The AI security control plane: Building Trust in the Age of AI



Table of contents

01	The emerging AI security challenge	5
02	Why securing the future AI enterprise looks so different from the past	7
03	You can't secure what you can't see and control	10
04	The AI security control plane: shaping security architecture for the AI era	14
05	Operationalising an AI security control plane	17
06	The AI security control plane	20
	Authors	26

AI is transforming how organisations create value, and demand keeps accelerating. From AI agents that enhance individual productivity to autonomous systems that transform entire business processes, the opportunity to unlock greater efficiency, faster innovation and new business models has never been greater. The organisations that capture the most value will be those that move with confidence, trusting the outcomes AI delivers and giving boards, customers and employees the assurance to deploy AI further and go bigger.

That confidence is the real differentiator. As AI capability advances at remarkable speed, the organisations that pull ahead are those that build trust into how they adopt it, so they can innovate quickly and safely. For the teams on the front line responsible for deploying AI, the goal is clear: turn that trust into a repeatable, scalable advantage rather than navigating each new use case without a map.



> This question goes beyond simply needing to secure AI. Organisations that get this right will support the business to innovate faster within the guardrails that they have put in place.

Ivo van Bennekom,
Partner, Cybersecurity and Privacy, PwC Netherlands



01

The emerging AI security challenge





Knowing where AI applications and models sit across the enterprise is increasingly challenging because AI is spreading across systems, applications and third-party solutions. And AI agents, sophisticated systems capable of performing tasks autonomously, analysing data and aiding decision-making with minimal human intervention, making that even harder. As autonomy increases, so does its potential, making robust security, identity management and governance essential to scaling AI safely and with confidence.

While cybersecurity in a large, complex enterprise has never been straightforward, the rise of AI makes it harder at a scale many organisations are not set up to manage. AI also presents a different set of vulnerabilities and risks from those familiar to IT and security teams. As AI adoption expands the enterprise attack surface, organisations face new vulnerabilities across data, identities, models and AI-enabled solutions. **PwC's 29th Annual Global CEO Survey** found that cyber risk was among CEOs' leading concerns globally. Organisations pursuing AI at scale should therefore treat security as a foundational enabler of their AI strategy.

Established frameworks for IT governance, security, and risk management were not designed for highly distributed, AI technologies. As a result, organisations face an unprecedented challenge over visibility and control. The instinct is to treat this as a trade-off between speed and safety. It isn't. Embedding security, governance and accountability by design is what turns AI risk management from a compliance constraint into a scaling advantage, providing the confidence to move quickly on AI and capture more value from it.

02

Why securing the future AI enterprise looks so different from the past





Historically, organisations sought to maintain visibility and control over technology environments through established governance, security and operational processes. Challenges emerged where technology services were introduced outside those processes, whether through unauthorised software, unmanaged infrastructure, or third-party solutions. The rise of AI opens a new frontier of value, and with it a new generation of this challenge to address, as agents, AI services and large language models spread across the enterprise. Organisations that bring these into a managed, well-governed environment will be the ones positioned to scale AI with confidence.

Realising that opportunity, however, means understanding what makes AI different, and AI agents in particular. Their adoption can be decentralised and fragmented across multiple applications, platforms and systems, including:



Embedded in the stack

Core business platforms, typically from third party vendors supporting functions from CRM to HR, now come with built-in AI capabilities. Often enabled with a simple click, these features can create new data flows and decision paths that IT cannot easily see or control.

