endava

Is Your Enterprise Security Profile AI-Ready?

Strategies to tackle new AI-based security threats

Table of contents

Consequences of failure to secure systems against AI-based attacks

Cybersecurity once focused primarily on protecting systems from malicious actors seeking to compromise data and disrupt organisations. Today, as organisations adopt advanced AI capabilities, new security risks are emerging across systems, data, customers and teams.

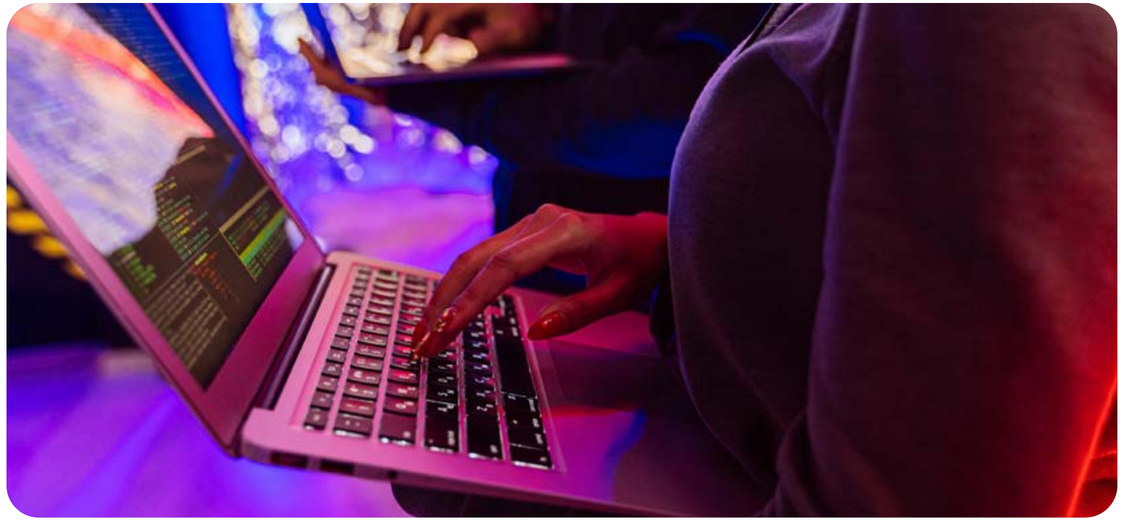
Without **appropriate safeguards**, AI-related vulnerabilities can expose sensitive data, risking reputational damage and undermining customer trust. A security strategy designed for yesterday's cyberthreats may not address the speed and scale of AI-enabled attack techniques.

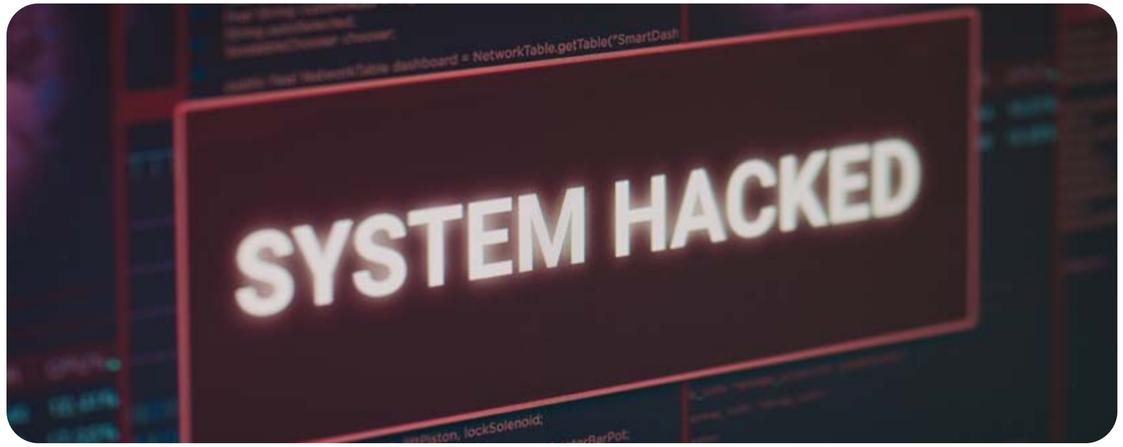
AI-powered threats introduce material operational, financial and regulatory risks. **Possible threat scenarios include:**

