Combating AI-driven threats: Inside Google's own machine-speed defense strategy

White paper
July 2026

Google Cloud
Security

Executive summary

The rapid advancement of AI has fundamentally altered the cybersecurity landscape, allowing adversaries to operate with unprecedented scale, speed and sophistication. Manual patching and traditional defense strategies are no longer sustainable against machine-speed threats that can identify thousands of vulnerabilities in minutes. Enterprises must move beyond reactive measures and take action now.

This paper outlines Google's approach to hardening existing software and structurally transforming the development lifecycle to build ongoing resilience. Based on insights and lessons learned, Google developed a framework that organizations can leverage and apply to help ensure they remain protected against the next generation of sophisticated cyber threats.

A changing landscape

AI has been rewriting the rules of business and cybersecurity. It is changing the scale, speed, and sophistication of threats: boosting adversaries, uncovering novel techniques, and introducing new surfaces to defend against.

Just a year ago it would have taken months or even years for a good application security team to find hundreds or thousands of vulnerabilities. Today, a team equipped with AI can find the same number in hours — or even minutes.

This expanded targeting is dramatically accelerating the threat cycle. While AI drives a massive surge in newly discovered vulnerabilities, it simultaneously empowers attackers to exploit them with unprecedented speed. With annual CVEs skyrocketing and the mean time-to-exploit plummeting, enterprises face a critical risk window that manual patching alone simply cannot close.

Today, a team equipped with AI can find [hundreds or thousands of vulnerabilities] in hours — or even minutes.

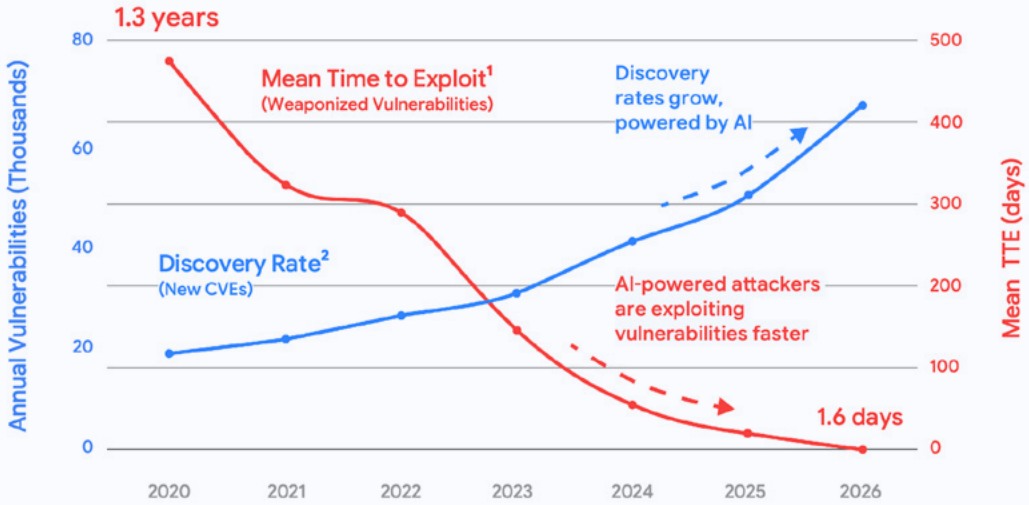

Figure 1: Vulnerability discovery rate and time to exploitation

[1] zerodayclock.com
[2] National Vulnerability Database (NVD); 2026 estimated

We realized a new defense strategy is needed

Google protects billions of users and devices every day. Operating one of the largest attack surface areas on the planet comes with an immense responsibility.

To combat the rise in AI-driven threats, Google recognized the need for a fundamentally new defense strategy, and we've been preparing for this moment for years. Securing the enterprise environment against machine-speed adversaries requires moving beyond a traditional approach to an innovative strategy capable of finding and eliminating vulnerabilities at an unprecedented scale by using multiple AI models, ensuring each class of threat is met with the optimal tool.

To achieve this, we implemented a dual strategy:

- **Leveraging AI to aggressively strengthen the security of our existing software**
- **Making structural improvements to our software development lifecycle (SDLC) to build ongoing resilience directly into the system**

How we did it and lessons learned

Across our products and services, we've found that a unified approach helps us protect our own environment at Google scale. Therefore, we standardize and enforce our defense practices across all business units. Our strategy handles every code tier across the ecosystem—including first-party Google software, open-source software, vendor-managed applications, and firmware.

1. Proactively reduce exposure

2. Find vulnerabilities in the software development lifecycle

3. Remediate with AI

4. Employ continuous detection and response

1. Proactively reduce exposure

To defend against AI-driven attacks, we focus on reducing our overall attack surface area. This includes Google infrastructure, services, products, software (e.g., Android and Chrome), and our business partners. Across these areas, we identify source code dependencies using a shared pipeline, ensuring libraries are scanned and remediated before promotion.

We then use a Configuration Graph to prioritize the vulnerabilities we find based on blast radius and accessibility. In parallel, engineering teams remove unused code, replace dependencies, and retire legacy products, while sandboxing known high risks, such as components written in non-memory-safe languages.

Supporting these defenses requires a structured operational framework — specifically, a centralized inventory, tracking, and dependency management system. This system provides the end-to-end visibility needed to monitor and secure our entire footprint across infrastructure, services, corporate systems, and third-party supply partners. Within this framework, we identify and risk-rank critical assets, including external products and suppliers. Google shares its findings via Open Source Insights, which is a service developed and hosted by Google to help developers better understand the structure, construction, and security of open source software packages.

