Google Cloud's

Product Vision for AI-Powered Security

May 2024

Table of Contents

Written by

Umesh Shankar
Chief Technologist, Google Cloud Security

Steph Hay
Senior Director, Google Cloud Security

Our vision for AI is to accelerate your ability to protect and defend against threats by shifting from manual, time-intensive efforts to assisted and, ultimately, semi-autonomous security — while providing you with curated tools and services to secure your AI data, models, applications, and infrastructure. We do this by empowering defenders with Gemini in Security, which uses SecLM, our security-tuned API, as well as providing tools and services to manage AI risk to your environment. Our Mandiant experts are able to help you secure your AI journey wherever you are.

Executive summary

The number and sophistication of cybersecurity attacks continues to increase, but generative AI (gen AI) has the potential to tip the balance in favor of defenders, with security agents providing help across every stage of the security life-cycle: prevention, detection, investigation and response. At Google Cloud Next 2024, we announced new AI-driven innovations across our security portfolio to help deliver stronger security outcomes for customers and enable organizations to make Google a part of their security team. In this paper, we explain our vision behind these innovations as we continue working to empower defenders with AI-powered security and look forward to working together to bring us closer to a safer, more secure and trusted digital world.

Introduction

The last year saw generative AI go from a big research advance to something that directly impacts real people and products. The multidimensional interplay between gen AI and security means we must protect AI-powered workloads against old and new risks, and apply gen AI to solve security problems. The resulting collisions pose new challenges to overcome:

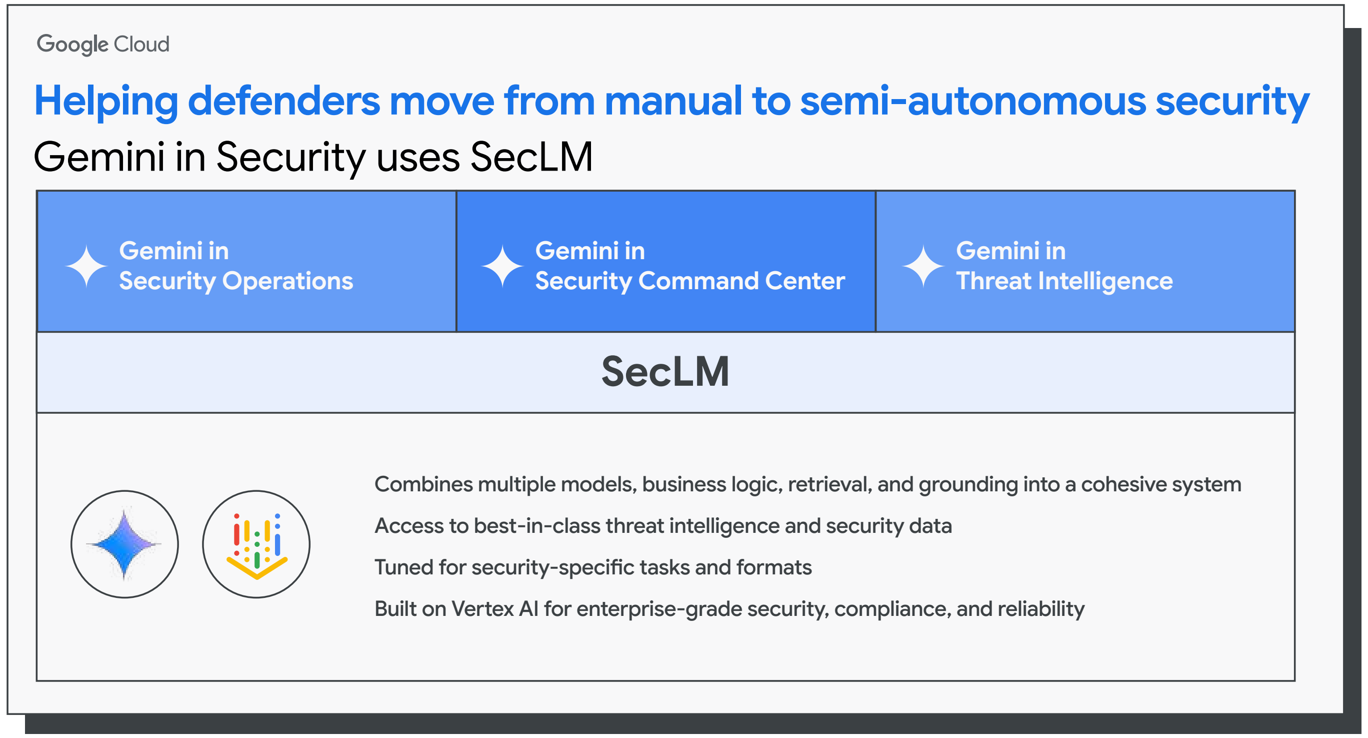
People expect AI to deliver on the initial hype with enduring value. Organizations and users expect AI-powered assistance to be more than a simple productivity boost; it needs to consistently provide correct answers and insights, and maintain context to actually get smarter over time.

