



Espandere Microsoft Security Copilot con agenti AI

Mario Serra

MW & Security Specialist @ Overnet
Microsoft MVP Security

Michele Sensalari

CTO @Overnet
Microsoft MVP Security



“

Security Copilot

”



Microsoft Security Copilot

Microsoft Security Copilot is an advanced cybersecurity tool that leverages artificial intelligence to enhance the detection, analysis, and response to cyber threats. It integrates various security data sources, including system logs, network events, and vulnerability information, to offer a comprehensive view of an organization's IT infrastructure security status.



Use cases include:

- Incident Summaries
- Guided Response
- Threat Intelligence
- Script Analysis
- Incident Reporting



Microsoft Security Copilot

Primary use cases



Incident Response — Summarize incidents, assess impact, and receive tailored remediation guidance, including for triage, investigation, and containment.



Security Reports — Summarize investigations, incidents, vulnerabilities, or threats in minutes and prepare the information in ready-to-share reports.



Reverse Script Engineering — Analyze complex command line scripts and translate them into natural language with clear explanations of actions.

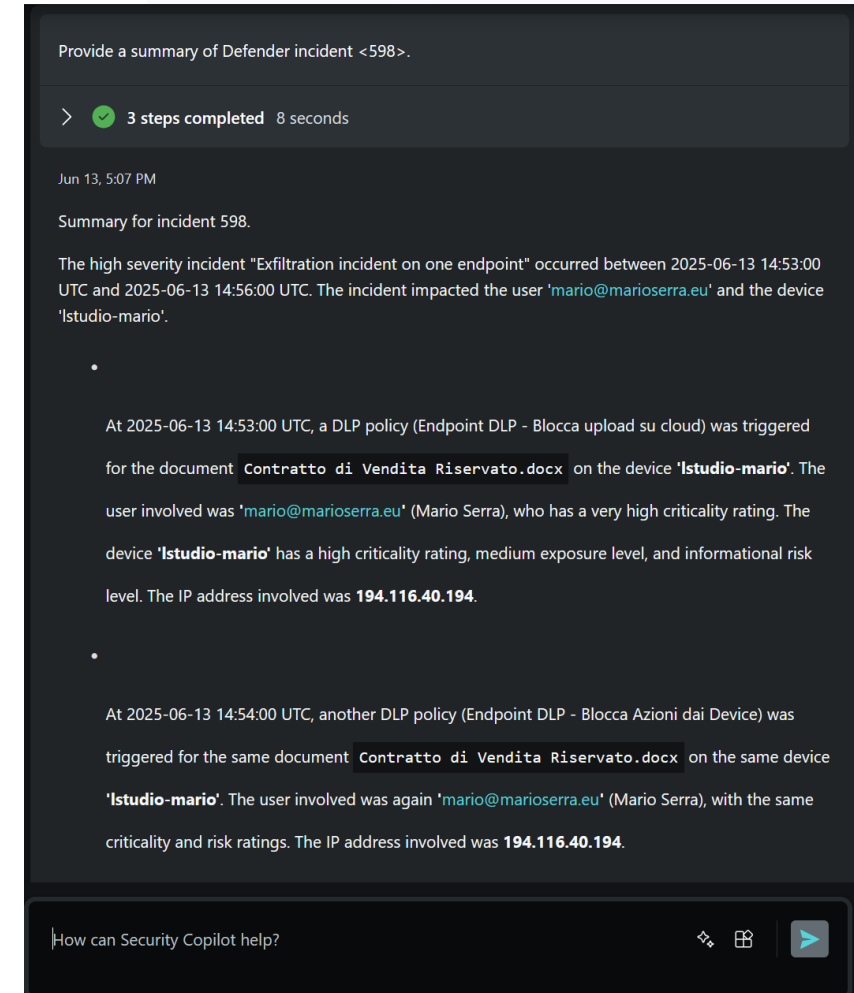
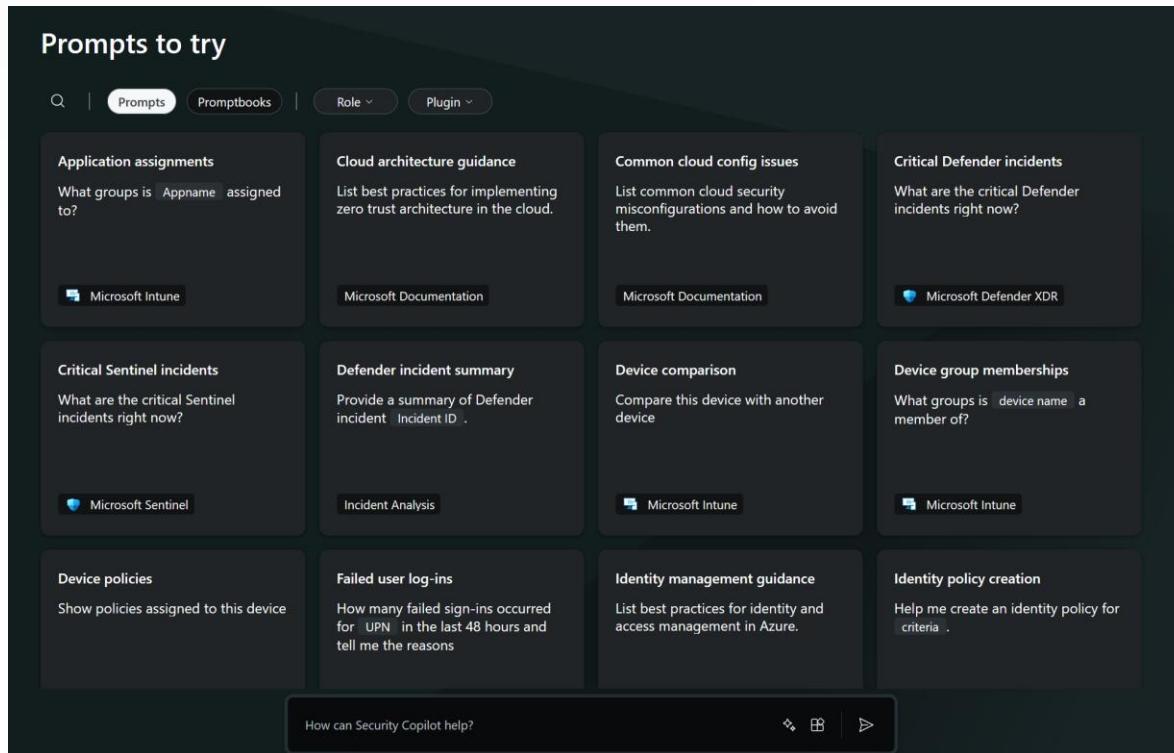


