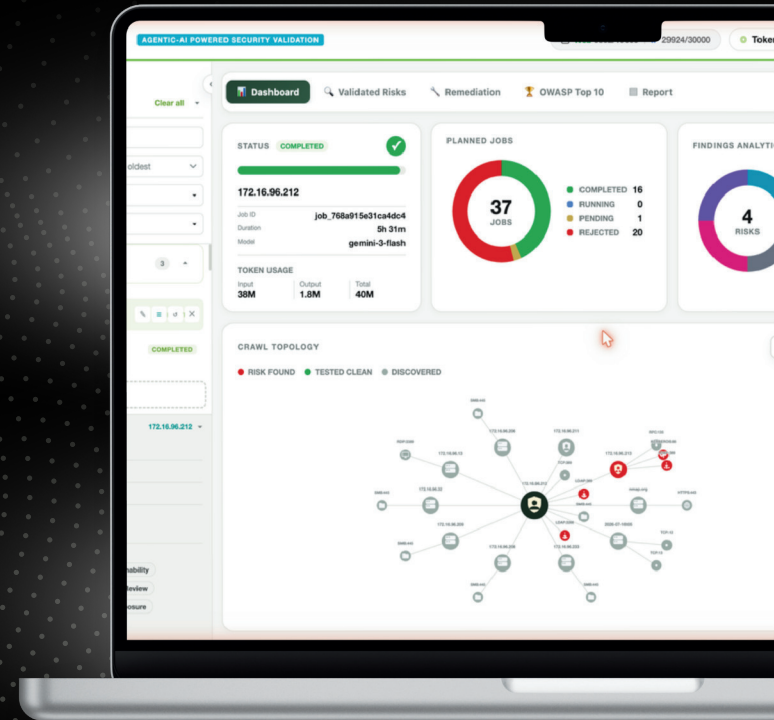


RIDGE SECURITY

AGENTIC AI OFFENSIVE SECURITY

Find. Prove. Fix.



AI HAS CHANGED OFFENSE.

In April 2026, Anthropic previewed Claude Mythos: an AI capable of discovering complex vulnerabilities in production-grade code, writing working exploits on the first attempt in over 83% of cases, and surfacing a 27-year-old flaw in OpenBSD during pre-release testing.

It followed a string of milestones. Google Big Sleep intercepted a live SQLite zero-day before any human defender. DARPA's AIxCC demonstrated end-to-end autonomous discovery and patching. AI-orchestrated offense has been observed in the wild, at scale.

But the headline question:

“Can AI find bugs? - is the wrong one.”

The answer has been yes for years. The question that matters for a CISO is what this means for your environment, right now.

Claude Mythos

Apr 2026

First-attempt working exploits in 83%+ of cases; thousands of zero-days across every major OS and browser.

Google Big Sleep

Jul 2025

Detected and foiled CVE-2025-6965 in SQLite before any public exploit, an AI first.

DARPA AIxCC

2025

End-to-end autonomous vulnerability discovery and patching, demonstrated live.

Offense in the wild

2025–26

AI-orchestrated attacks now observed operating at machine speed and scale.

Source: “Beyond Mythos”, Nick Mo (Ridge Security) & Ken Huang.

AI FINDS OLD BUGS. IT ALSO SHIPS NEW ONES.

The same systems that find vulnerabilities are writing an enormous amount of new vulnerable code into production every day. “AI finds old bugs” gets the headlines; “AI ships new bugs” gets a footnote.

Defenders now face two compounding pressures at once: an adversary that discovers and weaponizes flaws at machine speed, and a codebase growing more vulnerable with every AI-assisted commit.

The result is a validation crisis. Two-thirds of practitioners now spend more than half their time manually confirming findings, not fixing them.

AI IS WRITING VULNERABLE CODE AT SCALE

45% of AI-generated code samples failed security tests, introducing OWASP Top 10 flaws

~10× more security findings per month from AI-generated code by mid-2025 vs late 2024

1.7× more major issues in AI-authored code: broken auth, IDOR, XSS

AND DEFENDERS CAN'T KEEP UP

of security practitioners spend more than half their time manually validating findings instead of remediating them.

