital Twin simulation to remediate the rest at the root cause.

In its Emerging Tech Impact Radar: Preemptive Cybersecurity, published September 11, 2026, Gartner® named ZEST Security a Sample Vendor in two categories, Preemptive Exposure Management and Autonomous Exposure Remediation, and ZEST is the only vendor named in both. Gartner writes that autonomous exposure remediation (AER) "shifts security from reactive, manual patching to a preemptive discipline." Learn more at zestsecurity.io.

Gartner, Emerging Tech Impact Radar: Preemptive Cybersecurity, Elizabeth Kim, Charanpal Bhogal, Luis Castillo, Bryan Haley, David Senf, Tom Powledge, Ayelet Heyman, Neil MacDonald, 11 September 2026. Gartner does not endorse any company, vendor, product or service depicted in its publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner publications consist of the opinions of Gartner's business and technology insights organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this publication, including any warranties of merchantability or fitness for a particular purpose. GARTNER is a trademark of Gartner, Inc. and/or its affiliates.

RESOURCES

Sources and further reading

- 01 NIST National Vulnerability Database, CVE API 2.0, monthly publication counts, nvd.nist.gov (queried September 23, 2026).
- 02 OSV.dev, malicious package advisories (MAL-) for npm, osv-vulnerabilities.storage.googleapis.com/npm/all.zip (retrieved September 23, 2026).
- 03 Anthropic, "Claude Mythos Preview," April 7, 2026, anthropic.com/research/mythos-preview.
- 04 Anthropic, "Project Glasswing: Securing critical software for the AI era," April 7, 2026, anthropic.com/glasswing.
- 05 Anthropic, Coordinated Vulnerability Disclosure ledger, as of August 26, 2026, red.anthropic.com/2026/cvd.
- 06 P. Garrity, Anthropic Credited CVEs tracker, github.com/patrickmgarrity/Anthropic-Credited-CVEs (retrieved September 23, 2026).
- 07 VulnCheck, "The Anthropic Glasswing Receipts Are Starting to Trickle In," September 8, 2026.
- 08 UK AI Security Institute, "Our evaluation of Claude Mythos Preview's cyber capabilities," April 13, 2026, aisi.gov.uk.
- 09 AISLE, "AI Cybersecurity After Mythos: The Jagged Frontier," April 2026, aisle.com/blog.
- 10 Cloudflare, "Project Glasswing: what Mythos showed us," May 18, 2026, blog.cloudflare.com/cyber-frontier-models.
- 11 M. Cembalest, "Misanthropic," Eye on the Market, J.P. Morgan Asset & Wealth Management, April 13, 2026, am.jpmorgan.com/us/en/asset-management/liq/insights/market-insights/eye-on-the-market/misanthropic.
- 12 Mandiant (Google Cloud), "Analysis of Time-to-Exploit Trends: 2021-2022," and "M-Trends 2026," March 23, 2026, cloud.google.com/blog/topics/threat-intelligence.
- 13 Google Threat Intelligence Group, "Vulnerability Discovery and Exploitation Trends in the AI Era," October 1, 2026, cloud.google.com/blog/topics/threat-intelligence/vulnerability-discovery-and-exploitation-trends-in-the-ai-era.
- 14 Google Cloud Office of the CISO, "Cloud Threat Horizons Report H1 2026," March 2026, cloud.google.com/security/report/resources/cloud-threat-horizons-report-h1-2026.
- 15 CISA, Known Exploited Vulnerabilities Catalog, cisa.gov/known-exploited-vulnerabilities-catalog; FIRST, Exploit Prediction Scoring System (EPSS), first.org/epss.
- 16 Unit 42, Palo Alto Networks, "Shai-Hulud Worm Compromises npm Ecosystem in Supply Chain Attack," unit42.paloaltonetworks.com/npm-supply-chain-attack.
- 17 OpenAI, "OpenAI and Hugging Face partner to address security incident during model evaluation," July 21, 2026, openai.com.
- 18 J. Spracklen et al., "We Have a Package for You! A Comprehensive Analysis of Package Hallucinations by Code Generating LLMs," [arXiv:2406.10279](https://arxiv.org/abs/2406.10279), 2024.
- 19 Gartner, Emerging Tech Impact Radar: Preemptive Cybersecurity, September 11, 2026.

Further reading from ZEST

- 20 "Navigating the CTM Journey with ZEST Security: A Practical Guide," December 10, 2025, zestsecurity.io/blog/navigating-the-ctm-journey-guide.
- 21 "Cloud-Native CTM: Modern Exposure Management with AI," March 24, 2026, zestsecurity.io/resources/content/building-a-cloud-native-ctm-exposure-management.
- 22 "Exposure Management: What It Is, Why It Matters, and How to Get It Right," May 14, 2026, zestsecurity.io/resources/content/exposure-management-explained.
- 23 "Understanding Preemptive Exposure Management and Why it Matters," April 7, 2025, zestsecurity.io/blog/understanding-preemptive-exposure-management-and-why-it-matters.
- 24 "Beyond CVSS: Why EPSS and KEV Are Game-Changers for Prioritizing Vulnerabilities," February 13, 2025, zestsecurity.io/blog/beyond-cvss-why-epss-and-kev-are-game-changers-for-prioritizing-vulnerabilities.
- 25 "When 'Critical' Isn't: the Summer 2026 OpenSSL and curl Wave," July 15, 2026, zestsecurity.io/resources/content/when-critical-isnt-the-summer-2026-openssl-and-curl-wave.
- 26 "How AWS SCPs Mitigate Risk, Cut Noise, and Reduce Backlogs," August 18, 2025, zestsecurity.io/blog/how-aws-scps-mitigate-risk-cut-noise-and-reduce-backlogs.
- 27 "Your WAF blocks everything except your CVE," September 6, 2026, zestsecurity.io/resources/content/block-it-before-you-patch-it.
- 28 "Agentic Remediation: How AI Closes the Risk Gap," June 9, 2026, zestsecurity.io/resources/content/agentic-remediation-how-ai-closes-the-risk-gap.
- 29 "Preparing for Project Glasswing: Anthropic's Mythos AI Changes Everything," April 8, 2026, zestsecurity.io/resources/content/the-new-defensive-mandate-anthropics-mythos-model-and-project-glasswing.
- 30 "State of the CVEs: September 2026," September 30, 2026, zestsecurity.io/resources/content/state-of-the-cves-september-2026.
- 31 "Gartner® names ZEST twice in one report," September 22, 2026, zestsecurity.io/resources/content/gartner-preemptive-exposure-management-autonomous-exposure-remediation.

Contributors: ZEST Security Research. Platform data is aggregated and anonymized, and no customer is identified.

FREE MYTHOS EXPOSURE CHECK

If your team knows about more vulnerabilities than it can close, request a free Mythos exposure check at zestsecurity.io/book-a-demo. ZEST matches the CVEs credited to Anthropic's research and CISA's KEV catalog against your own vulnerability data, then shows which of them are exploitable in your environment, which your controls already mitigate and which fixes remove the most exposure.

