Safer with Google

Securing the Future: Policy Recommendations for AI and Cyber Defense

The rapid advancement of frontier models has altered the global cyber threat landscape, making the deployment of autonomous AI cyber defenses an urgent necessity. Google recommends governments prioritize rapid vulnerability remediation across critical infrastructure, adopt AI-driven cyber defenses and cloud technologies, and foster robust international collaboration and standardized security benchmarks for state of the art cybersecurity AI models.

The global cybersecurity landscape is undergoing a fundamental transformation. For decades, cyber defenders have been trapped in a cycle of reactive patching of software and operating systems. Defenders must protect an increasingly complex terrain and need to be successful at all times, where an attacker only needs to succeed once. We call this dynamic the "Defender's Dilemma".

Defenders can achieve a security advantage by deploying AI for vulnerability management and systems hardening – using AI code scanning tools to draw down the number of latent vulnerabilities in their code and integrating AI-assisted development tools to fix software bugs as the code is being written and compiled, thereby preventing new vulnerabilities from ever reaching the end product where they can be exploited. Autonomous code scanning and remediation is increasingly important as more software is "vibe-coded," or written by AI agents through natural language prompts.

At Google, we helped pioneer the use of AI for cyber defense and strong security standards for AI models. We are a proud industry leader in the deployment of AI tools for cyber defense across our platforms and products.

We envision a world where most newly created software is secure-by-design, where vulnerabilities are autonomously discovered and remediated by defensive AI agents, and the wider ecosystem is safer and more secure.

However, the same AI technologies could pose risks to the world's software and critical infrastructure if it falls into the wrong hands. Much of the world's critical infrastructure remains reliant on legacy systems with years of technical debt that are especially vulnerable to AI-assisted adversaries. The average time it takes for an attacker to first exploit a vulnerability after its initial disclosure, has fallen from 1.3 years in 2020 to only 1.6 days in 2026, largely driven by advances in AI tooling. It's therefore crucial that the ecosystem works together to evolve our defenses for the AI era, while also balancing the need for rapid innovation to ensure we remain ahead of these threats.

This paper outlines Google's policy recommendations for this era of advanced AI cyber capabilities, detailing the evolving threat landscape, our integrated AI defense ecosystem, and the global policy shifts required to ensure AI bolsters global cybersecurity.

The Evolving Threat Landscape

The transition to AI-integrated cyber threat activity introduces complexities that demand a new defensive architecture. Per Google Threat Intelligence Group's May 2026 AI Threat Tracker Report, attackers are no longer limited by human bandwidth – they are increasingly leveraging AI models and agentic AI tools to identify and exploit vulnerabilities. This shift has the following characteristics: AI tools to identify and exploit vulnerabilities. This shift has the following characteristics:

- ### Increased Velocity of Vulnerability Discovery

  Multiple AI systems can now scan massive codebases at machine speed, identifying subtle logic flaws and memory corruption issues that would take human researchers months and years to find. Open source maintainers and Google's vulnerability reward programs have reported a sharp increase in valid vulnerability reports since the start of 2026.

- ### Enhanced Exploit Generation

  Threat actors have started using AI to develop exploits, and state-of-the-art AI models have proven capable of reverse engineering and chaining exploits.

- ### The Rise of Agentic AI in Cyber Attacks

  Attackers are using agentic AI to operationalize multi-stage attack chains. This includes streamlining reconnaissance, vibe-coding novel exploits, and generating code that enables the download and execution of malware with minimal human guidance.

- ### AI-Integrated Malware

  Adversaries are now deploying AI-powered malware that can dynamically rewrite its own code and generate decoys in an attempt to evade traditional security defenses. These tools are becoming autonomous, allowing malware to independently analyze a victim's network and execute adaptive attacks with minimal human oversight.

