Testimony of Sandra Joyce
VP, Google Threat Intelligence
Hearing on "The AI Security Landscape: How Frontier Models, Agentic AI, and AI Coding Tools Are Reshaping Cybersecurity and Critical Infrastructure Resilience"
House Committee on Homeland Security, Subcommittee on Cybersecurity and Infrastructure Protection
June 4, 2026

Chairmen Garbarino, Ogles; Ranking Members Thompson, Ramirez; and Members of the Committee and Subcommittees: thank you for the opportunity to speak with you today. My name is Sandra Joyce, and I serve as Vice President of Google Threat Intelligence Group (GTIG). Our team relentlessly defends Google, our users, and customers by building the most complete threat picture to disrupt adversaries.

Thank you for holding this important hearing. We welcome the opportunity to provide information about Google's efforts to use artificial intelligence to strengthen cyberdefense and enhance our collective security.

Artificial Intelligence and Cybersecurity: Identifying Opportunities and Mitigating Risks

We stand at a critical technological inflection point. Rapid advances in artificial intelligence (AI) are unlocking new possibilities for the way we work and accelerating innovation in science, technology, and beyond. This technology has impacted cybersecurity in profound ways for the defender as well as the attacker.

While recent introductions to AI vulnerability research have caught the attention of mainstream audiences, Google has long anticipated and prepared for this impact. Over the years, our security teams and frontier AI research teams at Google DeepMind have introduced tools like CodeMender, which have shown the immense potential of AI to discover and mitigate vulnerabilities. For years we have used these tools to review and harden the open source software that many of us depend on.

While Google leveraged AI for defense, we anticipated that threat actors would abuse these capabilities to find and exploit vulnerabilities.[1] Those concerns were validated recently when we

[1] Google Threat Intelligence Blog | Google Cloud

discovered a criminal actor had used AI to develop a zero-day exploit.[2] We expect threat actors to continue attempting to use this technology to their advantage.

Over the past year, GTIG has identified an important shift, with adversaries not only leveraging AI for productivity gains, but successfully adopting AI to significantly enhance the scale, speed, and sophistication of their operations.

Looking ahead, we are particularly concerned about two future scenarios:

- **A deluge of vulnerabilities:** Though AI will be used by defenders to harden software and produce safer code, adversaries may have the initiative in the short term to find and exploit vulnerabilities at scale.
- **A swifter, scaled adversary that will overwhelm contemporary security:** Agentic orchestration allows threat actors to cheaply scale their operations and operate at unprecedented speed to take advantage of slow patch cycles, beleaguered security teams, and human response time.

An Evolution in Vulnerability and Exploitation

Along with Google, several of our peers have validated the potential of this technology to perform vulnerability and exploit development. Because exploits are the top attack vector in intrusions we observe,[3] this evolution has enormous consequences.

Though newer AI models can be used to find these vulnerabilities, less sophisticated models have been maliciously used for this purpose when coupled with a purpose-built harness, which is software designed to wrap around the model and make it operational. Adversaries do not require unprecedented breakthrough model capabilities; instead, they are deploying sophisticated harnesses to achieve an automated research and development capability.

Recently, GTIG discovered a campaign where cybercriminals utilized an AI model to support the discovery and weaponization of a zero-day vulnerability.[4] The actor planned to leverage the zero-day in a mass exploitation scheme. Google worked directly with the impacted vendor to responsibly disclose and patch this vulnerability before mass exploitation could occur.

Though this criminal example is the only confirmed development of a zero-day exploit using AI by a threat actor that we have thus far observed, we can infer that states have been doing

[2] Adversaries Leverage AI for Vulnerability Exploitation, Augmented Operations, and Initial Access | Google Cloud Blog
[3] M-Trends 2026 Executive Edition
[4] Adversaries Leverage AI for Vulnerability Exploitation, Augmented Operations, and Initial Access | Google Cloud Blog

significant research in this area and we expect that other zero-days developed with AI are already in use.

We are working to integrate AI directly into the development cycle and make code more difficult to exploit than ever; however, this transition period presents challenges. As we harden existing software with AI, threat actors will simultaneously use it to discover and exploit novel vulnerabilities.

The Agentic Shift

The most critical structural shift observed over the last year is the shift from users accessing LLMs through manual prompts to users deploying agents that can operate autonomously. Just as businesses are exploring the capabilities of agentic AI to automate their workflows, threat actors are experimenting with agentic capabilities to automate malicious activity, such as persistently probing a target or carrying out research and development on their behalf.

