YesWeHack

HOW YESWEHACK IMPLEMENTS ANTHROPIC'S PRESCRIPTIONS

READ →

Mythos Preview's zero-day discoveries may have sparked widespread alarm, but the response within the cybersecurity industry was more measured. Some experts even dismissed the frontier model's findings as overblown.

Either way, the current capability of Claude Mythos – or OpenAI's 'Cyber' equivalent – is ultimately beside the point. What matters is the industry's response to a reality that is not in dispute: these LLMs will only become more powerful over time.

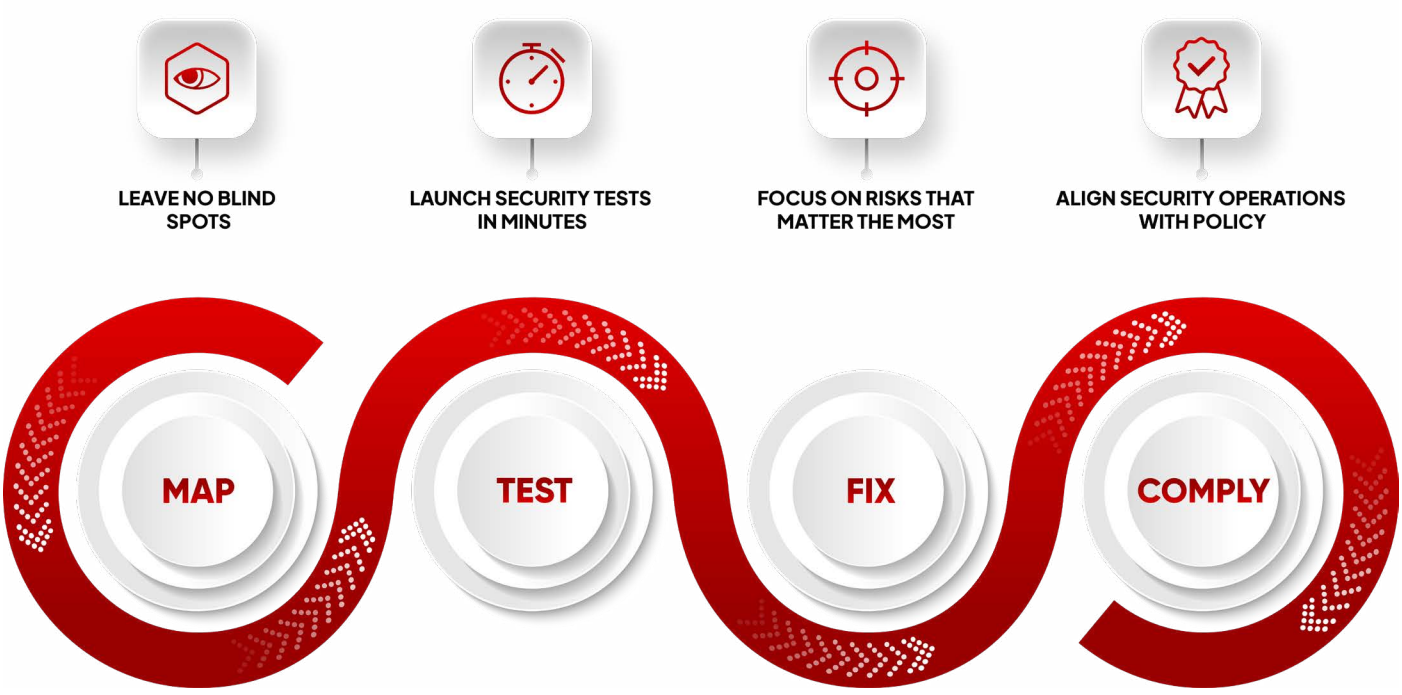
Countering accelerating attacks

For YesWeHack, the Mythos moment is simply the latest milestone in an automation trend that has long shaped our approach. Coding automation – first through DevOps and Infrastructure as Code (IaC) and now through tools like Claude Code – has accelerated the expansion of attack surfaces. And cyber-attack automation has progressed from mass scanning and exploit frameworks to LLMs that potentially further increase the speed and scalability of attacks while lowering the barriers to entry.

Built around a four-step cycle of MAP > TEST > FIX > COMPLY, our **offensive security and exposure management platform** delivers real-time visibility of fast-evolving attack-surfaces and rapid validation of the most urgent threats as exploitation timelines shrink.

This approach, which combines automated checks for actively exploited vulnerabilities with AI-assisted expert testing, aligns closely with Anthropic's recommendations for **'preparing your security program for AI-accelerated offense'**. Many of these prescriptions were, Anthropic notes, "already part of the existing security consensus"; the difference now is that the risks of not implementing them have become unignorable. Below, we break down how the YesWeHack platform implements Anthropic's recommendations.

RECOMMENDED: *COST, AI frontier models and more: A measured take on the future of security testing →*



Map, test, fix, comply - YesWeHack's approach is built on a continuous, four-step model

Close your patch gap

ANTHROP\C

"AI models are very effective at recognizing the signatures of known, already-patched vulnerabilities in unpatched systems. Reversing a patch into a working exploit is exactly the kind of mechanical analysis at which these models excel. This means that the window between a patch being published and an exploit becoming available is shrinking."

