endava

Understanding
AI Cyberthreats

Assess vulnerabilities, build resilience and respond to AI-driven threats with confidence

Table of contents

Preparing for AI-enabled cyberthreats

Many organisations have already experienced **near misses or emerging risks** in their cloud security environments. In the past, threats to data and operations were often associated with external attackers operating at the edge of your environment.

Today, AI is changing the shape of these risks. The same technology that helps teams move faster can also be used to bypass controls, find weaknesses and extract sensitive information from business-critical systems.

This guide helps you frame the right conversations about AI-driven cyberthreats. It's built to spotlight common exposure points and prompt practical questions about your security posture and can act as a starting point for strengthening controls and modernising your approach.

5 common AI cyberthreats

The existing threats such as phishing have not been removed, but they are no longer the only headline risk. AI is enabling new attack paths and accelerating old ones, which means security teams need to plan for a wider range of scenarios.

AI has made attacks faster and harder to spot. These five threats show where to focus first.

1/ Over-privileged AI agents

can reach data and systems they were never intended to access. That can create privacy and compliance exposure even without malicious intent. A common cause is shared service accounts or API keys with broad, long-lived permissions. If an agent can perform high-privilege actions, a user with limited access may be able to trigger those actions simply by asking the agent.

2/ Prompt injection

can happen through normal content: an email, a document, a support ticket or a web page. An AI system can follow malicious instructions embedded in that content and disclose information while appearing to behave as expected. This risk increases when large language models (LLMs) are connected to internal knowledge sources or tools and there is no strong input handling, output filtering or policy enforcement.

3/ Poisoned or biased training data

can lead AI systems to make confident but incorrect decisions at scale.

In some cases, attackers introduce ‘triggers’ (a phrase, pattern or signal) so the model behaves normally until it sees the trigger, then produces a harmful outcome.

4/ Shadow AI

is the use of AI tools that haven’t been approved by an organisation. A common outcome is sensitive data being shared with external platforms outside governance and audit controls. Because these tools bypass procurement and security review, they may not meet requirements for encryption, retention, data handling or auditability.

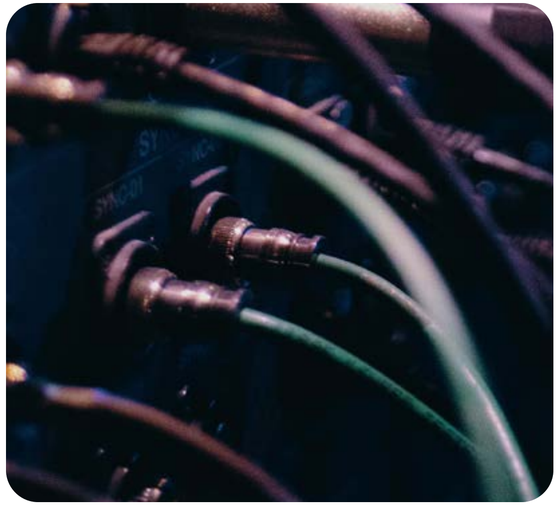
5/ Unmonitored AI workloads

AI workloads can create unusual API traffic, unexpected inference costs and operational strain before teams spot what’s happening. Without the right monitoring and controls, issues can cascade quickly. Traditional security and governance approaches may not detect these patterns early, especially where workloads scale automatically or agents run continuously.

Fortify your security operations centre (SOC)

A well-prepared security operations centre (SOC) helps orgnaisations detect AI-enabled threats early, respond decisively and learn quickly. To begin calibrating your operation model, **use the following questions:**

1. Who is on our team, what systems do they have access to, and who determines permissions?
2. What AI agents are we using?
3. What systems are accessible to our AI agents?
4. Who has oversight into:
    a. Users authorised to develop/deploy/ access AI agents
    b. Systems accessible by AI agents
5. What is our identity and access management (IAM) policy?
6. What is our policy regarding users accessing third-party or external AI platforms?
7. What is our zero trust policy? Does it operate consistently for both human and agentic AI users?
8. What training and communications do we have in place to keep teams up to date on our agentic AI security policies?
9. How do we monitor the wider internet for threats targeting our organisation?
10. If we face an AI-based related security incident, who is accountable for assessment, mitigation, communications and recovery?