Foundation Models aren't enough on their own to solve many real-world security problems. Organizations and practitioners have realized that tackling most tasks requires using multiple—often specialized—models, planning, orchestration, extensions, external memory, and other techniques together. Some problems may require open-ended, long-running agents that can run indefinitely to advance specific goals.

Amid these challenges, we continue to focus on the macro problems facing security today: threats, toil, and talent. With gen AI, we are aiming to cut down on the ever-expanding number of threats, reduce the toil security professionals face in doing so manually, and bridge the cyber talent shortage. And we believe we can continue to tackle those problems by setting ambitious goals for 2024 and beyond across our portfolio:

- Evolving the security life-cycle to be semi-autonomous, with Gemini
- Powering security use cases with our not-so-secret sauce: SecLM

The diagram below shows how we are infusing generative AI capabilities into Google Cloud Security products with SecLM, a security-specialized API.



Google Cloud

Helping defenders move from manual to semi-autonomous security

Gemini in Security uses SecLM

Gemini in Security Operations

Gemini in Security Command Center

Gemini in Threat Intelligence

SecLM

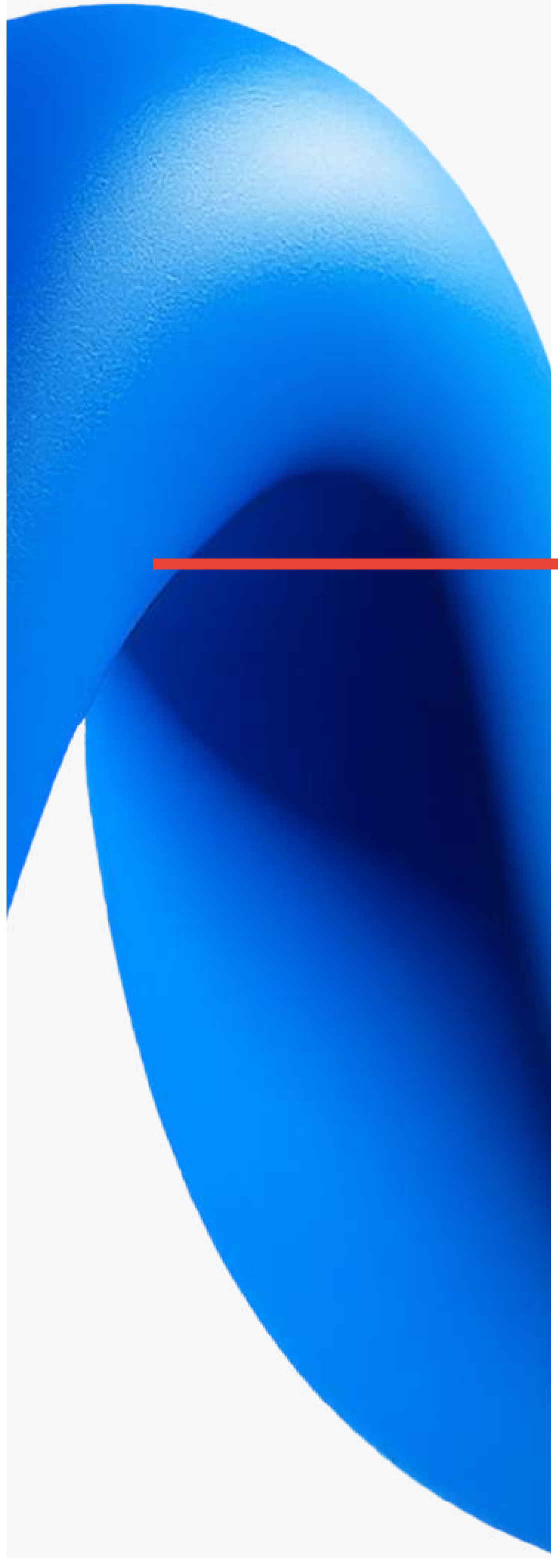
Combines multiple models, business logic, retrieval, and grounding into a cohesive system

Access to best-in-class threat intelligence and security data

Tuned for security-specific tasks and formats

Built on Vertex AI for enterprise-grade security, compliance, and reliability

In the following sections, we'll detail our strategy toward these goals and how we'll get there.