Security Posture Management — Learn if your organization is at risk from vulnerabilities and examine resources in your environment for signs of a breach.



Standalone Copilot Security Experiences

The Standalone experience is best when you need a centralized, immersive portal for security operations.



Embedded Copilot Security Experiences

The Embedded experience may be used when you want to enhance your workflow within existing Microsoft security products with Copilot's AI capabilities.

The screenshot displays the Microsoft Security Copilot interface for an incident titled "ID 598: Exfiltration incident on one endpoint". The interface is divided into several sections:

- Incident Details:** Shows the incident title, severity (High), status (Active), and last update time (Jun 13, 2025 5:03 PM). A critical asset tag is also present.
- Attack story:** A tab showing the incident's progression.
- Detection & Categories:** A section with a table of active alerts and categories.
- Alerts:** A list of alerts, including "DLP policy (Endpoint DLP - Blocca upload su cloud) matched for document (Contratto di Vendita Riservato.docx) in a device".
- Incident graph:** A visual representation of the incident, showing a communication between a user (mario) and a device (Istudio-mario).
- What Happened:** A section describing the incident, stating that Mario Serra copied a file to cloud "Contratto di Vendita Riservato.doc" on an endpoint device.
- Policy description:** A section explaining the policy, stating that it is a custom criterion created from scratch to protect content.
- Tasks:** A section showing the status of tasks, with a "Status: Any" filter and an "Add task" button.
- Summary:** A section providing a detailed summary of the incident, including the date and time of the incident, the user involved, and the device.

The interface also includes a "Copilot" button in the top right corner, indicating the AI capabilities available within the security products.



“

Demo

”



Microsoft Security Copilot Users Experience

Microsoft Purview

Search

Copilot

Home

Solutions

Agents

Learn

Settings

Data Loss Prevention

< > ⏮ ⏭

Welcome to the Microsoft Purview portal

Microsoft Purview brings together solutions across data governance, data security, and compliance so that you can govern and secure your data wherever it lives.

Supported cloud platforms:

Microsoft 365 Microsoft Azure Microsoft Fabric

Other cloud platforms

Data Catalog

Information Protection

Data Loss Prevention

Insider Risk Management

View all solutions →

Featured insights

Copilot

+

...

×

Get AI-generated summaries from Microsoft Purview product documentation.

Start with our suggested prompts or type your own to ask Copilot to summarize a topic or find related information from product documentation. Include solution names for relevant responses.

Troubleshoot

How do I fix issues with automatic labeling in MIP?

Learn

How do I setup a DLP policy?

Ask

What permissions are needed to get started with Insider Risk Management?