Define your security operations best practices

Security operations (SecOps) bridges the gap between IT operations and security. Done well, an organisation can move from reactive response to a more proactive, joined-up defence, including for AI-enabled threats.

A strong, sustained **SecOps** focus typically includes:

- **Proactive risk reduction**: Ensuring continuous monitoring across network traffic, endpoints and cloud workloads, plus regular remediation of security gaps.
- **Faster incident response**: Using security orchestration, automation and response (SOAR) to automate repeatable tasks and speed up containment.
- **Visibility across complexity**: Having a unified view across environments to reduce blind spots as the estate grows
- **Shared operational priorities**: A culture where IT and security work towards common outcomes without creating unnecessary friction.
- **Efficient resourcing and compliance**: Leveraging automation to triage alerts, reduce fatigue and support audit-ready processes.
- **Regulatory alignment**: Ensuring consistent adherence to relevant requirements (GDPR, HIPAA, PCI DSS and others).

Get the most from security automation

Automation can help your **SOC** reduce repetitive work, improve consistency and speed up response, without losing oversight.

Google SecOps combines security information and event management (SIEM) with security orchestration, automation and response (SOAR) to help teams detect, investigate and respond across hybrid and multi-cloud environments.

Capabilities include:

- **Faster investigation and response:** Playbooks can support triage and standard actions, such as isolating hosts or blocking known-bad indicators, while keeping analysts in control.
- **Improved scale and visibility:** The platform uses Google's infrastructure for unlimited data ingestion and 12 months of hot storage, a high-performance data storage tier for frequently accessed, time-sensitive data that needs instant retrieval.
- **Threat intelligence enrichment:** Additional context can help teams prioritise alerts and focus on what matters most.

Strengthen your AI threat intelligence

Threat intelligence is most valuable when it is timely, actionable and aligned to your business risks.
A practical approach typically covers:

Incident response support

Real-time, actionable data that accelerates the detection, containment, and remediation of active cyberattacks. Leverage threat indicators and forensics to minimise downtime and reduce the overall financial and reputational damage caused by a breach.

Dark web monitoring

Monitor unindexed sites, underground forums, and encrypted marketplaces to identify stolen credentials, leaked data, or discussions of upcoming attacks.

Early warning signals

Focus on indicators that suggest preparation for attack, such as lookalike domains or suspicious infrastructure.

Vulnerability intelligence

Contextualise software and hardware flaws by identifying which vulnerabilities are being actively exploited by specific actors. This enables your organisation to prioritise your response based on real-world exploitability and business risk.

Supply chain intelligence

Evaluate and monitor the security posture of third-party vendors and suppliers to identify risks relevant to your own environment.

Partner with security experts

Endava supports organisations embedding security into cloud and AI initiatives, from architecture and controls through to monitoring and operating model. To tailor guidance, we typically begin by **understanding your organisation's critical system**s, data stores and transfer paths, as well as AI maturity and ambitions.

Our experts help organisations design and run end-to-end security programmes across cloud and AI environments. As a **Google Cloud** and **Google Security partner**, we bring practical experience across architecture, hardening, monitoring, incident response and continuous improvement.

We work with you to build a security approach that is cost-effective, compliant with your regulatory context and aligned to your risk profile. The focus is a coherent operating model and a set of controls you can sustain, not a collection of disconnected tools.

To get started, consider these questions to help determine where you are in your security journey.

1. **Policy and governance:** Does your current security policy specifically cover the use and behavior of autonomous AI agents?
2. **Visibility:** Can your SOC team clearly distinguish between normal AI activity and a sophisticated "prompt injection" attack?
3. **Access control:** Are your AI agents restricted to only the specific data and systems they absolutely need to function?
4. **Incident response:** Do you have an automated plan to immediately stop and audit an AI agent if it begins acting maliciously?
5. **Data privacy:** Do you have full visibility into what sensitive company data is being shared with external AI models?
6. **Infrastructure scale:** Can your current security tools handle the massive volume of telemetry generated by thousands of AI interactions?
7. **Regulatory compliance:** Do you have a mechanism to audit AI decision-making to meet evolving AI regulations (like the EU AI Act or industry-specific mandates)?

Ready to strengthen your security profile for AI?
Talk to us today about an AI security assessment covering identity, data access, cloud controls and security operations.

↘ **Contact us**


endava