Evolving the security life-cycle to be semi-autonomous

In 2023, we imagined a future state in which every customer had a virtual security assistant who knows the user and their goal—having been trained by the world's leading security experts—and recommends the best path to deliver on that goal. This approach primarily removes manual toil from the security life-cycle, but it's not enough to transform security for good.

Climbing the ladder of autonomy

Gemini for Google Cloud is a new generation of AI assistants for developers, Google Cloud services, and applications. These assist users in working and coding more effectively, gaining deeper data insights, navigating security challenges, and more.

To realize its full value, Gemini will have to evolve the practice of security from primarily manual-driven toward more autonomous operation, so humans can operate higher up the value chain.

Accordingly, we define four levels of autonomy as follows:

01 **Manual**:
Humans cyclically and routinely execute tasks along a common security life-cycle: prevent issues from happening; detect issues that are happening; prioritize which issues require action; and respond to the most pressing issues to mitigate any fallout.

02 **Assisted:**
Humans still do their typical job, but Gemini boosts their productivity by answering questions, retrieving and generating insights, performing tasks on demand, and recommending next steps to take to achieve a goal.

03 **Semi-autonomous**:
Humans begin to offload mission-critical tasks to Gemini, which drives outcomes across each phase of the task life-cycle consistently well. People are still involved where the AI hasn't yet become trusted, or where it cannot automate with sufficiently high confidence or precision.

04 **Autonomous**:
Gemini is a trusted assistant driving the security life-cycle to positive outcomes on behalf of the customer. The customer interacts with the system primarily to set or update their goals and preferences.

The shift from manual to assisted is already underway. For example, Gemini in Security Operations translates natural language inputs to specialized syntax so a customer can search across their event data, and summarizes case meta-data for active alerts and recommends how to respond. Gemini in Security Command Center similarly summarizes complex security data–including simulated attack paths–and recommends the next steps to remediate areas of risk. These assistive features surface to the user in existing UIs (e.g., buttons, icons, input fields, and text show up by default or invoked by the user) and as multi-turn chat (e.g., Q&A, search refinement, and action selection or execution).

Moving to semi-autonomy poses an even greater opportunity to transform the security experience as Gemini takes primary responsibility for many security tasks, only asking for human assistance where needed.

That opportunity forms the basis of our vision:

Gemini evolves the security life-cycle to be semi-autonomous

Why not full autonomy? It will take time to build the necessary trust in the system. We believe that most customers will want the ability to be in the loop on important decisions. Quality is our guide to when we can climb the ladder of autonomy; until the system can be relied upon to make good decisions consistently for a given task, we will not advance. Still, we are already exploring use cases where full automation may be possible, for example running a response playbook for low-value repetitive issues or clustering alerts into prioritized cases.

Roadmapping our way to semi-autonomy

An example of this vision beginning to manifest is Gemini in Security Operations, which now conversationally assists analysts through their investigations from detection to response within the existing platform experience. Gemini automatically retrieves and incorporates context relevant to the investigation—both from the customer's unique data plus enrichments from Mandiant and VirusTotal threat intelligence sources—and navigates the user to the most relevant UIs for more detailed information. Gemini makes recommendations on what to do next, such as creating a new detection or running a playbook, and also creates new playbooks from natural-language inputs.

These capabilities help remove high-toil tasks like navigating multiple data sources (often across different surfaces) or manually constructing queries, detection rules, and response playbooks. And they lay the foundation for semi-autonomous workflows that offload the main workflow to AI, keeping people in the loop for key decisions.