- **Over-privileged AI agents** accessing sensitive datasets beyond their intended scope, creating compliance and privacy exposure.
- **Prompt injection** embedded in everyday documents or emails, causing AI systems to disclose internal data while appearing to operate as expected.
- **Poisoned or biased training data** that leads AI systems to make confident but incorrect decisions at scale.
- **Shadow AI usage** where sensitive business or customer data is shared with unapproved platforms, outside governance controls, increasing the risk of data exposure.
- **Unmonitored AI workloads**, which generate abnormal API traffic or inference costs, creating financial and operational risks

However, these risks can be addressed through a structured, end-to-end systems strategy.

In this e-book, we'll share how to approach each stage of a comprehensive security strategy, supported by relevant features of **Google Cloud.** From how to design your systems from the ground up for maximum resilience to future-proofing your security profile, we'll cover each crucial step to building a modernised **security operations centre (SOC).**





New AI capabilities, new AI attack vectors

As your organisation adopts new and compelling AI solutions to enhance **productivity** and improve the **customer experience**, it's crucial to be aware of the potential risks that this technology can bring.

As AI capabilities mature, the risks will continue to evolve. Staying ahead of these threats means first assessing your relationship to the AI systems you're using and applying the same principles that protect you from human-based threats.

Key AI-driven threat vectors include:

- **Indirect prompt injection**: A malicious instruction embedded in an email or document that causes an AI system to disclose data or perform unintended actions. For example, hidden text formatted to be invisible to a human reader may still be processed by the AI model.
- **Model inversion attacks**: An attack in which outputs from a model are analysed to reconstruct elements of the data used to train it.
- **Training data poisoning**: The deliberate manipulation of training data to influence model behaviour or introduce bias and security vulnerabilities.

Preparing your landing zone

Establishing a landing zone as a foundation of your security operations strategy is critical to fortifying your systems long-term. A landing zone establishes a secure architectural foundation for cloud operations.

A robust security landing zone can include:

- **Identity and access management (IAM)**: Systems to verify users, applications, and AI agents, and ensure that they have access to appropriate data.
- **Secure networking**: Infrastructure designed to ensure data integrity, prevent incursions of malware and other threats, and prevent and stop data exfiltration at a structural level.
- **Resource management**: Strategic allocation of resources aligned to put your organisation in the best position to repel cyber threats in alignment with your business goals.
- **Firewalls:** Google Cloud provides firewall protection around cloud objects, not just at the perimeter.
- **Cost management**: Aligning security investment to business risk, ensuring resources are directed where they strengthen resilience and reduce unnecessary spend.
- **Monitoring and logging**: Continuous visibility across systems, capturing events and anomalies to support rapid detection, investigation and response.
- **Policy-as-code**: Applying security policies programmatically across the cloud environment, enabling consistent enforcement of controls such as regional deployment restrictions to meet data sovereignty requirements in the UK, EU or US.
- **Shift-left security**: Embedding security controls early in the software development lifecycle, so risks are identified and addressed during design and development rather than after release.
- **Shift-down security**: Building secure-by-default platform environments that automate controls at runtime, reducing manual effort for developers while maintaining governance and compliance, allowing them to focus on code without manual security tasks.

Every component of your security landing zone should align to the same goal, **preventing unauthorised access** to your systems and data from uncredentialled users. While the principle is straightforward, the introduction of non-human, AI agents increases complexity.

