AI and agents Insights | Microsoft Security Blog

Published: 2026-04-02 · Archived: 2026-04-05 21:31:39 UTC

AI and machine learning help you identify threats sooner and respond more effectively. Learn how to safeguard your infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS) resources across multicloud and hybrid environments.

Filtered by

Clear All

- AI and agents

Refine results

- **Threat actor abuse of AI accelerates from tool to cyberattack surface**

Generative AI is upgrading cyberattacks, from 450% higher phishing click-through rates to industrialized MFA bypass.

- **Applying security fundamentals to AI: Practical advice for CISOs**

  Read actionable advice for CISOs on securing AI, managing risk, and applying core security principles in today's AI-powered environment.

-

- **Identity security is the new pressure point for modern cyberattacks**

  Read the latest Microsoft Secure Access report for insights into why a unified identity and access strategy offers strong modern protection.

- **Governing AI agent behavior: Aligning user, developer, role, and organizational intent**

  This research report explores the layers of agent intent and how to align them for secure enterprise AI adoption.

- **CTI-REALM: A new benchmark for end-to-end detection rule generation with AI agents**

  Excerpt: CTI-REALM is Microsoft's open-source benchmark for evaluating AI agents on real-world detection engineering—turning cyber threat intelligence (CTI) into validated detections.

- **Secure agentic AI end-to-end**

  In this agentic era, security must be woven into, and around, every layer of the AI estate.

- **New tools and guidance: Announcing Zero Trust for AI**

  Microsoft introduces Zero Trust for AI, adding a new AI pillar to its workshop, enhanced reference architecture, updated guidance, and a new assessment tool.

-

- **New Microsoft Purview innovations for Fabric to safely accelerate your AI transformation**

  As organizations adopt AI, security and governance remain core primitives for safe AI transformation and acceleration.

- **Detecting and analyzing prompt abuse in AI tools**

  Hidden instructions in content can subtly bias AI, and our scenario shows how prompt injection works, highlighting the need for oversight and a structured response playbook.

- **Secure agentic AI for your Frontier Transformation**

We are announcing the next step to make Frontier Transformation real for customers across every industry with Wave 3 of Microsoft 365 Copilot, Microsoft Agent 365, and Microsoft 365 E7: The Frontier Suite.