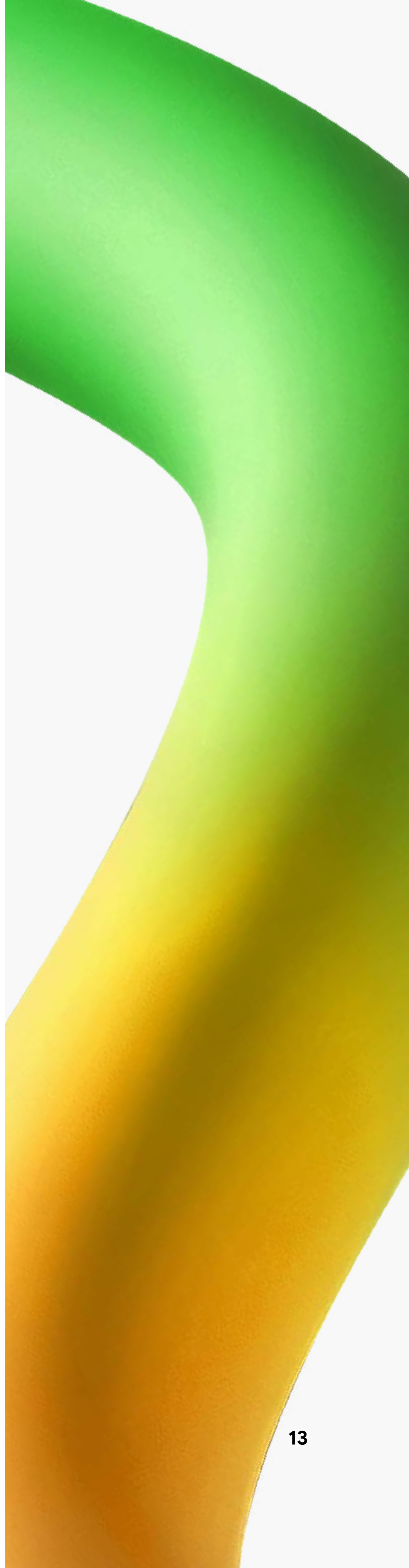
Gemini in Security Command Center leverages the same foundation for powering risk analysis and remediation for Cloud Security across multi-cloud footprints. For example, Security Command Center's Risk Engine constructs a graph-based model of a customer's multi-cloud estate for attack path simulation, which Gemini summarizes to identify the most vulnerable areas of risk to tackle first. Gemini also summarizes alerts on critical and high-priority misconfigurations and vulnerabilities plus recommends the next steps to remediate, including running automated playbooks.

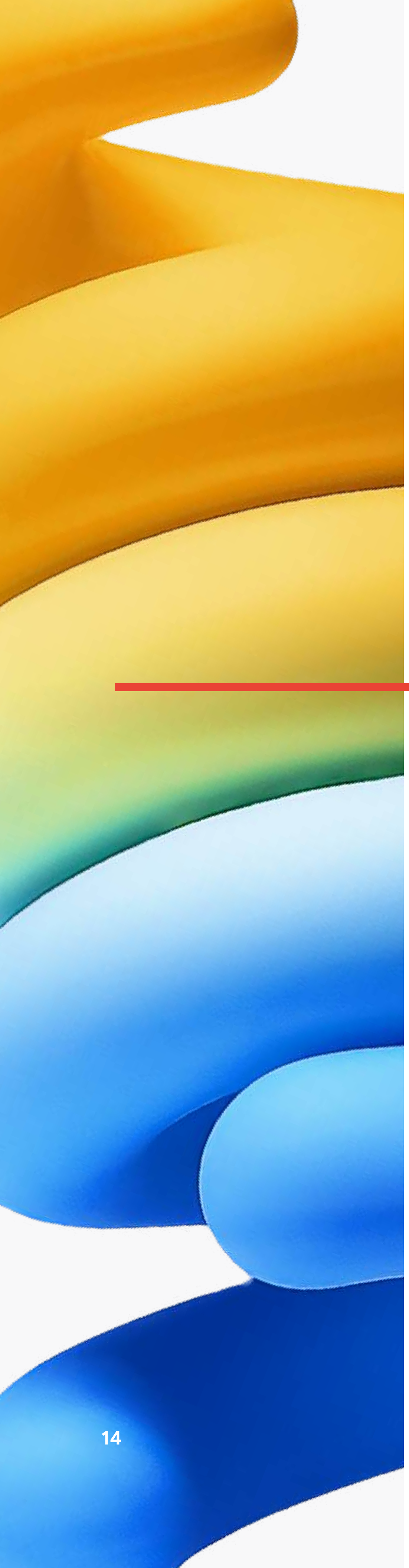
Another example is with Gemini in Threat Intelligence; last year we announced Code Insight in VirusTotal, which can identify malicious code in more than 200 different types of files. With the advancement of the Gemini Pro 1.5 model's 1 million token context window, this capability now can reverse engineer entire binaries in one go, identifying malicious aspects of code and recommending responses. Indeed, this advancement marks the initial steps toward Gemini's autonomy in malware analysis.