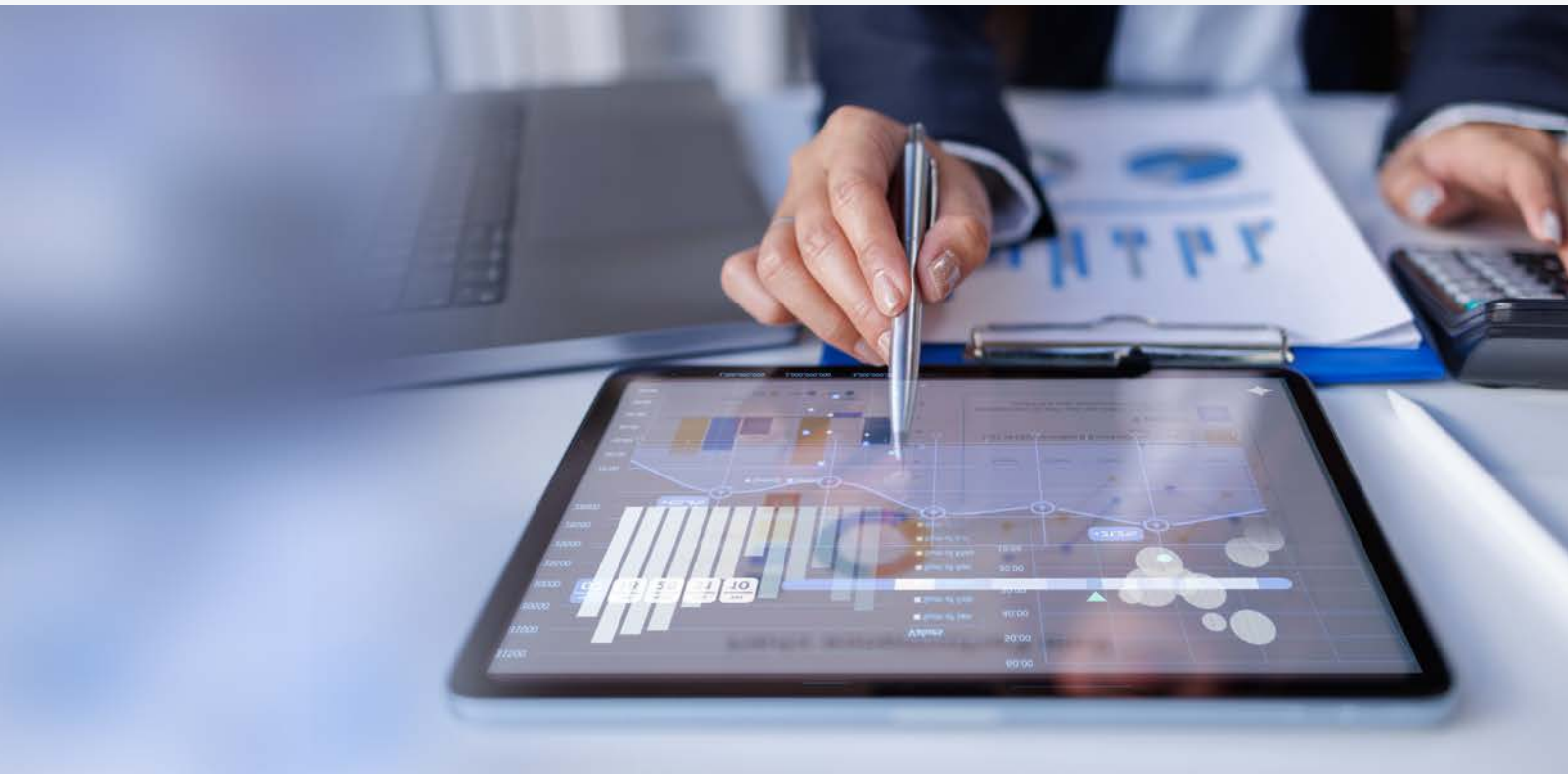


Democratised via assistants

With AI Agents and assistants, every employee now has uniquely powerful tools at their fingertips. Integrated into daily workflows, these assistants access, process, and generate content based on sensitive corporate data in ways that are nearly impossible for security teams to track or govern.

Procured on the fringes

While shadow IT has always been an issue, ‘shadow AI’ has far more worrying implications. Eager for a competitive edge, business units may connect directly to third-party AI models and APIs, upload proprietary data to public tools, or build unsanctioned micro-applications. That creates risk outside formal oversight.



The visibility challenge is not limited to any one type of organisation. Digital-native enterprises often operate across rapidly growing ecosystems of cloud, SaaS and AI services, while more established organisations may need to manage a combination of legacy, hybrid and cloud-based environments. In both cases, the distribution of data, identities, applications and AI systems creates significant challenges in maintaining consistent visibility, governance and control.