View prompts

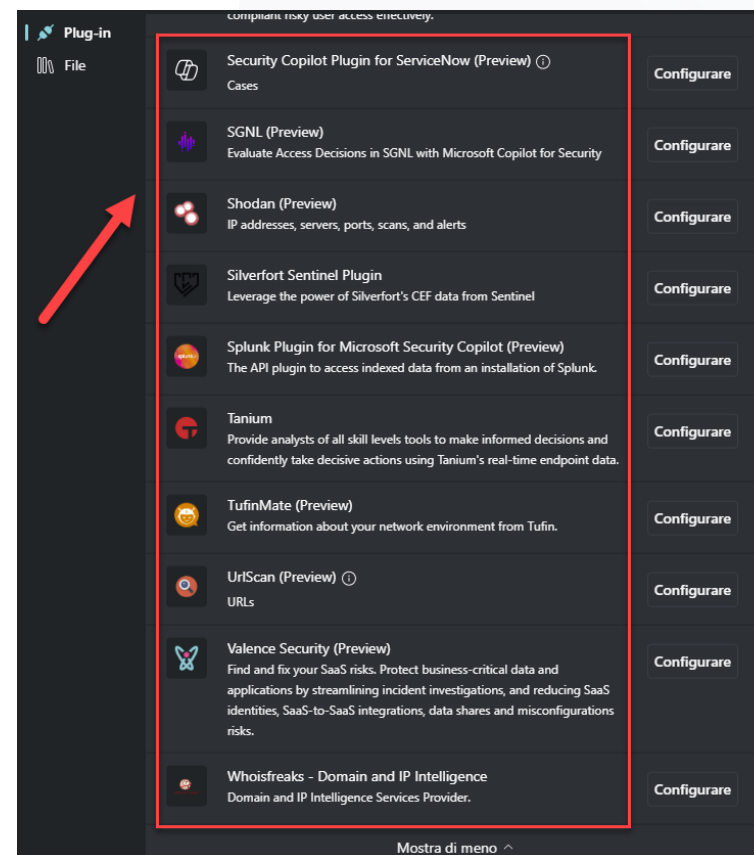
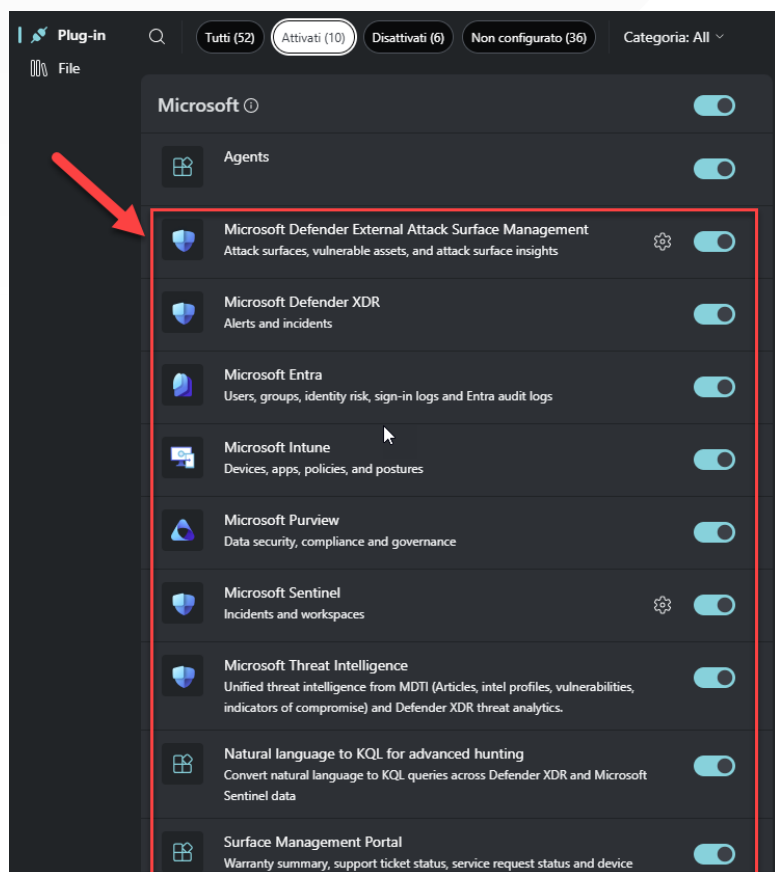
Ask a specific question about a topic in Microsoft Purview documentation

0 / 2000

➤

Plugin

Copilot for Security comes with many default plugins and supports several non-Microsoft plugins. You can also extend Copilot for Security's capabilities by adding or creating your own plugin. The Copilot for Security platform enables developers and users to write plugins that can be invoked to perform specialized tasks.