Zero trust: not just for humans

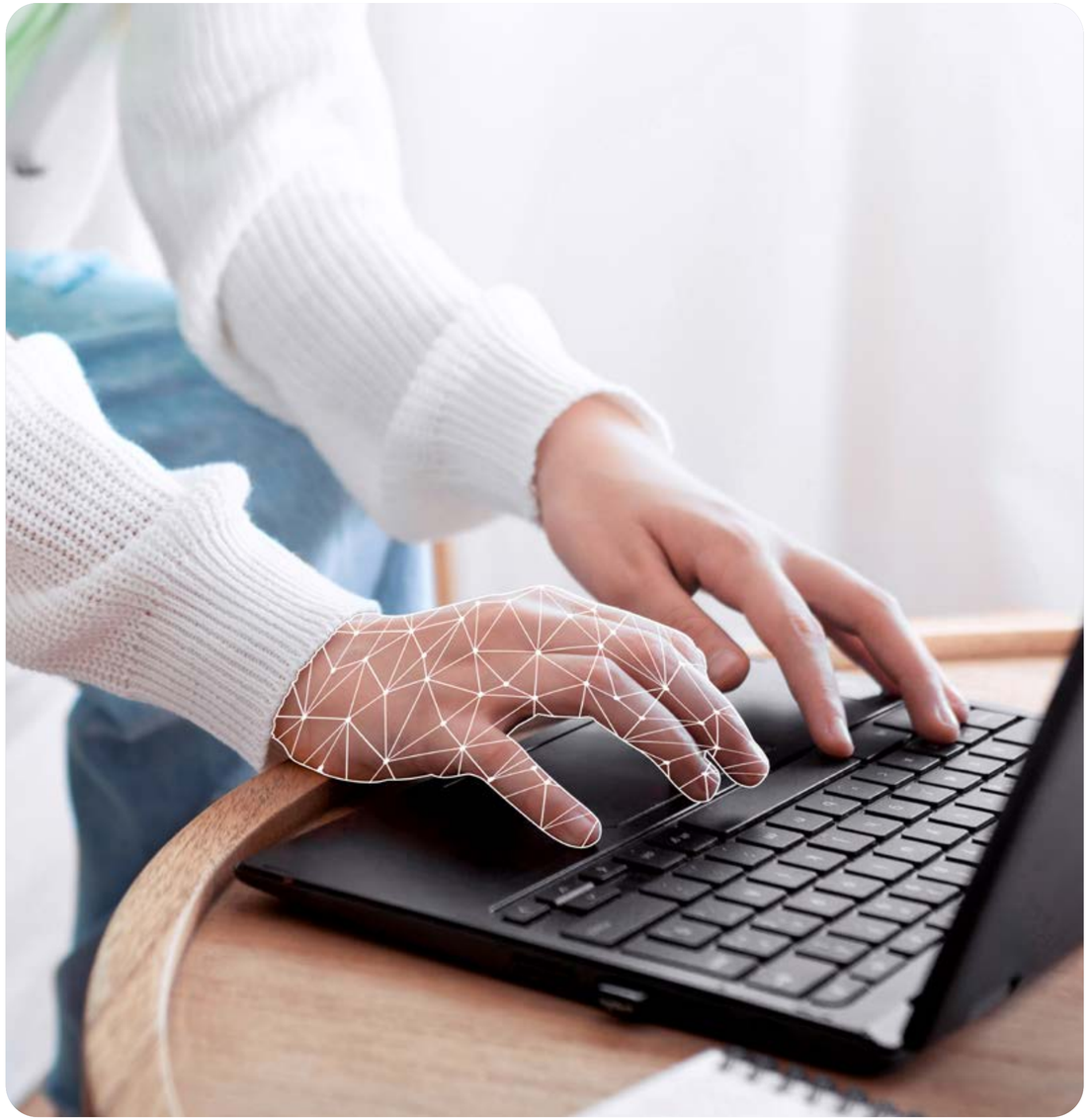
Securing your systems for resilience against AI attacks starts with a simple principle: insist that AI agents observe the same rules that you apply to your human users.

This means applying IAM principles to agentic AI, ensuring that only the right users with the right credentials have access to the data designated to them. This involves authentication that users are who they say they are, and authorisation regarding what they can do with the data.

Zero trust includes using software-defined firewalls for continuous, context-based validation in compliance with the core principle of **never trust, always verify**. A strong security profile assumes that a breach is inevitable, with response plans available that employ segmentation and isolation to limit the blast radius.

Agentic AI should follow the same rules. A **zero trust framework** begins with the assumption that any user in a system, human or otherwise, is subject to strict verification and ongoing validation.