03

You can’t secure what you can’t see and control

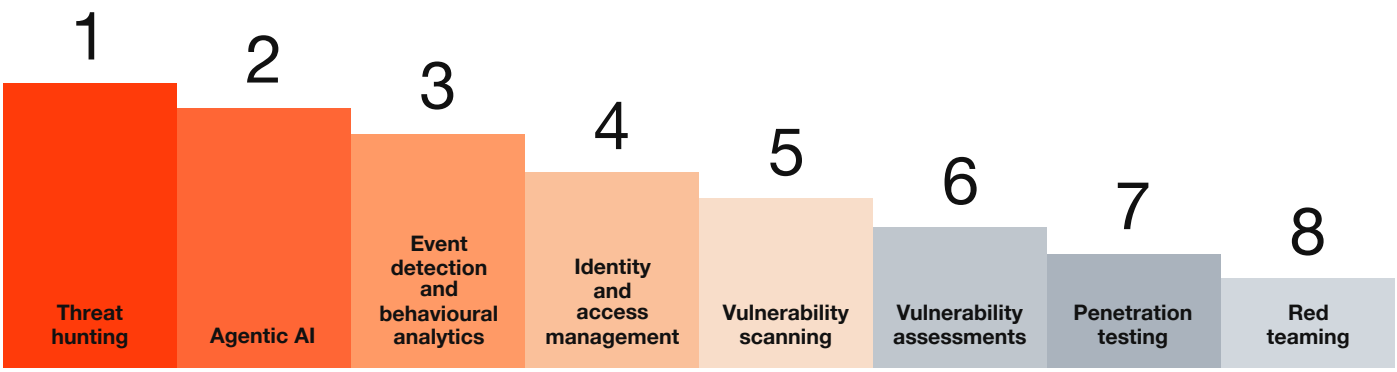


This critical lack of visibility, along with fragmented governance, leads to the risk of an even more threatening outcome: a systemic loss of overall control. The leap from AI providing assistance, to AI agents working with minimal human intervention, introduces new risks that current control mechanisms were not designed to address.

These include AI agents taking unpredictable actions. They also include agents interacting with other agents, compounding errors, and amplifying flawed decisions in ways that can lead to cascading failure and make root causes hard to trace. Agents could inadvertently leak sensitive information to third-party systems they interact with, including supply chain partners. And threat actors may take advantage of AI agents' autonomy to manipulate them into taking malicious actions on their behalf through prompt engineering and other emerging attack vectors. **PwC's 2026 Global Digital Trust Insights: C-suite playbook and findings** reports that agentic AI ranks among the leading AI security capabilities organisations are prioritising over the next 12 months, with planned uses including cloud security, data protection, cyber defence and operations.

Agentic AI among the top prioritised AI security capabilities

(Ordered based on those who ranked as their top priority)



Q18. Which of the following AI security capabilities will your organisation prioritise over the next 12 months?
Base: Security leaders=1740

Source: PwC 2026 Global Digital Trust Insights

When organisations use a mix of commercial, private and third-party LLMs they expand the attack surface by creating additional pathways for sensitive data movement. In the absence of strong AI security controls, including prompt governance, data segmentation, access management and continuous monitoring, organisations increase the likelihood of data leakage and compliance risk.

Fotis Sofronis,
Director, EMEA Microsoft Cybersecurity Leader, PwC Greece



The lack of visibility into AI and the consequential loss of control is one of the defining security challenges of the AI era. The critical question for every leader is no longer if AI is operating within their organisation, but where it is, what it is doing, and who or what is truly in control.

The answer is to build an approach to control AI securely that embeds new forms of AI-specific security, governance and risk as upfront design considerations. This offers comprehensive oversight and control everywhere it's needed. We refer to this as an **"AI security control plane"**.



PwC's 2026 Global AI Performance Study found that AI leaders achieve 7.2 times more AI-driven revenue and efficiency gains than their peers.

The study links stronger performance to a broad set of AI foundations, including strategy, investment, data and technology, workforce, governance and risk, and innovation. Security and governance therefore support confident innovation and scale.