66%

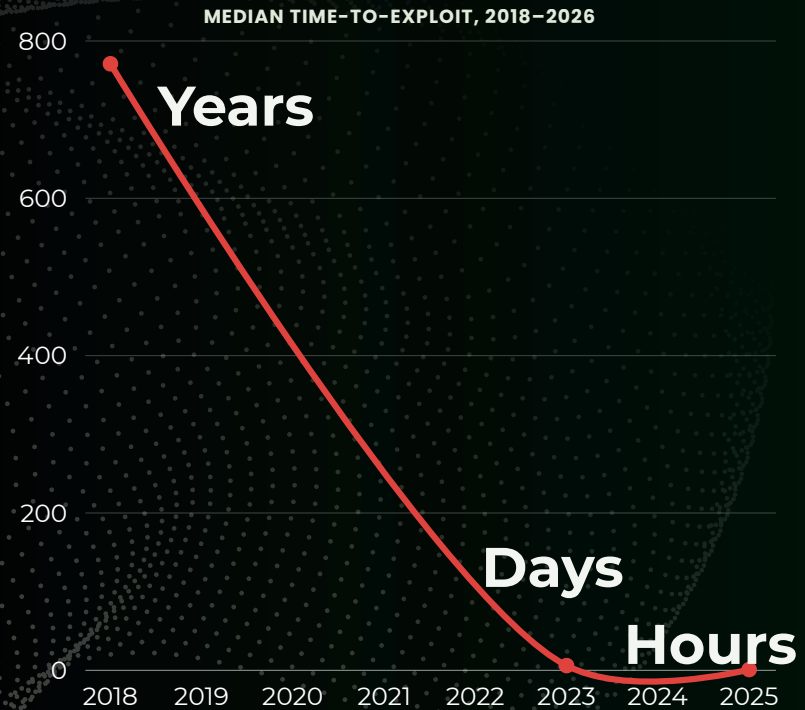
Sources: Veracode, Apiiro, CodeRabbit, ProjectDiscovery (2025–26), via Beyond Mythos.

THE EXPLOITATION WINDOW HAS COLLAPSED.

The speed asymmetry between attackers and defenders has widened for years. AI turned a trend into a structural crisis. The Zero Day Clock, tracking 83,000+ CVE-exploit pairs, tells the story every vulnerability-management conversation now has to start from.

In 2018, teams had over two years to patch. By 2024, exploitation arrived in hours. In 2025, the majority of CVEs were weaponized before public disclosure.

A collapse from 771 days to hours. Point-in-time testing was built for a slower world.



Sources: Zero Day Clock, Verizon DBIR, Mandiant M-Trends, CISA KEV.