Following the principle of **least privilege**, in which users are given only the minimum degree of access needed to complete tasks, helps reduce the risk of an AI agent exceeding the bounds of its specific instructions.



Dimensions of a security profile

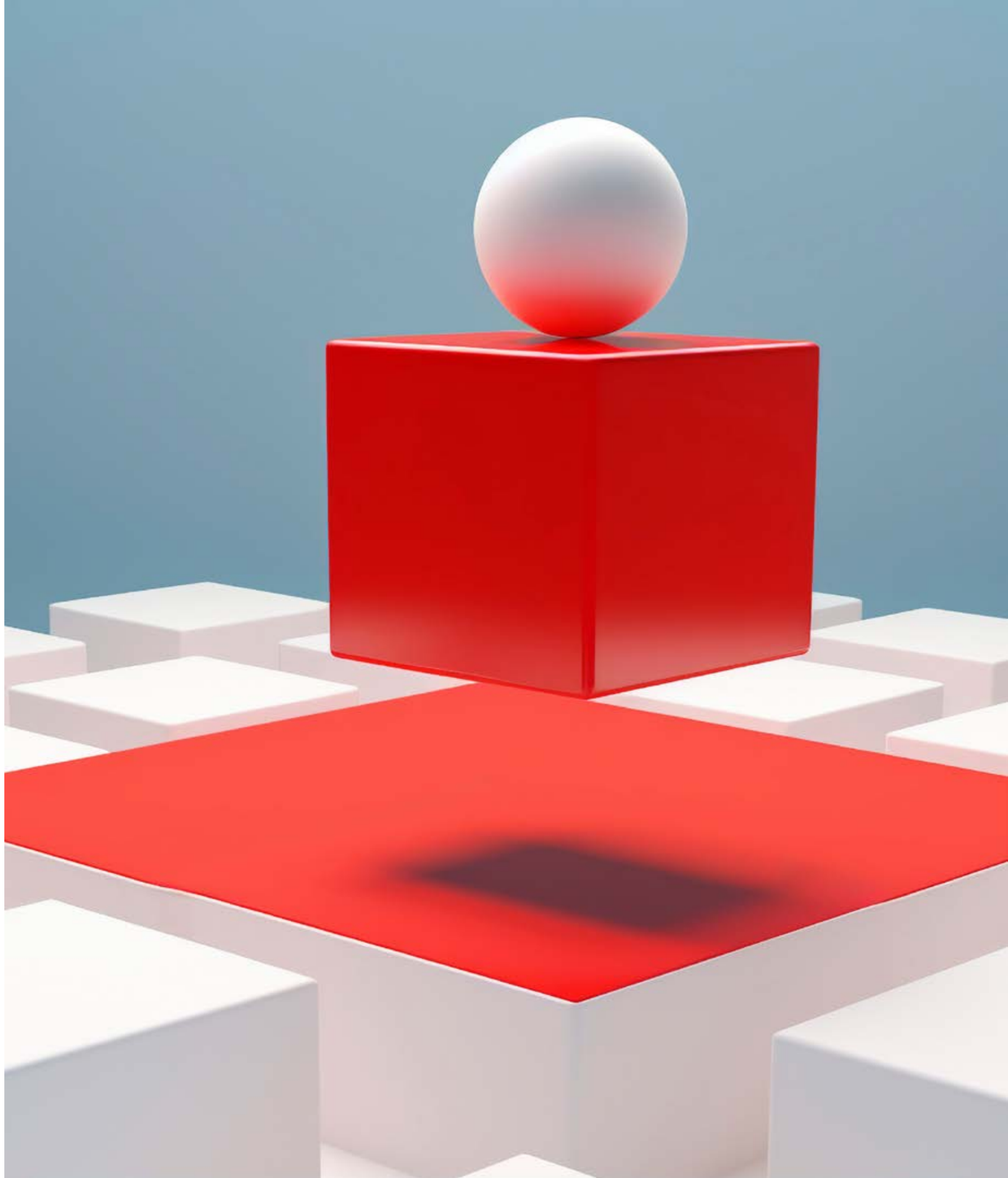
While certain **AI-driven attacks** are new, and evolving quickly, the core principles of enterprise security still apply. Design systems to withstand compromise, strengthen controls where risk is highest, monitor continuously and maintain a clear incident response plan.