Data flow for Microsoft Copilot for Security

Prompting in Microsoft Security solutions

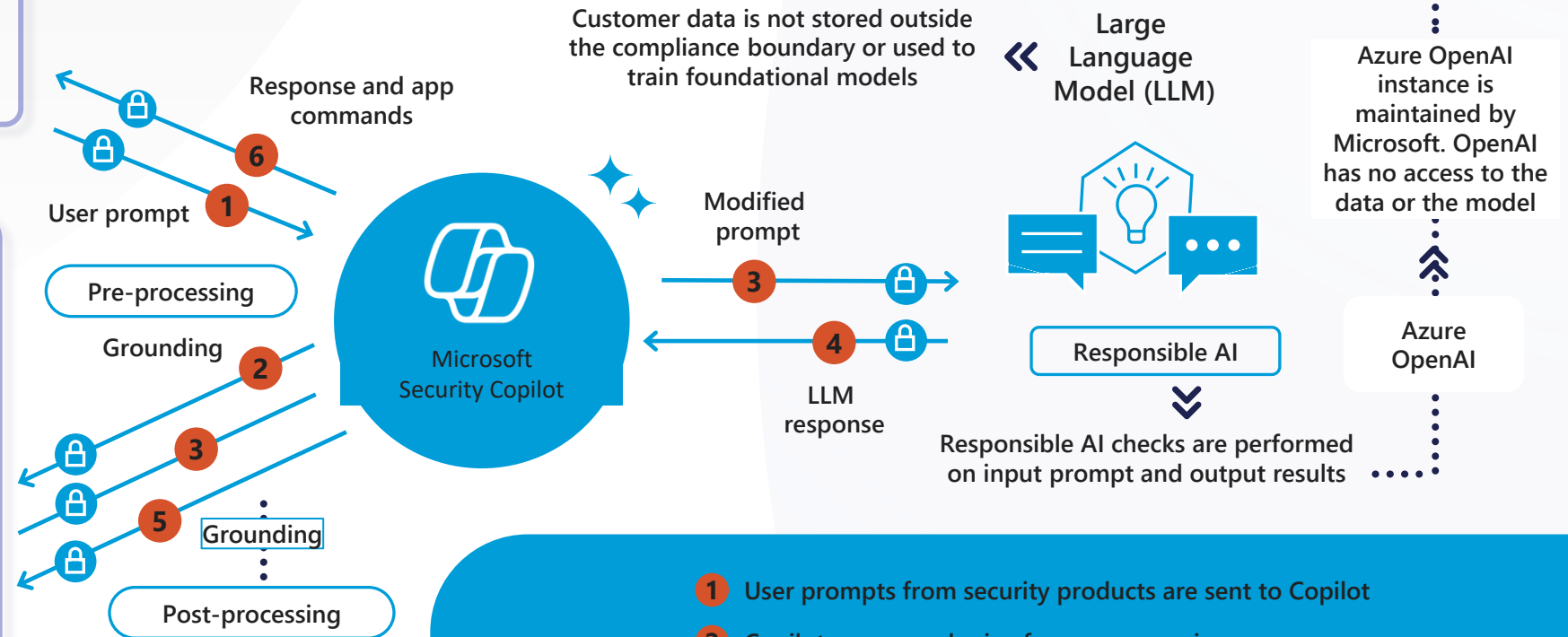


Plugins for Microsoft and third-party security products



splunk> servicenow ...

Your context and content
Event logs, alerts, incidents, & policies



Data flow

(🔒 = all requests are encrypted via HTTPS)

- 1 User prompts from security products are sent to Copilot
- 2 Copilot accesses plugins for pre-processing
- 3 Copilot sends modified prompt to LLM
- 4 Copilot receives LLM response
- 5 Copilot accesses plugins for post-processing
- 6 Copilot sends the response, and app command back to security products