To support continuous analysis, we explicitly budget for the required tokens, compute infrastructure, teams, and resources, and we integrate these scanning and remediation steps directly into the new product release process. And, our multi-AI strategy enabled us to effectively manage costs by using lighter-weight, faster models for broad, continuous coverage, and reserve frontier models for the highest-risk applications and findings.

Finally, we create product-specific response plans. As the volume of AI-discovered vulnerabilities increases, manual patching processes become unsustainable. By building response plans that are inclusive of automated agents, we ensure the reliable rollout of patched software packages. We employ human-on-the-loop operations where AI agents independently execute workflows, but humans maintain high-level supervision. This approach allows teams to deploy fixes systematically, maintaining the defense posture while preventing operational burnout among security and engineering personnel.

2. Find vulnerabilities in the software development lifecycle

To effectively identify vulnerabilities, we deploy a scanning harness directly into the software development lifecycle (SDLC) pipeline. While the underlying AI models are important, the test harness that orchestrates them is even more critical for consistent and accurate execution.

We establish specific metrics for completeness and depth, executing scans and tracking their coverage across models and products. We scan using multiple AI models, recognizing that different models may be better suited for detecting certain classes of vulnerabilities. Beyond standard code scanning, this process continuously assesses system configurations, identifies toxic combinations, and maintains continuous posture management.

Concurrently, we reevaluate supply chain security and enforce strict code hygiene. This includes scanning binaries released to customers, as well as firmware and code provided by contractors. We also require our external suppliers to formally validate their own testing and remediation practices. To maintain ongoing code hygiene, engineering teams actively remove unused code, retire legacy products, and leverage AI to rewrite aging components into modern languages.

Once findings are generated, we categorize vulnerabilities by severity and distinct vulnerability classes. We analyze the dependencies of the critical assets identified during the preparation phase understand specific remediation requirements. Testing is prioritized for security-critical libraries and infrastructure first, followed by Google-owned OSS, first-party code, non-Google-owned OSS, binaries, and firmware. We then coordinate the remediation of common libraries across the company. Finally, to validate these defenses, we perform AI red teaming to simulate agentic attacks originating from both inside and outside the network.

3. Remediate with AI

To execute remediation effectively, Google uses multiple models, including Gemini, to recommend code patches. Our remediation strategies extend beyond simple patching to include fixing, rewriting, and removing code entirely. When we identify classes of issues, such as memory vulnerabilities, we rewrite the affected code using memory-safe languages.

Remediation efforts are tracked centrally to ensure adherence to service-level objectives (SLOs) for patching. Within a central repository, we label which libraries have been scanned, the specific vulnerabilities found, and the exact AI model and version that identified them. This tracking is paired with automated exploit validation to confirm the risk. We then deploy 24/7 autonomous patching workflows, maintaining human review and robust rollback SRE capabilities to ensure operational stability.

Finally, we build systemic resilience into the remediation lifecycle. Because AI models are continually updated, we rely on continuous scanning to catch newly discoverable vulnerabilities over time. To minimize operational disruption, we accelerate recovery times from SRE teams by automatically detecting and reversing any service degradations caused by new patches. Furthermore, Google has invested heavily in resilient infrastructure; for example, our global datacenter network fabric enables Google teams to patch core network to patch core network and CPU components during normal operational cycles with little to no service disruption.

4. Employ continuous detection and response

To sustain long-term resilience, we have established a continuous feedback loop across the environment. We look for indicators of change volumes to detect system strain or ecosystem complexity, which helps identify operational bottlenecks early. This monitoring focuses on identifying hot spots, such as high-risk open-source libraries and firmware. Furthermore, we consistently monitor for new AI models and capabilities in the market to ensure our defense capabilities evolve alongside state-of-the-art technology, allowing us to track and maintain long-term remediation health.

We maintain rigorous tracking through a centralized asset inventory that monitors our overall scan posture and the completeness of model coverage for remediation. This ensures all systems strictly adhere to product-level vulnerability service-level objectives (SLOs). Operational infrastructure is designed to handle this continuous cycle seamlessly; rolling patches allow even core datacenter hardware to be updated continuously without service interruption.

To ensure these updates are successful, we deploy specialized software agents to verify patch efficacy and assist with ongoing coding and monitoring tasks. We also scan with new models and capabilities. Because models produce non-deterministic results, we conduct iterative runs with the same model and version to potentially uncover additional findings.

Finally, we automate and standardize playbooks to ensure an instantaneous response to emerging threats, including emergency containment procedures to handle machine-speed threats at enterprise scale. To proactively improve our defense systems, key internal libraries are moved into Gemini training workflows to improve pattern recognition and enhance security-aware coding patterns in our code generation agents. We also continuously enhance detection rules and signatures around our core protection platforms. This entire posture is regularly validated through continuous red-teaming of our core infrastructure and corporate systems.

Sharing with the community

As a leader in the technology industry, Google recognizes its responsibility to help the broader ecosystem and open source community defend against emerging cyber threats. A good example is becoming a founding member of the Linux Foundation's Alpha-Omega Project.

We actively share our learnings and capabilities with the community in two primary ways. First, we publish the best practices we develop, empowering enterprises to integrate these strategies directly into their own software development processes and security teams. Second, we make solutions available that deliver the same technology and expertise Google relies on to secure its own infrastructure.

In this spirit, Google Cloud has introduced Google AI Threat Defense, an automated security system designed to help you continuously monitor for and stop AI-powered threats before they can impact your business. By externalizing our internal defense capabilities, we aim to ensure that enterprises of all sizes have the tools they need to proactively protect themselves from the next generation of AI-driven threats.

Learn more about Google AI Threat Defense

Contact us to get started today

Google Cloud Security