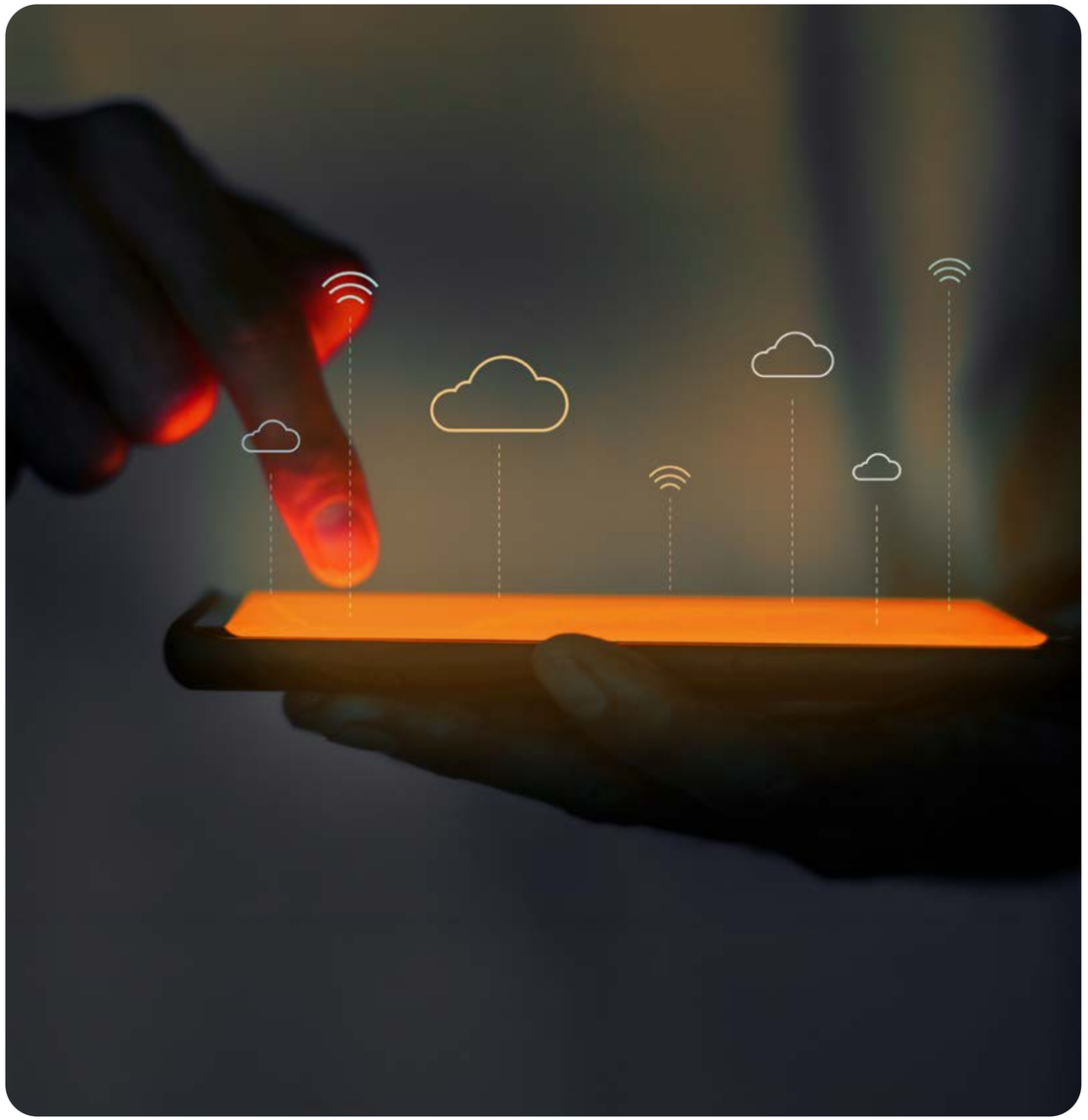
This approach requires attention to past decisions, present exposure and future risk. Organisations must be **confident** in their controls while remaining alert to emerging risks.