“

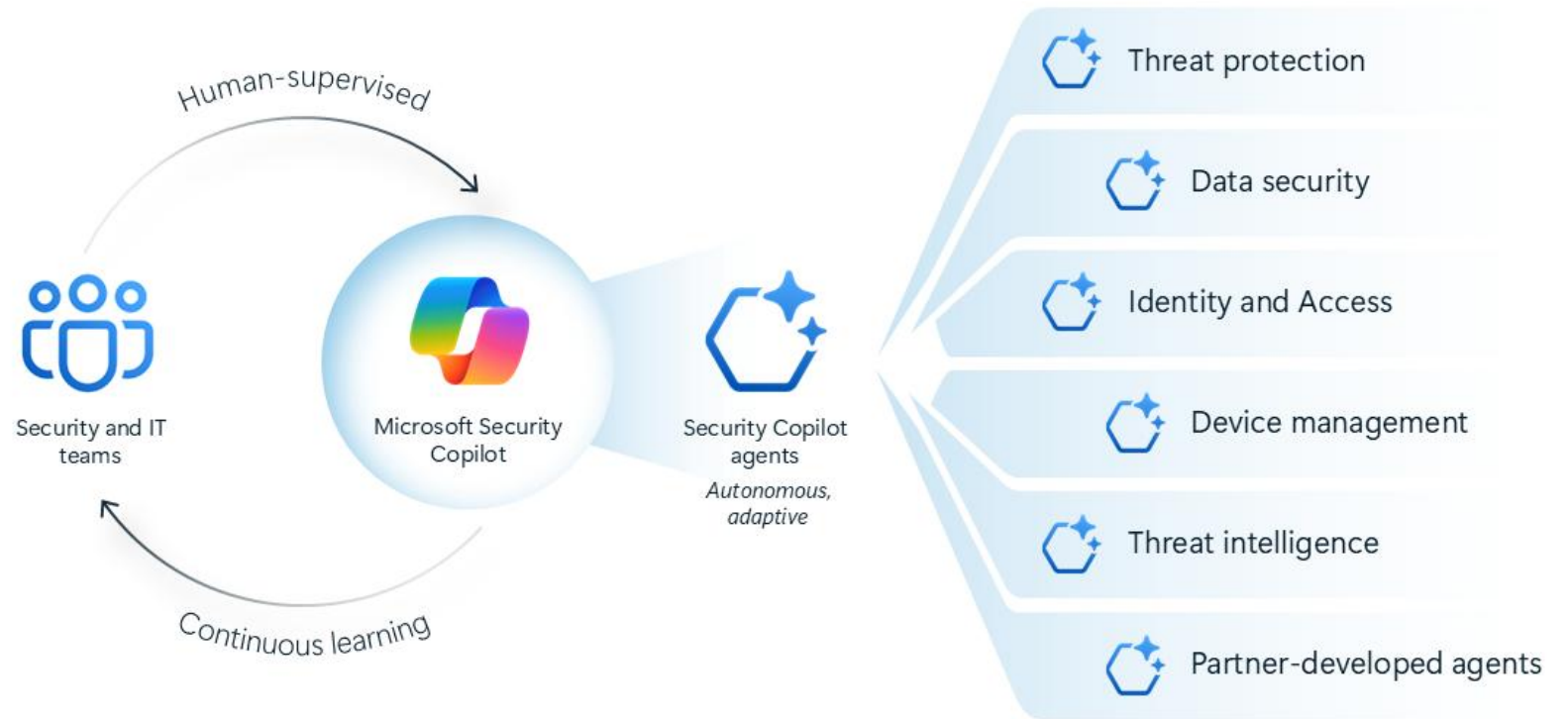
Security Copilot AI Agent

”