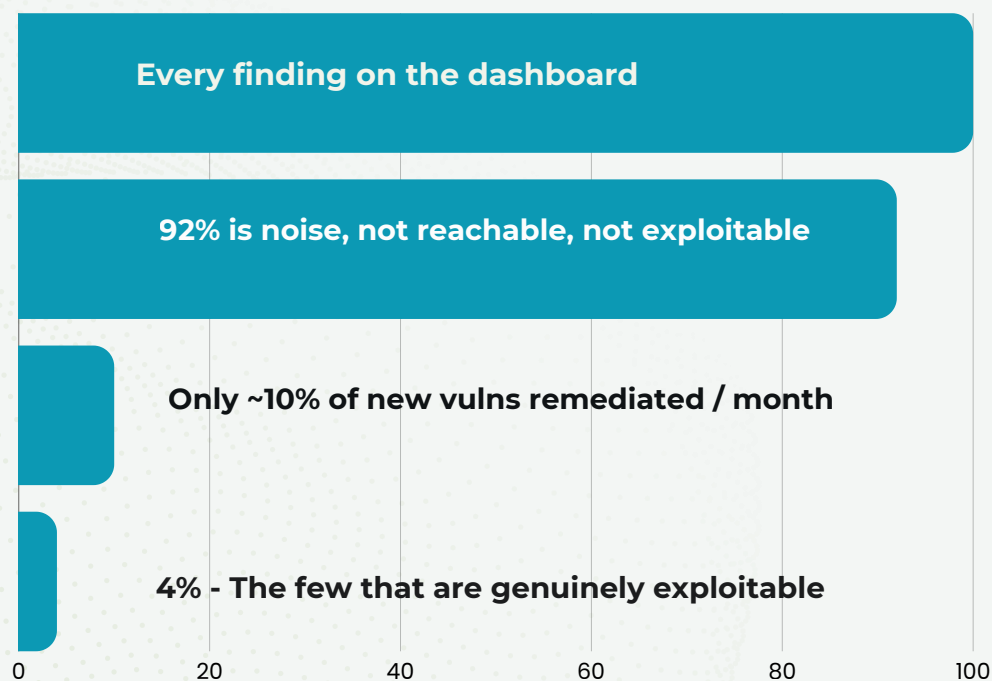
MOST OF WHAT YOUR TOOLS FLAG IS NOISE.

For two decades the industry answered risk with more visibility: more scanners, more SIEMs, more posture tools. Every tool adds signals; few add clarity.

The gap between knowing a vulnerability exists and knowing whether an attacker can reach it in your environment has never been wider. In context, the overwhelming majority of dashboard findings are not reachable, not exploitable, and not a real risk.

AI widens the gap. Attackers now search the possibility space faster than defenders can investigate their own alerts.

Ridge Security validates which findings an attacker can actually exploit, cutting the noise.



Sources: Cyentia Institute; Beyond Mythos analysis.



Find

Broad, continuous discovery of the real, reachable attack surface across your entire estate.



Prove

Evidence-verified exploitability. No noise, no false positives, only what an attacker could actually do.



Fix

Click-away, evidence-backed remediation, collapsing mean time to remediate from months to minutes.

CONTINUOUS, AUTONOMOUS VALIDATION.

The answer is not another scanner or a once-a-year pen test. It is continuous, autonomous adversarial validation that does three things traditional tools cannot:

Confirms exploitability in context: Empirically, which findings an attacker can actually reach and exploit in your environment.

Chains attack paths: Combines low-severity weaknesses into the multi-step routes real adversaries use.

Runs continuously: Every change, deployment, and integration, revalidated, not sampled once a year.

Ridge Security provides the solution, built for breadth, depth and speed, able to find, prove and fix security vulnerabilities at machine speed.



TWO LOW CVEs. ONE FULL TAKEOVER.

In January 2026, Ridge Security's research team documented a chained attack against n8n, a widely deployed workflow automation platform.

Two CVEs that looked limited in isolation combined into complete compromise.

A scanner flagged both, independently. It could not tell whether the chain was reachable. Signature and version tools see each finding alone.

RidgeBot proved the whole chain. End-to-end, with zero false positives and no system state altered, confirming exploitability in its real, deployed context.



