Enterprise Strategy Group by TechTarget

SHOWCASE

FEBRUARY 2025

Accelerating Secure, Private, and Compliant AI Transformation

Dave Gruber, Principal Analyst

Abstract: As organizations race to uncover where and how AI can supercharge growth while driving unprecedented efficiencies, most are concerned about the potential risks that AI can bring to the reputation, protection, and privacy of their sensitive data; their ability to meet regulatory and sovereign needs; and the security of their core operating systems and applications. Standards bodies are working in parallel to establish frameworks to guide AI program security and development, but organizations are left to fabricate their own AI implementations, slowing the realization of AI's potential. In order to rapidly iterate with AI to achieve the greatest potential, new security solutions are needed to help mitigate risks unique to AI. Solutions like AI Protection, Model Armor, and Confidential Computing from Google Cloud can help organizations confidently build, deploy, and run AI workloads by enabling the secure, private, sovereign, and compliant use of AI.

Overview: Securing the AI Era

As organizations race to gain advantage through AI transformation, a new level of balance between the speed of innovation and risk management is needed. Yet, with the AI-era moving at a breakneck pace, most are struggling to implement effective guardrails to ensure applications, data, systems, and reputations are protected. For some, internal centers of excellence (CoE) are emerging to develop strategy, rationalize decisions, and manage AI initiatives.

Securing AI

As organizations race to gain advantage through AI transformation, a new level of balance between the speed of innovation and risk management is needed. Effective guardrails are needed to ensure applications, data, systems, and reputations are protected.

The majority of organizations (85%) are already utilizing or developing large language models (LLMs) for proprietary use, whether as part of a production generative AI (GenAI) environment or an ongoing development project, while another 13% report that they are in the process of developing one.[1] Beyond LLMs, organizations are well underway in applying GenAI in areas like IT operations, software development, sales, and research. In parallel, some are also being careful about the use of public AI resources, taking early measures to block access to specific GenAI sites. Overall, most organizations have developed or are developing acceptable use policies and governance controls that might provide needed guardrails for GenAI use, though security professionals admit that these controls have abundant weaknesses.

At the same time, CISOs worry about meeting sovereignty and compliance requirements, as their organizations struggle to implement strong data and model security and governance controls. Many organizations already building GenAI applications are further concerned about data loss and brand protection due to risks such as hallucination, response inaccuracy, prompt injection, and models potentially outputting offensive content.

[1] Source: Enterprise Strategy Group Research Report, *Generative AI for Cybersecurity: An Optimistic but Uncertain Future*, April 2024. All Enterprise Strategy Group research references and charts in this showcase are from this report.

This Enterprise Strategy Group Showcase was commissioned by Google and is distributed under license from TechTarget, Inc.

© 2025 TechTarget, Inc. All Rights Reserved.

Regulatory bodies are rapidly establishing requirements to ensure safe, ethical use of AI. As these requirements become more widely enforced, security, risk, and compliance teams will need ready access to AI inventories, verification of controls, and audit data. The speed of innovation is accelerating, and inventory data will constantly be changing, which requires highly automated mechanisms to monitor all aspects of the operating environment.

Expanding Cloud Security Requirements

AI development and runtime environments add new infrastructure, data and data pipeline management requirements, and data privacy, compliance, and security requirements.

Building on top of established cloud development and security practices, AI development and runtime environments add new infrastructure, data and data pipeline management requirements, and data privacy, compliance, and security requirements.

Yet, security teams are still very early in their understanding of GenAI and need help in identifying, prioritizing, and remediating the biggest risks. They need new mechanisms that can continuously monitor and inventory AI-related applications, assets, and use to identify areas of potential risk and response mechanisms capable of mitigating that risk. Detective and preventative controls are also needed to enforce policy and identify malicious activity.

Adding complexity to the problems, these policies and associated enforcement mechanisms involve multiple teams, including app development, security, and risk management—all commonly members of CoEs.

Core Building Blocks

The three foundational building blocks of AI security include:

- The discovery and visibility of all AI-related assets across AI development environments, AI operating infrastructure, and AI data pipelines.
- Control mechanisms capable of both setting and enforcing governance policies associated with the use of AI and preventing AI-specific threats.
- Detection and investigation mechanisms that enable security operations teams to see and respond to AI threats and vulnerabilities.

Inventory of AI Assets

Security teams need the ability to continuously see and understand what AI workloads are in use within their infrastructure. This includes AI applications, models, and data sets. AI security also requires visibility into both public and private models in use, the data pipeline that feeds those models, and the configuration and operating status of all active controls used to protect and govern them. This “AI inventory” is core to the next two foundational building blocks.

Key Capabilities

- AI model discovery.
- AI data set discovery.
- AI workload discovery.
- AI application, data, and infrastructure mapping.

AI Controls

AI security controls require both proactive controls to ensure model content quality and policy enforcement and active controls that can protect applications and models from malicious activity and attacks.