Gemini in reCAPTCHA now helps organizations intuitively understand their fraud risk by analyzing logs to find patterns indicative of global and client-specific fraud and surfaces those insights to customers in digestible summaries. This information can help businesses identify and take action against attackers more quickly, reducing the financial and reputational impact of fraud.

These advances are made possible by leveraging Google DeepMind Research techniques, which we use across our products even outside the context of Gemini assistance. For example, our Assured Open Source Software uses generative AI for code generation and automatic test-harness creation for fuzzing to find zero-days across hundreds of OSS packages. We are continually exploring the application of novel techniques such as Graph Neural Nets and Time Series forecasting, to transform the processes underlying security tasks that we want to automate–including threat detection, alert precision, and User and Entity Behavior Analytics (UEBA).

We also are leveraging agentive technologies; to achieve semi-autonomy, Gemini will orchestrate a cast of security agents (open-ended, long-running, stateful resources) that can reason about and execute processes to achieve custom security goals. These agents codify security expertise from sources including Google, Mandiant, VirusTotal, and SafeBrowsing–thereby enabling Gemini to call upon them as contextually relevant per use case, to receive the most precise and substantive information or action to deliver on the ideal outcome.

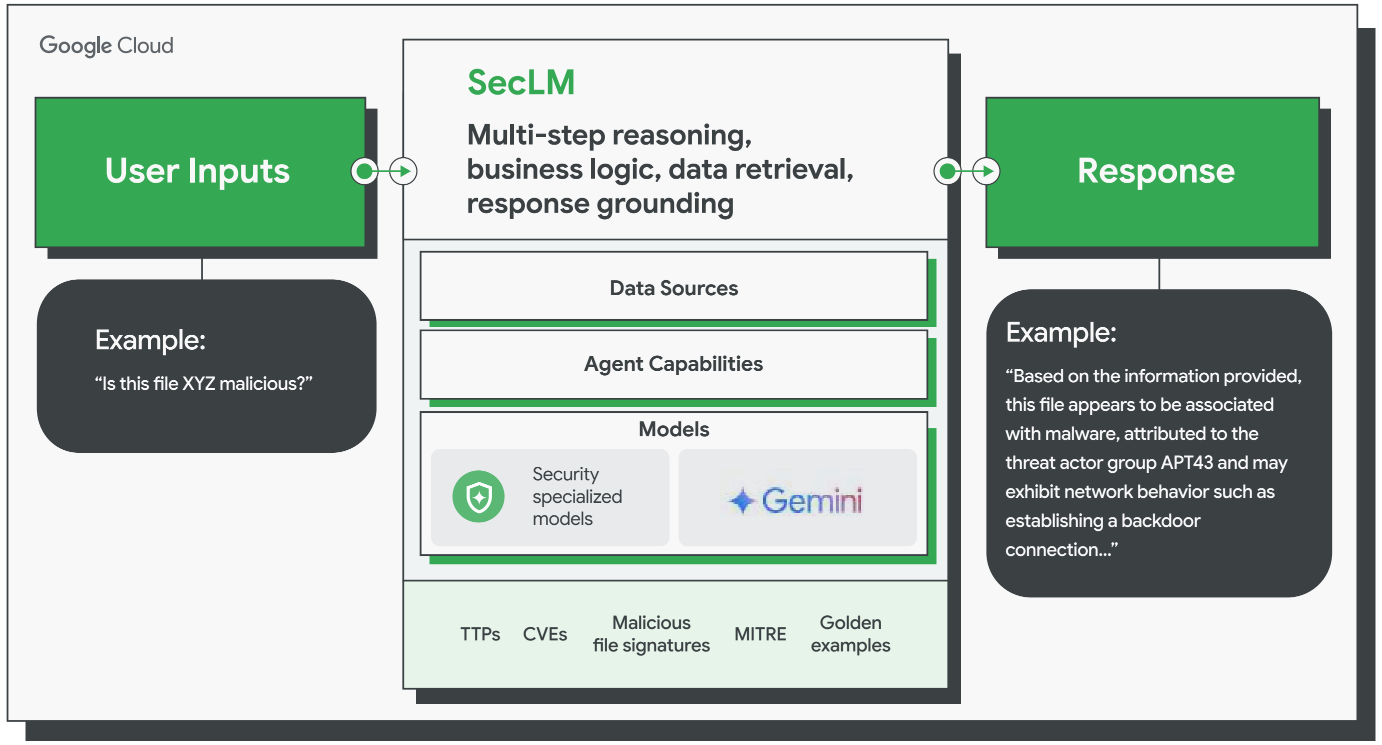




Powering security use cases with our not-so-secret sauce: SecLM

General purpose foundation models are helpful to overcome challenges, but in practice are not enough on their own to solve real-world security problems. Ensuring high quality and continuity across Gemini in Security requires a platform approach–one that enables our products to deliver a consistent experience across surfaces.

To gain horizontal leverage across Gemini in Security, we have been building a security-specialized AI API. This API, which we call SecLM, combines multiple models, business logic, retrieval, and grounding into a cohesive system that can help solve security related tasks with minimal input. SecLM is tuned for security-specific tasks and is benefiting from the latest advances in AI from Google DeepMind as well as Google's industry-leading threat intelligence and security data. By using the Vertex AI serving platform, SecLM is able to offer enterprise-grade privacy, security, and compliance guarantees—including a promise not to save customer prompts and responses or use them to train our models.



Google Cloud

User Inputs

Example:

"Is this file XYZ malicious?"

SecLM

Multi-step reasoning, business logic, data retrieval, response grounding

Data Sources

Agent Capabilities

Models

Security specialized models

Gemini

TTPs CVEs Malicious file signatures MITRE Golden examples

Response

Example:

"Based on the information provided, this file appears to be associated with malware, attributed to the threat actor group APT43 and may exhibit network behavior such as establishing a backdoor connection..."