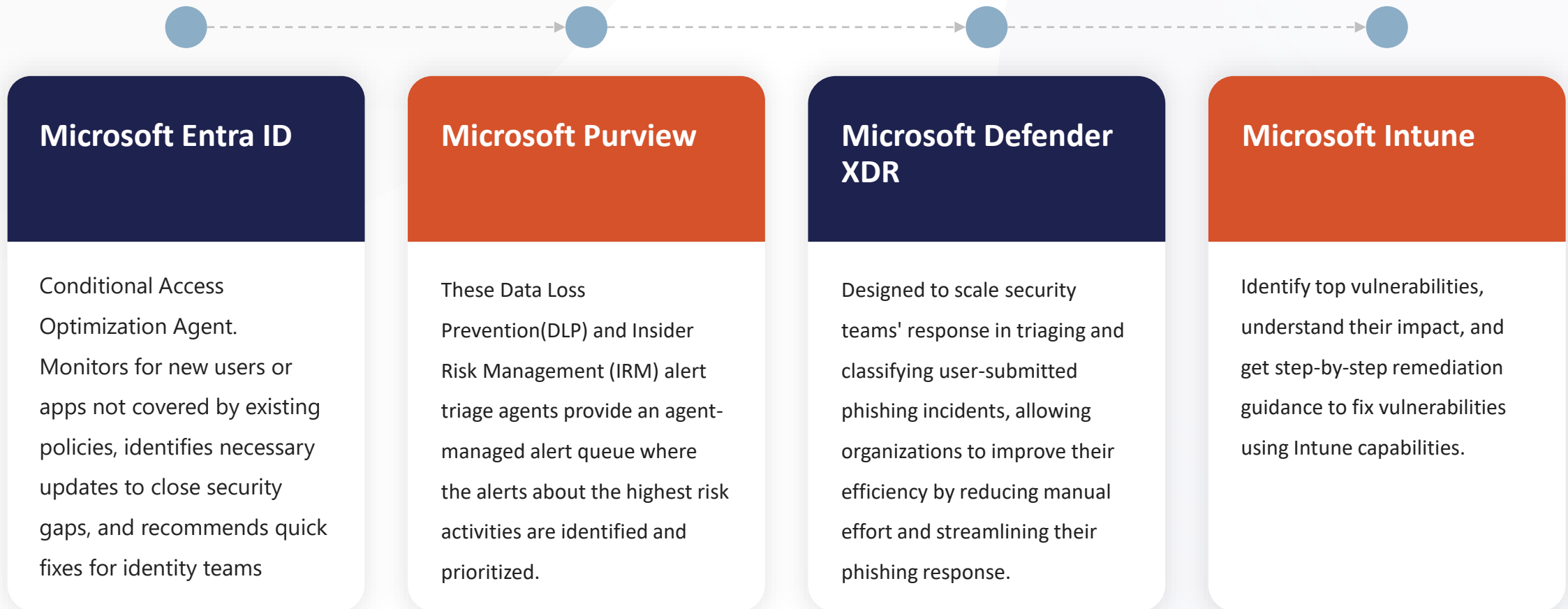


Security Copilot Agents

- Microsoft Security Copilot agents automate repetitive tasks and reduce manual workloads
- These agents handle high-volume, time-consuming tasks by pairing data and code with an AI language model.
- Agents utilize SCUs to operate just like other features in the product.



Available Agent in Security Copilot

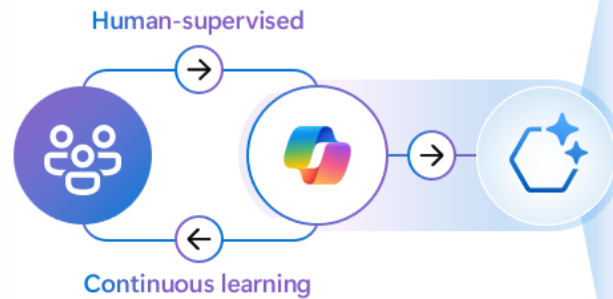


Standalone experience - **Threat Intelligence Briefing** Agent in Security Copilot: generates threat intelligence briefings based on the latest threat actor activity



Partner ecosystem

Agents to empower roles across security and IT



Microsoft Security

- Phishing Triage
- Alert Triage in DLP
- Alert Triage in IRM
- Conditional Access Optimization
- Vulnerability Remediation
- Threat Intelligence Briefing