04

The AI security control plane: shaping security architecture for the AI era



An “AI security control plane” provides the architectural foundation for securing AI across the enterprise. It spans AI development, deployment, operations and lifecycle management, coordinating capabilities to address AI-specific security risks. Rather than replacing traditional IT security and governance, it connects and extends existing capabilities, tools and structures to support a Zero Trust architecture for secure AI adoption. It enables these expectations to be enforced and demonstrated at scale. Importantly, the “AI security control plane” is a foundation of Trusted AI, not the whole framework. Trust also depends on governance, testing and continuous observability.

AI Security Control Plane

A unified set of security controls that wraps the AI platform to enable trusted, secure outcomes.



Illustration of an ‘AI Security Control Plane concept’ for representation only.



Rather than relying on a single tool, a control plane addresses the multiple touch points of AI workflows and agent journeys across the organisation. It's also important to recognise that organisations at different stages of AI maturity and operating in very different industry and regulatory contexts may be variously buying, building and/or consuming AI tools and solutions. They need an approach that can support the secure use of all these options.

So above all, an "AI security control plane" needs to be able to adapt in line with changing requirements and the rapid pace of tech development. It's a dynamic and modular architecture, not a static blueprint, that addresses the different types of AI and their maturity, from tools offering personal productivity, to the custom and autonomous agents that organisations want to deploy, provided by the following layers:

- **AI Security Posture**
- **AI Threat Management**
- **AI Platform Security**
- **AI Data & Model Security**
- **Agent Identity & Access**
- **AI Runtime Security**

05

Operationalising an AI security control plane





As organisations embed AI into business processes, security capabilities, governance structures and operating models must evolve to keep pace with these emerging risks. Responsibility for managing those risks no longer sits with the CISO alone. Making the “AI security control plane” work effectively requires more than implementation; it requires an operating model that can be sustained.

As a key part of this, it’s important to define clear responsibilities that control plane capabilities require from different roles, from the CISO to the wider C-suite.

Different security silos dealing with their specific responsibilities won’t fix the fundamental underlying enterprise-wide challenges. Teams exploring AI from a risk perspective, security operations looking at the new signals coming in, different vulnerabilities falling to the identity and offensive security teams, and so on, won’t get the job done.

What’s needed is a shared operating model, common telemetry, and clear decision rights across functions.

Securing AI at enterprise scale cannot sit with one person in the C-suite. It requires the CISO, CIO, CDO, DPO, CTO, model risk, legal, technology and business leaders to work from a common shared responsibility model. The challenge is not just whether AI is being used, but where it is being used, what data it can access, what actions it can take, and who is accountable when something goes wrong.

Christopher Dean,
Director, AI Security, PwC UK



06

The AI security control plane



The next question is how organisations can operationalise that control plane in a way that is practical, scalable and aligned to their existing technology environment. For organisations already invested in Microsoft, Microsoft 365 E7 provides one route to bringing identity, security, data protection and AI governance closer together. For those not yet on the Microsoft stack, it offers an equally credible starting point, allowing the control plane to be built in from the outset rather than retrofitted later. For leadership teams, the “AI security control plane” is the foundation for scaling AI with confidence. It brings together the capabilities organisations need to maintain visibility, control and accountability and shows how Microsoft enables them through a platform approach anchored in Microsoft 365 E7.