Arbitrary file read

CVE-2026-21858 · Content-Type confusion



Credential decryption

Encryption keys read from local disk



Admin session

Hijacked via decrypted secrets



Remote code execution

CVE-2025-68613 · sandbox escape



Full workflow takeover

Complete compromise of the platform

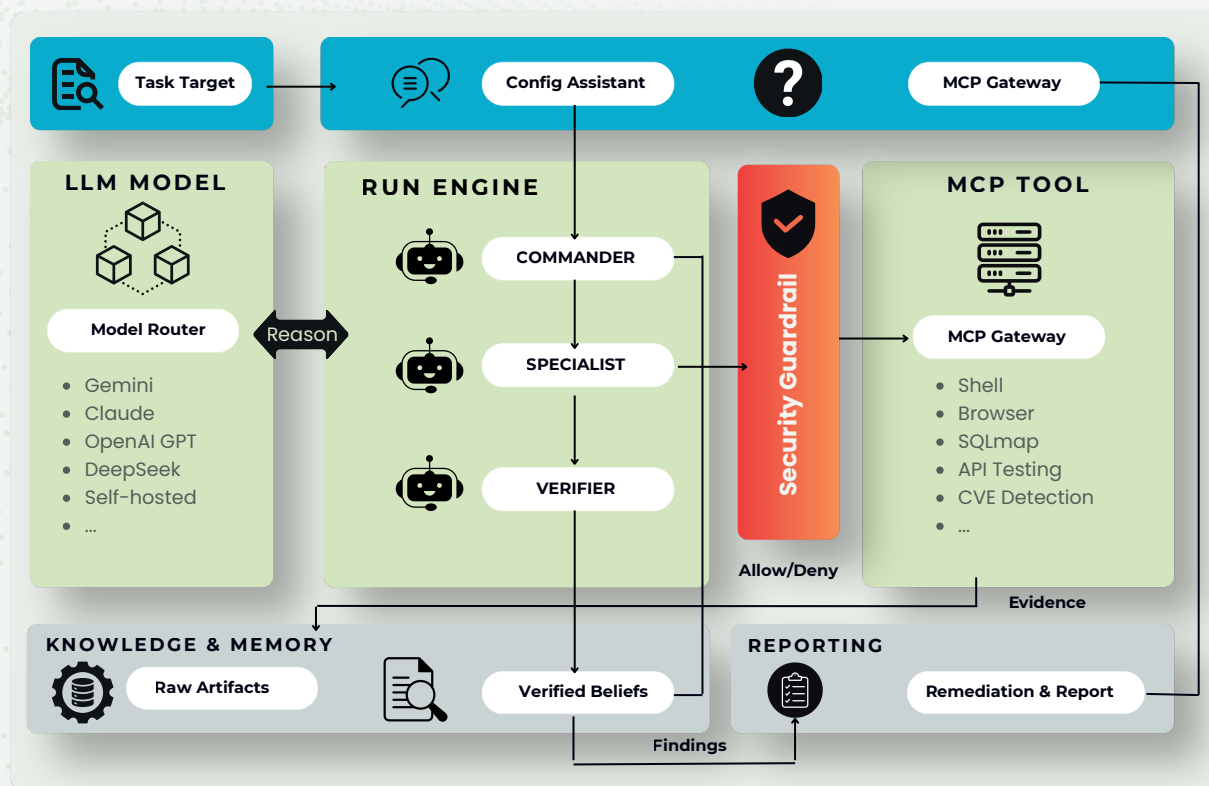
Source: Ridge Security research team, Jan 2026 (published in Beyond Mythos).



THE BRAIN BEHIND THE SOLUTION.

Driving Ridge Security's agentic platform is NG-RidgeGen, a purpose-built orchestration framework. The LLM provides creative, human-like reasoning; RidgeBrain binds it to deterministic test paths, explicit authority, enforced guardrails, and a fully auditable evidence trail.

An LLM can explain an attack. RidgeGen can autonomously validate one, and prove it.



Model-agnostic

The best LLMs, or your own

Platform-agnostic

IT, IoT, OT, and AI infra

On-premises

Small footprint; data stays inside

BYOA

Bring your own agent



PROVEN, NOT JUST CONFIDENT.

Autonomous testing carries a high trust burden: the system inspects real systems, makes decisions mid-run, and explains results afterward. Ridge Security's answer is not "the model is confident." It is a runtime chain that answers four key questions.

Evidence is a structural gate. A banner, version match, or agent assertion is a signal, never an automatic finding. Only claims supported by preserved artifacts are promoted to customer output.

Safety sits outside model discretion. Guardrails are the product safety contract, enforced by trusted controls and recorded as proof you can inspect.

01

Authority trust

Was this allowed?

Targets, policy, credentials, approvers, and launch authority from explicit Rules of Engagement.

02

Behavior trust

Was it controlled?