- **Patch everything on the CISA Known Exploited Vulnerabilities (KEV) catalogue immediately.** YesWeHack's automated testing leverages the KEV catalog and Exploit Prediction Scoring System (EPSS) to target CVEs under active exploitation. Vulnerabilities are prioritised according to real-world exploitability, ensuring rapid remediation of urgent threats.
- **Use EPSS to prioritize the remaining CVEs.** Asset mapping correlates each detected technology with associated CVEs and ranks them using EPSS scores, CVSS and (as defined by the client) asset criticality. Our platform generates a contextualised, prioritised vulnerability queue – as suggested by Anthropic
- **Reduce time-to-patch on internet-exposed system.** Automated tests run every three days and generate actionable reports as soon as a vulnerability is confirmed. The platform and its integrations facilitate routing toward patching systems. The solution covers rapid detection and alerting, but the patching decision and execution remain on the client side.
- **Automate patch deployment and reboots.** Reports have standardised formats compatible with automated integration into patching workflows (ITSM, CI/CD), enabling downstream automation.

RELATED: *Introducing Autonomous Pentest: identify actively exploited vulnerabilities across your attack surface* →

Handle a much higher volume of vulnerability reports

ANTHROP\C

"Over approximately the next two years, the processes you use to receive, prioritize, and fix vulnerabilities (both in your own code and in the software you buy from vendors) will be under far more pressure than they are today. Your Vulnerability Management process should plan for many more patches, from vendors and upstream."

- **Plan for an order-of-magnitude increase in finding volume.** YesWeHack's unified platform is architected to absorb high volumes through its workflows, statuses and assignment mechanisms. Automated prioritisation based on real-world exploitability streamlines triage and vulnerability management. The triage service systematically deduplicates and qualifies findings, delivering only confirmed and actionable items to the client – turning a massive inflow into a short, prioritised list.
- **Check the security of open-source dependencies.** Asset mapping and fingerprinting identify technologies and their versions on exposed assets, enabling detection of vulnerable open-source components accessible from the internet.

- **Apply the same expectations to third-party vendors.** The testing scope can be extended to surfaces exposed by vendors in the supply chain. Asset mapping identifies internet-accessible third-party technologies and their associated CVEs.
- **Speed up triage.** Our in-house triage team systematically deduplicates, validates, enriches, and prioritises every finding, delivering complete contextual summaries that account for system exposure and real business risk. Clients receive only relevant, confirmed vulnerabilities.
- **AI automation of upgrades and fixes downstream.** Anthropic notes frontier models increasing utility for generating and validating patches for "clear and thorough" vulnerability reports, with proofs-of-concept cited as helpful too. Standardised, detailed YesWeHack reports are readily integrated with LLMs or automation tools for these purposes.

BOOK A **DEMO** →

Find bugs before you ship them

ANTHROP\C

"Prevention is always better than cure. You should assume that bugs that reach production will eventually be found, so your security testing needs to happen well before."

- **Add automated penetration testing to the CD pipeline.** YesWeHack's automated tests simulate real attack scenarios against exposed environments. Bug Bounty Programs can also include staging or pre-production environments within the defined scope, in black-box, grey-box or white-box mode.
- **AI vulnerability scanning.** Our automated tests precisely simulate attacker exploits by leveraging known CVEs, KEV and EPSS. Bug Bounty goes further by deploying AI-assisted researchers to uncover vulnerabilities specific to the implementation, configuration and architecture of the client's systems.
- **AI patch generation.** Frontier models can change "the developer's job from 'understand the bug and write a fix' to 'verify a proposed fix is correct'" – accelerating remediation. Our open architecture and detailed, standardised reports can facilitate this downstream integration.

RECOMMENDED: Continuous Pentesting with zero false positives: a fully managed, platform-driven approach →

Find the vulnerabilities already in your code

ANTHROP\C

"Patching addresses known vulnerabilities in software you depend on. But your own codebase contains unknown ones. Most long-running production code has been reviewed by humans many times, but has never been examined by a frontier model, and that kind of analysis ***tends to surface new, previously-overlooked issues****. Proactively scanning can identify vulnerabilities that are within the reach of modern LLMs before attackers discover them themselves."*

- **Prioritise by exposure.** YesWeHack's asset mapping identifies which assets are internet-facing and what technologies they run. Combining CVSS, EPSS and client-defined business criticality, prioritisation ensures efforts are focused on exploit paths with the greatest risk – such as authentication, untrusted input parsing, exposed interfaces.
- **Include legacy code.** Bug Bounty Programs can cover exposed legacy systems in black-box or grey-box mode without requiring internal documentation. Automated tests can also cover legacy systems, which are typically the least well maintained.
- **Budget for remediation.** The triage service delivers only confirmed, actionable vulnerabilities, providing a reliable basis for remediation planning and resource allocation. Every reported finding is real, prioritised, and tied to a concrete action – eliminating noise and unnecessary tickets.
- **Proactive AI-assisted scanning to surface long-overlooked vulnerabilities.** Automated tests scan exposed assets for actively exploited vulnerabilities every three days. AI-assisted Bug Bounty researchers can also identify vulnerabilities specific to an organisation's code, configuration and architecture that historic scans and pentests may have missed.