"Google Cloud's SecLM [part of Gemini in Security] provides unique capabilities for security teams, including access to threat intelligence and simplifying complex tasks. We look forward to bringing SecLM to our customers and collaborating with Google Cloud as they continue to drive innovation with generative AI."

Prakash Venkata, Principal, PwC

We focused on a couple of core areas of value for SecLM:

Capable orchestration

- **Task coverage and output quality**. We leverage multi-step reasoning and multiple models to return the highest quality result for security and related use cases out of the box—so users of the API can provide minimal prompts in plain language.
- **Grounding**. Responses from SecLM are grounded in Mandiant Threat Intelligence, VirusTotal, and Safe Browsing data, greatly reducing hallucination risk for many security concepts.
- **Orchestration across external data sources and APIs**. SecLM calls out to external APIs and uses Retrieval-Augmented Generation under the hood to pull necessary context and return useful, grounded responses.

Ease of consumption

- **Minimal prompt engineering required**. SecLM reduces the need for prompt engineering to find the "right incantation" for system instructions, to do example retrieval for common tasks, or check that output is in the right format.
- **Reliability and security**. SecLM offers production-grade reliability, security, and compliance guarantees. It will not save customer prompts and responses or use them to train our models.

Customization

- SecLM also transforms our SecOps, Security Command Center, and Threat Intelligence offerings into open, extensible platforms that customers can leverage to extend Gemini to their surfaces. This will be a powerful platform for our SI and MSSP partners to bring tailored solutions to customers.

"SecLM [part of Gemini in Security] can help our clients transform their security operations by bringing security insights to generative AI solutions in an open and extensible platform. We see our SOC modernization efforts particularly benefiting from the faster time to value enabled by SecLM's customization and orchestration capabilities. We are excited to collaborate with Google Cloud to deliver these new generative AI innovations to our clients."

Upen Sachdev, Principal, Deloitte & Touche LLP

For more information on SecLM, listen to our recent podcast Beyond Regular LLMs: How SecLM Enhances Security and What Teams Can Do With It.

Conclusion

With rapid advances in AI technology, the art of the possible is always changing. And early signals from our customers continue to fuel our vision.

When Gemini semi-autonomously offloads the manual tasks practitioners deal with today, we free cyber talent to focus on more unique and strategic risks. And by building SecLM to be the best API for security-specific use cases, we boost the power of LLMs with specialized domain training and grounding to stop toil from expanding relative to the attack surface.

The combination of these strategies will help transform security for good, shifting the balance in favor of the defender. And that's a future we're hungry to bring about.

To learn more about our work to supercharge security with AI, visit our Security with Generative AI web page.

Google Cloud

This report includes extensive research from dozens of sources and comes in print and online versions. The online version contains links to relevant sources.