Guardrails apply trusted decisions at safety boundaries, never left to prompt obedience.

03

Finding trust

Is the claim supported?

A candidate becomes a trusted finding only when evidence and artifacts verify it.

04

Explanation trust

Is the answer grounded?

Reports answer from run records, findings, and artifacts, not vague model memory.

Source: RidgeGen Technical White Paper, §7 (The trust model).

AUTONOMY WITH CONTROLS.

RidgeGen delivers autonomy with control. Guardrails are the product safety contract: every safety-relevant boundary is enforced outside model discretion, and recorded as proof you can inspect.

The same principle protects credentials. SafeBox keeps real customer credentials out of the model's prompt, memory, and evidence trail, applying them only at the tool boundary, outside the model's view.

Even a compromised target cannot make the agent reveal what it never held.





BROAD ENOUGH TO FIND IT. DEEP ENOUGH TO PROVE IT.

Ridge Security is one AI-native strategy delivered through two complementary platforms, orchestrated by a single system, RidgeBrain. RidgeBot and RidgeGen act as security peers, but built for different jobs:

RIDGEBOT brings breadth. Machine learning AI, delivering continuous, production-safe validation across the whole estate, on compliance cadences.

RIDGEGEN brings depth. Agentic AI that augments human red teams, proving what is genuinely exploitable on a single authorized target.

Together they deliver comprehensive, continuous, agentic AI security.

RIDGEBOT

BUILT FOR BREADTH WITH
MACHINE LEARNING AI

- Continuous, production-safe
- Networks, apps, APIs, cloud, OT, identity
- CVE-centric, at scale
- Compliance cadences

RIDGEGEN

BUILT FOR DEPTH WITH
AGENTIC AI

- Agentic AI red teaming
- Multi-step attack chains
- Business-logic & non-CVE risk
- Periodic deep validation

WHEN PAIRED, NOTHING COMPARES.

RidgeBot and RidgeGen are designed for different jobs.

RIDGEBOT brings breadth: Machine learning AI capability, delivering continuous, production-safe validation across networks, applications, APIs, cloud, OT, and identity.

RIDGEGEN brings depth: Agentic AI that reasons through multi-step attack chains and uncovers the business-logic and non-CVE risks a scanner never sees. Run either alone and you get a partial picture. Run them together and the gaps close significantly.

When paired, they deliver one continuous loop from find, to prove, to fix, with shared guardrails and click-away remediation that turns validated findings into resolved risk in minutes.

RIDGEBOT

Architecture

Machine learning AI

Positioning

Autonomous penetration testing

Approach

Breadth, at scale

Coverage

CVE-centric (OWASP & MITRE)

Cadence

Continuous, compliance cadences

RIDGEGEN

Architecture

Agentic AI

Positioning

Augments human red teams

Approach

Depth, on specific targets

Coverage

Business-logic & non-CVE risk

Cadence

Periodic deep validation

SECURITY OUTCOMES YOU CAN SEE.

Ridge Security aligns to Continuous Threat Exposure Management (CTEM), the operating model for enterprise-scale exposure reduction.

Automated validation is what makes each stage viable.

- **Cost out**
retire scanners; cut testing spend
- **Capability in**
red-team depth, no new headcount
- **Months to minutes**
Mean time to remediate collapses on validated fixes
- **Business resilience**
protect revenue and reputation

ALIGNED TO THE CTEM CYCLE

- 1 Scope**
Crown jewels, regulated & exposed systems
- 2 Discover**
Continuously inventory assets and map vulnerabilities
- 3 Prioritize**
Rank by CVSS + business context + real reachability
- 4 Validate** **Ridge Security→**
Confirm exploitability through safe adversarial testing
- 5 Mobilize**
Feed validated findings into patch & ticketing workflows

Source: Gartner CTEM framework; Ridge Security implementation.



RIDGE SECURITY

AGENTIC AI FOR OFFENSIVE SECURITY

FIND | PROVE | FIX

Request a live demo →



Contact your Sales Consultant for more information