We recently observed a suspected PRC-nexus threat actor deploying agentic capabilities against Asian tech firms.[5] By incorporating open source tools and a memory system, the agent could map the target and autonomously pivot between tools based on its internal reasoning. Simultaneously, the actor leveraged a multi-agent penetration testing framework to automate the identification and validation of vulnerabilities. This approach suggests a transition toward autonomous reconnaissance that can scale the probing of targets with minimal human oversight.[6]

In another incident, we observed the North Korean actor, APT45, sending thousands of repetitive prompts to recursively analyze exploits.[7] This resulted in a more robust arsenal of exploit capabilities that would be impractical to manage without AI assistance.

In addition to the scale advantage conferred by agentic capabilities, threat actors are able to move rapidly before and after gaining access to a network using AI. Threat actors are using AI to take advantage of recently disclosed vulnerabilities before patches can be applied. After gaining access, agentic attacks can move rapidly through a network. In both cases, industry standards are based on human response time and insufficient to mitigate the threat.

---

[5] Adversaries Leverage AI for Vulnerability Exploitation, Augmented Operations, and Initial Access | Google Cloud Blog
[6] GTIG AI Threat Tracker: Distillation, Experimentation, and (Continued) Integration of AI for Adversarial Use | Google Cloud Blog
[7] Adversaries Leverage AI for Vulnerability Exploitation, Augmented Operations, and Initial Access | Google Cloud Blog

Defending the AI Supply Chain

As organizations continue integrating LLMs into production environments, the AI software ecosystem has emerged as a primary target for exploitation. While frontier models themselves remain highly resilient to direct compromise, adversaries are deploying traditional supply chain tactics against the orchestration layers—including open-source wrapper libraries, API connectors, and configuration tools.

The implications of this vector were demonstrated by a prominent cybercriminal cluster tracked by Google as UNC6780 (also known publicly as TeamPCP). The actor claimed responsibility for multiple supply chain compromises of popular GitHub repositories and associated GitHub Actions.[8] The compromise highlights the expanding attack surface of AI platforms and the potential for impact across the software supply chain. Given the package's widespread use, this incident could lead to considerable exposure of AI API secrets from affected victims, which could be used to gain further access to systems for traditional intrusion operations.

To help mitigate these supply-chain risks, we recommend that AI agents integrate automated security scanning directly into their public skill marketplaces. Every skill published to the repository should be analyzed to detect unauthorized network operations, malicious payloads, or unsafe embedded instructions. Based on this security-focused analysis, skills should be either approved as benign, flagged with user warnings, or blocked entirely, providing an essential layer of defense against ecosystem abuse.

Protecting the integrity of this supply chain cannot be treated as a standard, retroactive patch-management exercise; it requires a coordinated, structural shift in governance and defense. To secure these critical building blocks, organizations must establish comprehensive incorporation of AI capabilities into software supply chain security practices, including efforts to ensure transparency into model lineage, training datasets, and orchestration components. Organizations must also enforce strict least-privilege, zero-trust guardrails around AI data pipelines, ensuring autonomous agents operate in segmented environments without authority to elevate permissions or communicate with untrusted networks.

Autonomous Cyber Defense: Transitioning to Continuous, Lifecycle Security

Finding a vulnerability is only a single element of a broader operational challenge; to effectively tilt the cybersecurity balance in favor of defenders, we must close the exploit window entirely. Historically, patch management has been a retroactive, human-paced race against adversaries.

[8] Adversaries Leverage AI for Vulnerability Exploitation, Augmented Operations, and Initial Access | Google Cloud Blog

In the current threat landscape, where attackers use AI to discover and target design flaws at scale, traditional, siloed security tools fail to keep pace.[9] Our threat intelligence demonstrates that modern digital risk is no longer confined to isolated software code errors. Real-world attack paths routinely emerge from the complex, real-time friction between cloud application interfaces, infrastructure configurations, permissions, and network identities. If a security team is handed an unprioritized list of thousands of software flaws, they face immediate patch fatigue. For critical infrastructure operators and public sector networks, defense-at-scale requires an automated mechanism that shifts focus away from mere bug hunting and toward comprehensive environmental exposure management.

The Operational Framework: An Autonomous Defensive Control Loop

To address this collapse of the exploitation timeline, Google has pioneered an always-on four-step framework designed to help enterprises to implement an autonomous defensive control loop: Prepare, Scan & Prioritize, Remediate, and Monitor.[10] Our aim is to shift the industry away from reactive response and toward active prediction and accelerated remediation.