An example of a proactive control is the ability to scan and detect malicious files or unsafe URLs that are contained within models in use. Because many public models are utilized, active monitoring and filtering for these same types of malicious or unsafe content could be necessary.

An example of an active infrastructure control is the ability to detect and suggest the use of confidential compute to enable encryption in use that prevents the ability for insiders or external actors to access model weights. Another example of active AI controls is the ability to monitor for outside threats, such as prompt injection and jailbreak attempts.

Key Capabilities

- Content safety: Detect categories of potentially unsafe content.
- Malicious content detection: Detect malicious files and unsafe URLs.
- Sensitive data protection: Detect personally identifiable information (PII) and filtering.
- Inappropriate data filtering: Monitor for explicit content and materials.
- Prompt safety: Monitor for prompt injection and jailbreak attempts.
- Confidential compute: Control AI workload infrastructure to prevent model weight theft.

AI Threat Investigation and Response

Security operations teams need the ability to rapidly detect, investigate, and respond to active AI-related threats. Because AI threat investigation requires access to more than the commonly used security telemetry, new visibility into model interaction, model contents, and model controls is needed.

Beyond external threats, issues involving the data pipeline may require security teams to work closely with application developers and AI model developers to respond to model data-related issues.

Key Capabilities

- Monitor AI security posture: Continuously track and understand overall posture.
- Detect AI security posture issues: Prioritize, investigate, and respond to findings.
- Detect and respond to model misuse.
- Identify and respond when sensitive data is used within AI workloads.

Ideally, all these capabilities will be integrated with primary detection and response mechanisms, enabling automation and workflow support.

Introducing Google Cloud AI Protection

AI Protection from Google Cloud helps customers accelerate their AI journey by providing end-to-end capabilities for holistically managing risk across the AI lifecycle. AI Protection helps security teams:

- Gain a clear understanding of their AI workloads.
- Mitigate risk and safeguard AI assets.
- Detect, investigate, and manage threats to AI systems.

AI Protection helps security and application development teams collaborate in mitigating risk and safeguarding AI workloads across clouds. It also provides a centralized AI security posture view. Organizations can access AI Protection capabilities through Security Command Center, which enables them to discover AI inventory, improve security and compliance, and respond to threats targeting GenAI applications regardless of whether they use first-party, third-party, or open source models. Capabilities in AI Protection, like Sensitive Data Protection and Model

Armor, help organizations monitor for sensitive data used in training sets and mitigate risks such as sensitive data leakage, prompt injection, and offensive content. With AI-specific intelligence from Mandiant frontline analysts, teams can detect and respond to threats against their AI applications and systems with confidence.

Google further offers Mandiant AI Security Consulting services to help organizations secure their use of AI. Mandiant consultants help security teams assess risks to AI systems by evaluating the end-to-end security of AI deployed in production to identify weaknesses before attackers do.

Table 1. Google Cloud Security: AI Protection

Discover AI Inventory	Secure AI Assets	Manage AI Threats
Gain a clear understanding of potential issues in the environment	**Mitigate risk and safeguard AI assets by implementing controls, policies, and guardrails**	**Detect, investigate, and respond to threats against AI systems**
Integrated in Security Command Center for centralized management of AI risks		

Source: Google Cloud Security and Enterprise Strategy Group, a division of TechTarget, Inc.

Conclusion

Security leaders face massive responsibility in ensuring the secure development and operations of AI applications and the data used within them. At the same time, they have an opportunity to become a key enabler of AI transformation within their broader functional organization. Establishing early, foundational AI security infrastructure and controls will enable leaders to grow and keep pace with the rapid innovation occurring within line-of-business functions.

Informa TechTarget's Enterprise Strategy Group recommends security leaders and AI CoEs explore how solutions like AI Protection and Model Armor from Google Cloud can accelerate AI transformation, enabling continuous visibility and control while empowering developer teams to build and deploy AI systems without friction.

©TechTarget, Inc. or its subsidiaries. All rights reserved. TechTarget, and the TechTarget logo, are trademarks or registered trademarks of TechTarget, Inc. and are registered in jurisdictions worldwide. Other product and service names and logos, including for BrightTALK, Xtelligent, and the Enterprise Strategy Group might be trademarks of TechTarget or its subsidiaries. All other trademarks, logos and brand names are the property of their respective owners.

Information contained in this publication has been obtained by sources TechTarget considers to be reliable but is not warranted by TechTarget. This publication may contain opinions of TechTarget, which are subject to change. This publication may include forecasts, projections, and other predictive statements that represent TechTarget's assumptions and expectations in light of currently available information. These forecasts are based on industry trends and involve variables and uncertainties. Consequently, TechTarget makes no warranty as to the accuracy of specific forecasts, projections or predictive statements contained herein.

Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of TechTarget, is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact Client Relations at cr@esg-global.com.

About Enterprise Strategy Group
TechTarget's Enterprise Strategy Group provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

contact@esg-global.com
www.esg-global.com