Begin by assessing your system's architecture, starting with an audit of your **IAM strategy**. Identify all AI agents with access to enterprise data and map to the accountable human owner. Review whether the level of access granted to each AI agent aligns with the principle of least privilege applied to human users.

Next, use these findings to address any excessive access and strengthen controls where gaps are identified. Once access is rationalised, review the tools and processes used to detect and respond to emerging threats.

A comprehensive security strategy may draw on frameworks such as the **Google Cloud Well-Architected Framework**, and the **Google Secure AI Framework**, alongside established practices in data security, encryption and governance, network security, infrastructure security and application security.





The hyperscaler advantage

Our experts have delivered cloud security programmes across **AWS**, **Azure** and **Google Cloud**. In this e-book, we focus on **Google Cloud** because it offers strong foundations for securing data and AI workloads, with controls to reduce risk, demonstrate compliance and respond faster when something does go wrong.

Google Cloud is used to protect highly sensitive information across regulated sectors such as financial services, healthcare and the public sector. For security teams, this means the platform is built with defence-in-depth in mind: identity-first access, secure-by-default services and native capabilities for monitoring, detection and response.

Google is also investing heavily in AI through services such as **Gemini** and **Vertex AI**, as well as **AI features** in **Google Workspace**. As a result, Google Cloud continues to strengthen how it secures AI systems across the model, data and infrastructure layers, helping organisations adopt AI more safely and with clearer governance.

That said, strong design is only part of a mature approach. Effective security relies on continuous visibility, prioritisation and disciplined operations.

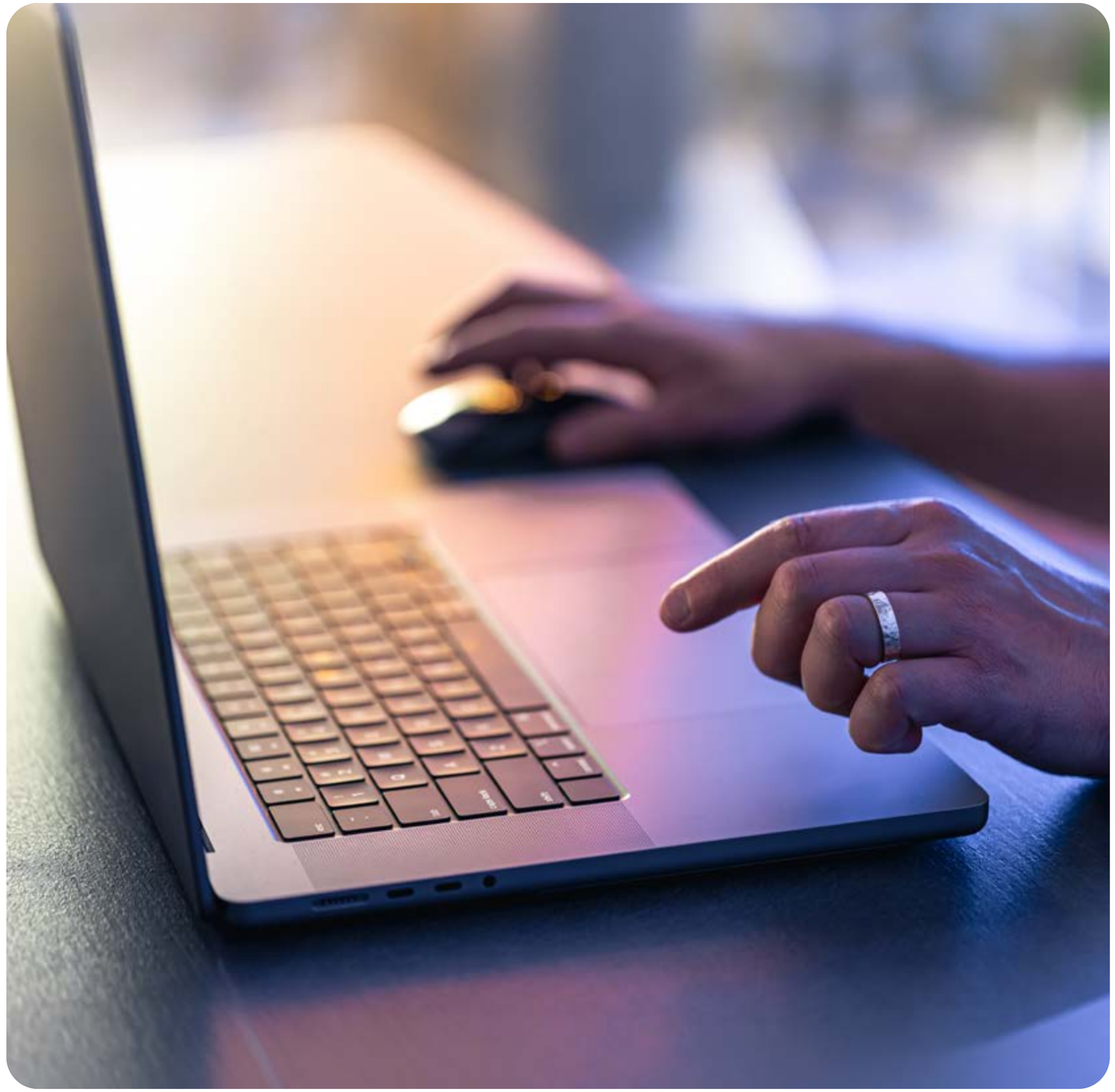
Wiz: cloud native application protection platform (CNAPP)