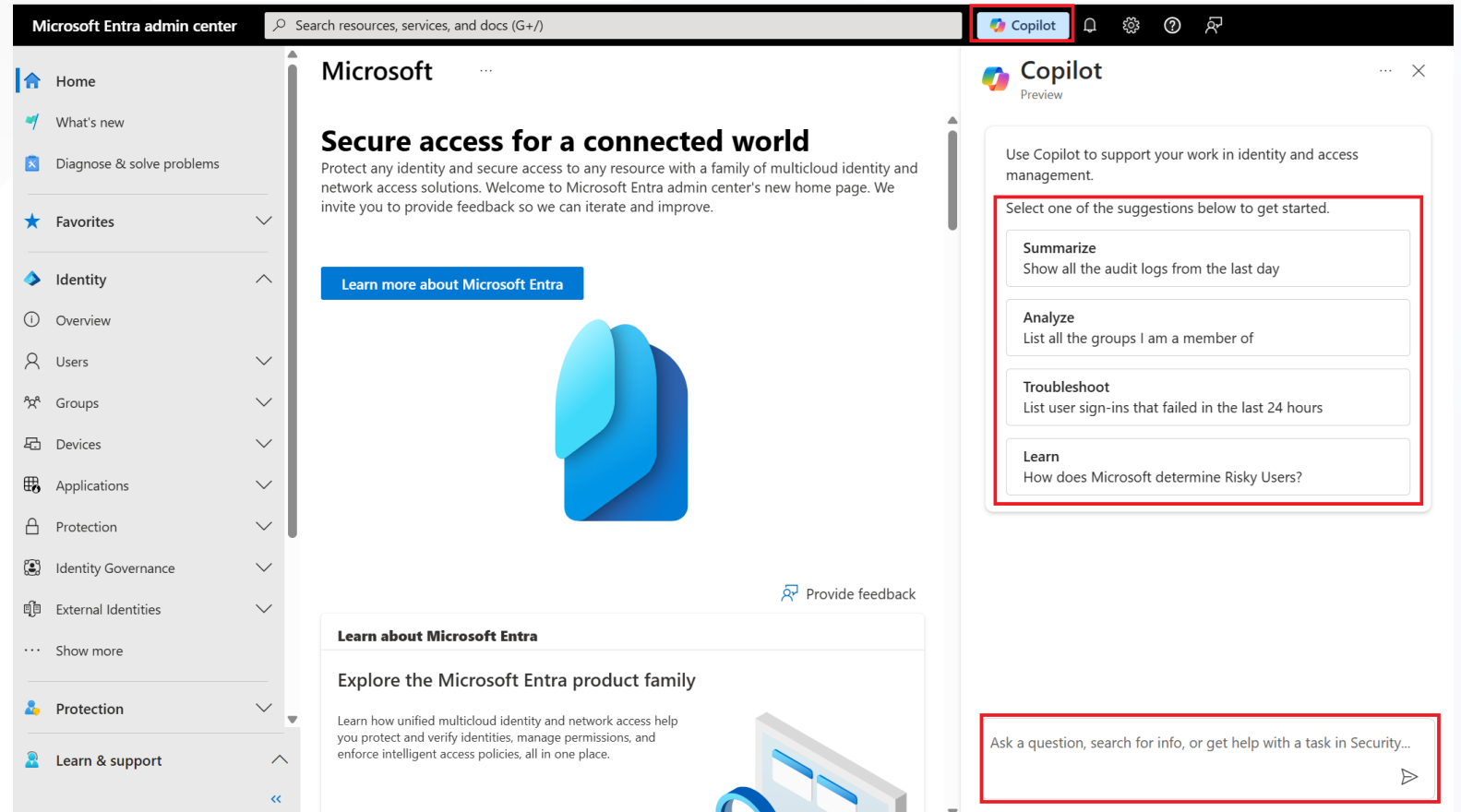
Partner ecosystem

- Privacy Breach Response by OneTrust
- SecOps Tooling by BlueVoyant
- Network Supervisor by Aviatrix
- Alert Triage by Tanium
- Task Optimizer by Fletch
- Email Threat Analyst by Performanta
- IAM Supervisor by Performanta



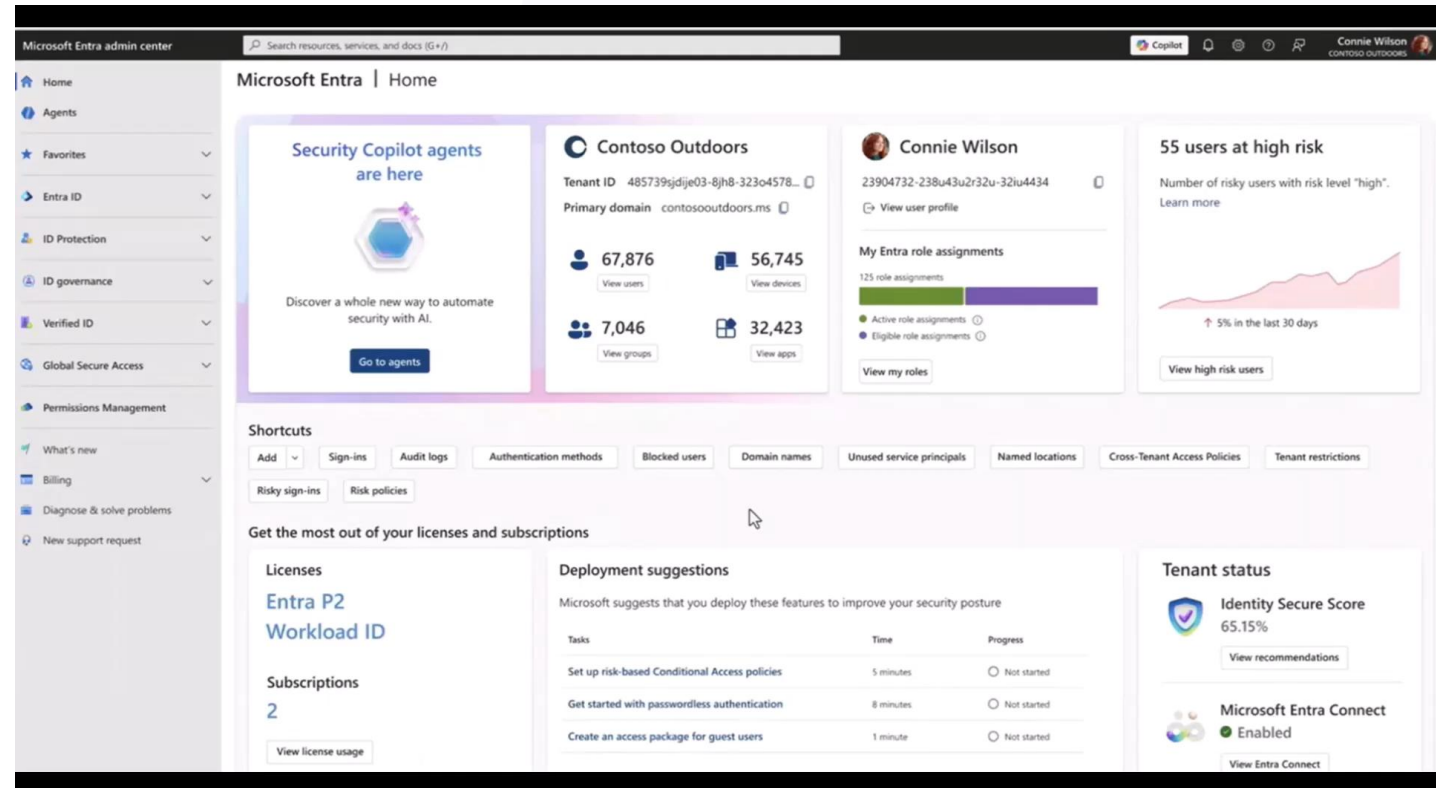
Microsoft Security Copilot scenarios in Microsoft Entra