Actionable vulnerability reports

ANTHROP\C

"If you are scanning code—your own dependencies, open-source projects, or vendor products—and reporting findings upstream, the quality of those reports determines whether anyone acts on them. Open-source maintainers are already receiving large volumes of low-quality automated reports, and many have started ignoring anything that looks AI-generated. Adding to that volume without adding signal makes the problem worse for everyone, including you."

- **Quality of vulnerability reports submitted to third parties.** YesWeHack's triage service fulfil's Anthropic's prescriptions: systematic human verification, clear description of the vulnerability and impact, documented code path, proof of concept, proposed patch. Every delivered report is signed off by an expert as truly actionable – not an automated pattern match submitted at volume.

READ MORE: *Validation: Your path to overcoming alert fatigue in vulnerability management* →

Design for breach

ANTHROP\C

"Attackers will try to get a foothold somewhere. You need to limit what they can reach from there."

- **Replace long-lived secrets with short-lived tokens.** YesWeHack's Automated tests and Bug Bounty Programs surface exposed secrets or credentials on internet-accessible surfaces (eg hardcoded API keys, tokens visible in HTTP responses, credentials in cleartext).

Reduce and inventory what you expose

ANTHROP\C

"This section is based on two important principles. First, you cannot defend systems you don't know about. Second, the smaller the exposed surface, the less there is to attack."

- **Maintain a current inventory of all internet-facing assets.** YesWeHack's asset mapping and fingerprinting capabilities provide real-time visibility of internet-exposed assets and associated technologies, which are then correlated against known vulnerabilities to mirror attacker reconnaissance.
- **Decommission unused systems.** Asset mapping can surface end-of-life technologies or forgotten assets with no clear owner.
- **Minimise what each service exposes.** Asset mapping reveals exposed risks – including open ports, detected technologies, exposed endpoints and service versions – enabling security teams to identify and mitigate unnecessary exposure.
- **AI-assisted pruning of stale code and systems.** Asset mapping identifies poorly maintained assets and internet-facing end-of-life technologies – guiding patching, upgrade or decommissioning decisions.
- **Autonomous external red-teaming.** Our automated tests simulate the scenario described by Anthropic: scanning exposed environments without credentials to identify reachable and exploitable assets. Bug Bounty extends this approach through researcher-led testing capable of chaining vulnerabilities to gain an initial foothold, providing deeper analysis than automated scanners alone.

Shorten your remediation timeline

Anthropic's guidance includes a section on focuses on shortening your incident response time. Many of the same principles also apply upstream to offensive security and exposure management, where reducing the time between vulnerability discovery and remediation is similarly critical.

- **AI model at the front of the alert queue.** The YesWeHack triage service processes all findings, deduplicating, validating and prioritising them to direct remediation efforts toward confirmed, actionable vulnerabilities. The platform also facilitates integration with SIEMs and alert management tools.
- **Exposure visibility.** Tests recurring every three days provide continuous visibility into exposed assets and vulnerabilities, reducing the time between vulnerability emergence and discovery – enabling rapid remediation.
- **Automated bookkeeping around incidents.** Structured, standardised reports with workflows and statuses facilitate automatic documentation of incidents. The enriched format enables downstream integrations with ticketing or incident management systems.
- **Driving the detection flywheel.** Recurring automated tests create a continuous detection loop, incorporating new vulnerability signatures and exploitation tech-

niques as they emerge to keep remediation workflows aligned with the evolving threat landscape.

- **First-pass triage at 100% coverage.** The triage service analyses all findings without minimum severity thresholds. Every vulnerability receives risk characterisation, business context and clear remediation guidance.
- **Proactive hunting across your environment.** Automated tests and Bug Bounty Programs provide systematic, recurring searches for vulnerabilities and misconfigurations across the exposed surface – including forgotten hosts, exposed management interfaces, and default credentials.

What if you lack a dedicated security team?

ANTHROP\C

"Most of the above advice assumes that your organization has a dedicated security function. If you are a small organization, a solo developer, or an open-source maintainer, the same risks apply but the actions are simpler."

- **For organisations without a dedicated security team.** YesWeHack automates detection, triage and prioritisation, enabling smaller teams to gain the visibility and protection capabilities they could not realistically build or maintain internally.

RELATED: *Map, test, fix, comply: unveiling our unified approach to offensive security* →

See the YesWeHack platform in action

If you're looking to expand or improve your security testing program, YesWeHack can help.

YesWeHack provides a full range of automated and human-led testing capabilities that can be combined and customised to fit your security and compliance needs.

Contact YesWeHack for a no-obligation live demo and review of your testing needs.

BOOK A **DEMO** →

YesWeHack

Offensive Security and Exposure Management Platform

YESWEHACK.COM →

FOLLOW US