Many organisations use **Wiz** alongside **Google Cloud** to improve visibility and reduce exposure across cloud workloads, including AI services. **Wiz** helps teams identify what matters most by connecting misconfigurations, vulnerabilities and identities to the assets they could put at risk, then guiding remediation. It can support a proactive posture through capabilities such as:

- **Risk mapping**: Beyond listing vulnerabilities, Wiz identifies attack paths to critical assets, such as BigQuery training datasets.
- **Secure foundation checks**: Wiz helps validate that services such as Vertex AI Notebooks, GKE clusters, and storage buckets are configured in line with policy before deployment.
- **Agentless visibility**: It can also surface risk across the underlying infrastructure that hosts AI agents (for example Cloud Run and serverless functions), including container image vulnerabilities and configuration gaps.

For organisations managing rapid AI adoption, **Wiz may also help with**:

- **Shadow AI discovery**: The **AI security posture management (AI-SPM)** identifies unapproved or unmanaged AI usage to bring it back under governance and audit.
- **Identity and data integrity**: Highlights over-privileged AI agents and scans sensitive data sets for exposure risks, including data that should not be used for training.
- **Runtime protection**: Supports detection active exploitation patterns, including prompt injection-style attacks, and flags critical gaps during operation.



Google Security Operations: monitoring and response at scale

Wiz helps you understand and reduce your cloud exposure. **Google Security Operations (SecOps)** helps you detect, investigate and respond to threats using enterprise telemetry and automation. Google SecOps brings together **security information and event management (SIEM)** and **security orchestration, automation and response (SOAR)** so teams can:

- **ingest and correlate signals** across the estate
- **investigate alerts** with consistent context
- **trigger automated playbooks** for containment and remediation

One considerable advantage of Google SecOps is that it draws from resources native to the greater Google ecosystem, including **Google Threat Intelligence (GTI)** and Madiant for incident response to evolving global threats.

Building a solid security operations centre

The **integrity of your data** is crucial to your continued success. A modern SOC brings skilled teams, clear processes and the right tooling together to reduce risk over time. The goal is not to assemble a patchwork of solutions, but to run security as a coherent capability, with clear ownership, measurable outcomes and repeatable controls.

We support organisations with end-to-end security strategies, from architecture and hardening through to monitoring, incident response and continuous improvement. As a Google Cloud and Google Security Partner, we draw from a deep bench of global cloud security expertise. These experts help organisations shape a programme that is practical, compliant and aligned to their risk profile, then embed the operating model needed to sustain it.

If you're looking to **strengthen your SOC for AI-enabled threats**, we can help you assess current exposure, prioritise the highest-impact fixes and strengthen governance and operational resilience using platforms such as Google Cloud, Wiz, Google SecOps and Mandiant.

Contact us


endava