- ### Adversarial Distillation Accelerates Threat Proliferation

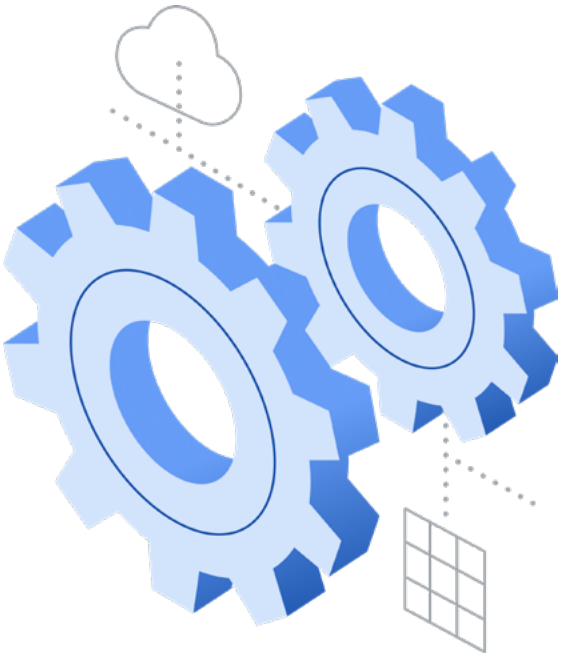
  The risks are further amplified by a rise in distillation attacks, which lowers the economic and technical barriers to entry for threat actors. Bypassing the resources, time, and talent needed to develop advanced capabilities, adversaries are extracting state-of-the-art capabilities directly from frontier models. Adversarial distillation allows malicious actors to bypass traditional barriers to entry and rapidly gain access to advanced dual-use AI capabilities.

To mitigate against these risks, Google is committed to safe model development as embodied in our Frontier Safety Framework, a set of protocols that aims to address severe risks that may arise from the high-impact capabilities of frontier AI. We develop threat models to identify potential vulnerabilities and create new evaluation and training techniques to address misuse. We investigate abuse of our products, services, users, and platforms, including malicious cyber activities by government-backed threat actors, and work with law enforcement to disrupt adversaries when appropriate. Moreover, our learning from countering malicious activities are fed back into our product development to improve safety and security for our AI models.

Google's AI for Cyber Defense Ecosystem

Google's AI innovation is anchored in a long-standing commitment to Secure-by-Design. Our Secure AI Framework (SAIF) provides a standardized, secure-by-design approach for industry to build and deploy AI systems that are resilient to manipulation and secure against traditional cyber threats. SAIF is modeled after Google's internal processes with which we build and deploy our models, agents, and systems.

With AI increasing the pace of vulnerability detection and exploitation, our philosophy is that defense needs to move as fast as offense. That's why we launched Google AI Threat Defense: a solution that combines the best of Google's world-class cybersecurity tools and services to help organizations accelerate their security processes and prepare for threats operating at machine speed. Google AI Threat Defense fuses the reasoning power of Gemini and other frontier models, the contextual risk prioritization of Wiz, the code remediation capabilities of Gemini and CodeMender, and the frontline expertise of Mandiant.



CodeMender: Autonomous Vulnerability Remediation

CodeMender represents a shift toward autonomous security engineering. It is Google's defensive AI agent that not only debugs vulnerabilities within codebases but autonomously generates, tests, and suggests patches for those vulnerabilities. This reduces the time-to-remediation from days to minutes.

Mandiant: Frontline Expertise

Mandiant serves as the intelligence engine of our defense ecosystem, feeding unparalleled frontline incident response data directly into Google's AI models. By combining real-world visibility with our AI capabilities, we empower defenders to validate true vulnerabilities and proactively neutralize threats based on the latest adversarial tactics.

Wiz: Comprehensive AI & Cloud Security

Wiz helps defenders connect code, cloud, and runtime into a single end-to-end security picture to automate risk reduction and threat response, and enable security teams to operate at AI speed. Wiz's Red, Green, and Blue agents autonomously scan customers' environments for vulnerabilities, investigate potential threats to determine severity, and suggest remediations and compensating controls.

Global Policy Recommendations

The technological shift must be matched by a policy shift. We urge governments and international bodies to develop comprehensive strategies to increase preparedness for advanced cyber threats through efforts to accelerate vulnerability discovery and remediation, modernize cyber defense with AI, and enhance cross-sector coordination and information sharing:

Accelerate Vulnerability Discovery and Remediation

Harden Critical Infrastructure Against AI Threats

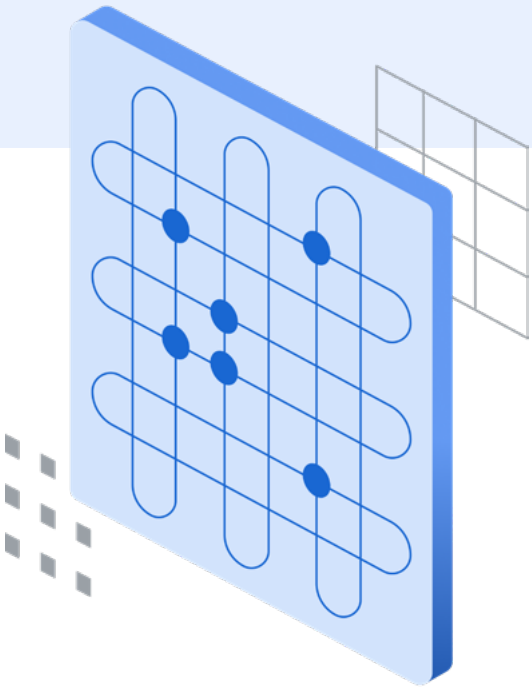
Use policy and procurement levers to require government entities and critical infrastructure providers to reduce their backlog of high-severity and critical vulnerabilities. Conduct a 90-day sprint to remediate known vulnerabilities. In an era of AI-accelerated attacks, “vulnerability debt” is a national security risk, while AI-enabled defenses can assist in speeding up remediation times.

Strengthen Open Source Security

Open source software is integral to the global software supply chain. Governments should assess their software inventories to understand the open source dependencies of government-managed and critical infrastructure systems. It is critical to work with industry and open source groups, such as the Linux Foundation and the Alpha Omega project, to prioritize the scanning of key open source libraries for immediate vulnerability identification and remediation. This can help support low-resource entities that maintain key digital public goods and reduce duplicate efforts amongst industry.

Increase Societal Preparedness and Fund Resource-Constrained Sectors

Prepare government entities, businesses, and critical infrastructure operators for reducing exposure to AI-enabled threats and adapting to a faster cybersecurity cadence. Provide funding to support vulnerable, resource-constrained sectors, such as rural critical infrastructure providers that have traditionally been underserved. Hold consultations to create buy-in of approaches to vulnerability remediation, and delineate clear roles and responsibilities. Communicate the importance of security best practices, including phishing-resistant multi-factor authentication, zero trust architecture, principle of least privilege, etc.



Modernize Cyber Defense with AI

Arm Defenders with AI

Governments should adopt and actively incentivize autonomous defensive AI capabilities for defenders to analyze threats and remediate vulnerabilities at machine speed. This includes utilizing different models and tools based on the varying capabilities and cost considerations of defenders. Governments should also incentivize the procurement of software that is secure-by-design from vendors that have integrated agentic AI tools in their software development lifecycle for automated patching.

Eliminate Barriers to Adopting State-of-the-Art Cyber Defense

Policymakers should reconsider regulatory barriers that could inadvertently slow or limit government or critical entities from accessing agentic cyber defense technologies at scale.

Enhance Security Baseline through Technology Modernization

Many public and private sector entities can uplift their security baselines by migrating to modern cloud-based platforms and software (i.e., PaaS and SaaS), which deliver strong inherited security controls and push software updates to customers automatically. Organizations managing aging or end-of-life systems should identify opportunities to either migrate assets to the cloud, or alternatively, deprecate systems that cannot reasonably be migrated to the cloud or protected using compensating controls.

Enhance Cross-Sector Coordination and Information Sharing

Establish Common Cyber Benchmarks

Governments should partner with industry and international bodies, such as the International Network of AI Safety Institutes and the Frontier Model Forum (FMF), to establish publicly available, standardized, and globally recognized evaluation frameworks for assessing the offensive and defensive cyber capabilities of frontier AI models. Developing these shared benchmarks is critical to accurately measuring and understanding both offensive risks and defensive potential of AI.

Strengthen International Coordination and Information Sharing

Cyber threats do not recognize borders. We propose the establishment and promotion of common standards and international norms for AI security, such as the Secure Access to Frontier AI (SAFA) Framework for secure model access, to prevent the proliferation of offensive AI tools for malicious use and foster defensive collaboration. By supporting coalitions such as the FMF and Coalition for Secure AI (CoSAI), and engaging in public-private intelligence sharing, governments can bolster the ecosystem's access to collective knowledge and best practices necessary to safely deploy frontier models and mitigate emerging security risks.

Conclusion

The era of purely manual, reactive cyber defense is over. AI is empowering our adversaries, but it also provides defenders with a potent set of tools to help secure the future.

By embracing AI-enabled defense, investing in secure-by-design architectures, and modernizing our global policy frameworks, we can flip the script and ensure that defenders have the upper hand.

The transition to an AI-driven security paradigm requires more than just technology; it requires public-private partnerships and international collaboration. Together, we can use AI as the ultimate "force multiplier" for defense, protecting the digital ecosystem for everyone.