AI Security Control Plane Layer	Business Risk if Unmanaged	Microsoft Security Capabilities
AI Security Posture	Organisations may lack visibility into where AI is being used, what data it accesses and which agents are operating across the environment, increasing the risk of unmanaged AI adoption, compliance breaches and governance failures.	Delivers continuous visibility, discovery and security posture management across AI environments through Microsoft Agent 365, Microsoft Defender and Microsoft Purview.
AI Threat Management	New attack techniques targeting AI systems may go undetected, increasing the likelihood of security incidents, business disruption and financial impact.	Detects, investigates and responds to AI-related threats through Microsoft Defender, Microsoft Sentinel and threat intelligence capabilities.
AI Platform Security	Vulnerabilities, misconfigurations or weak controls within AI platforms can expose critical services, reduce operational resilience and increase cybersecurity risk.	Secures AI platforms, applications and supporting infrastructure across Microsoft Azure, Microsoft Azure AI Foundry and Microsoft Power Platform through consistent security controls, policy enforcement and resilience capabilities.
AI Data & Model Security	Sensitive data, intellectual property and AI models may be exposed, altered or used inappropriately, leading to financial loss, regulatory scrutiny and reputational damage.	Protects sensitive data and AI assets through Microsoft Purview, delivering Data Security Posture Management (DSPM) for AI, information protection, data governance and compliance controls to reduce the risk of oversharing, data leakage and unauthorised access.
Agent Identity & Access	Over-privileged agents and automated workflows may gain access to information or systems beyond their intended scope, increasing the risk of unauthorised actions, data exposure and control failures.	Establishes and governs trusted identities for AI agents, applications and users through Microsoft Entra and Agent 365, enabling secure authentication, least-privilege access, policy enforcement and lifecycle management across AI-driven workloads.
AI Runtime Security	AI applications and autonomous agents may behave in unexpected ways, execute actions outside approved policies, or create operational, legal and reputational risks if not continuously monitored and controlled.	Monitors and protects AI applications and agents at runtime through Microsoft Defender and Azure AI Foundry guardrails, helping ensure behaviour remains aligned with organisational policies, safety requirements and approved operating boundaries.

> Trust is what lets organisations move faster with AI. When you can see every agent, govern what it can access, and switch it off in seconds if something goes wrong, AI stops being a risk you manage and becomes an advantage you can scale.

Narayan Kumar Gupta,
Senior Manager, AI Security Architect, Global Microsoft Security Association Lead, PwC Ireland

There is no shortage of products claiming to support an "AI security control plane". But for leadership teams, the real question is not which single tool to buy. It is whether the organisation can bring identity, security, data governance and AI oversight together in a way that is consistent, scalable and accountable.

Agentic AI shifts the security conversation from monitoring systems to governing autonomous decision-makers. For leaders, that means the "AI security control plane" cannot remain a theoretical framework. PwC helps translate it into an operational capability: one that contains the blast radius of agentic systems, enforces accountability at the agent level, and gives leaders the visibility and control they need before autonomy is scaled.

Working alongside organisations, we help establish the capabilities, processes, and ownership structures required to embed AI security into day to day operations, calibrated to their AI maturity, industry, and regulatory obligations. This includes treating agents as semi-autonomous identities subject to the same governance, access controls, and monitoring as users and APIs, not as exceptions to existing security frameworks.



“

AI has enormous potential, but organisations can only unlock that value when security, governance, and trust are built in from the start. By bringing together industry expertise and Microsoft's AI and security capabilities, partners like PwC help customers innovate with confidence and accelerate their transformation responsibly.

Leticia Gammill

EMEA Security Solutions & Programs Leader, Microsoft



PwC's global alliance with Microsoft strengthens this operational foundation, combining advanced security technology with PwC's governance risk, and industry expertise, and our track record in leading large-scale technology transformation. Together, we help leaders enforce controls, satisfy regulatory requirements, and maintain a defensible security posture, while unlocking the broader business transformation that agentic AI makes possible.

The result is not innovation at the expense of control. It is innovation because of it.

Set out a pathway for secure and responsible agentic AI deployments with PwC and Microsoft.

Authors



Ivo van Bennekom

Partner,
Cybersecurity and Privacy

PwC Netherlands



Christopher Dean

Director,
AI Security

PwC UK



Fotis Sofronis

Director,
EMEA Microsoft
Cybersecurity Leader

PwC Greece



Narayan Kumar Gupta

Senior Manager,
AI Security Architect, Global Microsoft
Security Association Lead

PwC Ireland

Contributors



Vladyslav Dunajevski

Director,
Cyber Security,
Cloud Security, and Privacy

PwC Germany



Tom Corkett

Senior Manager,
AI Security Lead

PwC UK



Leticia Gammill

EMEA Security Solutions
& Programs Leader

Microsoft



© 2026 PwC. All rights reserved. PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see **www.pwc.com/structure** for further details