- **Proactive Attack Surface Reduction (Prepare):** Before a vulnerability is ever targeted, defensive systems must establish a real-time exposure map of an entire network. By utilizing AI-driven, context-aware simulation agents, defenders can continuously pressure-test their own infrastructure. These agents behave exactly like an active adversary, mapping out complex, multi-stage attack paths to discover whether a sensitive asset is actually reachable from untrusted external networks. If a flaw cannot be reached, its immediate risk drops, allowing organizations to aggressively shrink their actual, real-world attack surface without relying on manual triage.

- **Contextual Risk Validation (Scan & Prioritize):** When new vulnerabilities are surfaced by AI code security agents—such as CodeMender, which leverages large language models to locate dormant, pre-authentication flaws—they must be cross-referenced with live operational data. By merging deep threat intelligence with cloud-posture data, AI-driven defense identifies which vulnerabilities represent an existential threat to critical workflows, sorting out benign software anomalies from active crisis points.

- **Machine-Speed Mitigation (Remediate):** Once an exploitable, high-priority risk is validated, autonomous code remediation layers—operationalized through autonomous security agents like CodeMender—step in to analyze the underlying architecture. These agents automatically generate, validate, and apply secure code patches directly to software libraries at machine speed. This effectively eliminates the human bottleneck,

[9] Defending Your Enterprise When AI Models Can Find Vulnerabilities Faster Than Ever | Google Cloud Blog
[10] Introducing Google AI Threat Defense | Google Cloud Blog

allowing critical systems to self-heal before an agile adversary can capitalize on a known vulnerability.

- **Continuous Detection (Monitor):** Even with a hardened foundation, true resilience across critical infrastructure requires constant vigilance during runtime. While pre-deployment, code-level scanning pipelines are excellent at catching flaws before software is pushed live, they are fundamentally incapable of blocking an active, zero-day exploit in real time. To counter this, our framework shifts defensive operations away from solely manual, human-paced oversight and toward machine-speed detection and real-time behavioral defense guided by frontline threat intelligence. Entities should utilize specialized, autonomous agents to triage suspicious behavior, and respond to live network intrusions at machine speed and an agentic Security Operations Center (SOC) functionality, to automate the detection, investigation, and tracking of emerging anomalies across their complex network, identity, and application telemetry.

By seamlessly tying real-world network exposure to automated, intelligent patch generation, we are establishing a comprehensive blueprint for modern cyber resilience. This continuous loop ensures that private and public entities can finally outpace the scaling volume and velocity of AI-driven adversarial operations.

Securing Artificial Intelligence through Bold and Responsible Innovation

We believe our approach to the frontier of artificial intelligence must be both bold and responsible. This means developing and deploying technology in a way that maximizes positive societal benefits while proactively engineering systems to withstand and mitigate modern adversarial pressures. For more than 20 years, Google has pioneered a Secure by Design approach, meaning we embed security into every phase of the software development lifecycle - not just the beginning or the end. Guided by Google's core AI Principles—originally published in 2018 and systematically updated to address the changing technology ecosystem—we design our AI systems from the ground up with robust security measures and strict safety guardrails.

Google's software and AI development pipeline relies on advanced threat modeling to proactively identify emerging threat trends and systemic risks, and to explicitly design our products for inherent safety. Rather than treating security as an afterthought, we continuously enhance safeguards inside our active products to offer scaled, adaptive protections to enterprise users and critical infrastructure operators across the globe.

GTIG partners closely with Google DeepMind to feed lessons learned from countering malicious activity directly into engineering processes. This creates positive feedback loops and

allows us to continuously improve the baseline safety and security of our AI foundation models. These intelligence-driven enhancements are applied dynamically at both the input/output classifier levels and deep within core model architecture. This rapid integration process is essential to maintaining extreme agility in our defensive postures and preventing sophisticated threat groups from abusing our technologies.

Conclusion

Cybersecurity has never been an environment where absolute perfection is possible. It will remain a fiercely contested, highly dynamic domain for years to come, demanding continuous innovation, speed, and structural agility to defeat adaptive adversaries.

As this Committee looks to secure our homeland and fortify the digital architecture supporting American critical infrastructure, Google stands ready to serve as a committed, transparent partner. By combining public-sector authority with private-sector technical innovation, we can harness the immense potential of artificial intelligence to tip the scales of cybersecurity permanently in favor of the defender.

Thank you for the opportunity to testify today. I look forward to answering your questions.

***