- Investigate recommendations
- Summarize a user's risk level
- Investigate access reviews
- Investigate insights within entitlements management
- Investigate alerts in Scenario Health Monitoring
- Microsoft Entra Role Based Access Control (RBAC)
- License Usage
- etc



Microsoft Entra Conditional Access optimization agent

- The Conditional Access optimization agent ensures all users are protected by policy. It recommends policies and changes based on best practices aligned with Zero Trust and Microsoft's learnings. In preview, the agent evaluates policies requiring multifactor authentication (MFA), enforces device-based controls (device compliance, app protection policies, and Domain Joined Devices), and blocks legacy authentication and device code flow.
- The agent runs every 24 hours but can also run manually.
- It runs in the context of the administrator who configured the agent.



Phishing remains one of the most persistent cybersecurity threats

- 91% of cyberattacks start with a phishing email, according to Microsoft Report
- Threat actors constantly refine their tactics: reviving old techniques or inventing new ones to bypass detection
- Attackers use creative phishing techniques to move faster, smarter and more persistently.
- New vectors:
 - QR code phishing
 - Collaborative tools as Teams
- In Microsoft 365 organizations with mailboxes in Exchange Online, users can report phishing and suspicious email in Outlook.
- SOC analysts are overwhelmed



Microsoft Security Copilot Phishing Triage Agent in Microsoft Defender

This AI-powered virtual agent is designed to scale security teams' response in **triaging and classifying user-submitted phishing incidents**, allowing organizations to improve their efficiency by reducing manual effort and streamlining their phishing response.

The screenshot displays the Microsoft Defender console interface. The main section, titled 'Incidents', shows a summary of attack disruptions (21) and incidents resolved by the agent (95%). Below this, a table lists various incidents, including those reported by users as malware or phishing. The right-hand panel provides detailed information for a selected incident, including its classification as 'False positive' and the latest agent activity.

Incidents Summary:

- Attack disruptions: 21 (Last 30 days)
- Incidents resolved by agent: 95% (Last 30 days)

Incident Table:

Incident name	Incident ID	Tags	Severity	Assigned to
Email reported by user as malware or phish	23575642	Agent, Credential Phish	Low	Phishing Triage Agent
Email reported by user as malware or phish	2356358	Agent, Credential Phish	Low	Phishing Triage Agent
Account enumeration reconnaissance on one endpoint	2861358		High	Logan Edwards
Exfiltration incident involving multiple users	7515358		Low	Sam Centrell
Multi-stage incident involving Initial access & Lateral movement including ransomware on mu...	2681673	Ransomware +2	High	Ellis Turner
Email reported by user as malware or phish	2681323	Agent, Credential Phish	Low	Phishing Triage Agent
Email reported by user as malware or phish	8906478	Agent, Credential Phish	Low	Phishing Triage Agent
Email reported by user as malware or phish	2681353	Agent, Credential Phish	Low	Phishing Triage Agent
Multi-stage incident involving Initial access & Lateral movement including ransomware on mu...	2681673	Ransomware +1	High	Logan Edwards
Email reported by user as malware or phish	2681348	Agent, Credential Phish	Low	Phishing Triage Agent
Email reported by user as malware or phish	2681395	Agent, Credential Phish	Low	Phishing Triage Agent
Multi-stage incident involving Initial access & Lateral movement including ransomware on mu...	2681673	Ransomware +1	High	Logan Edwards
Email reported by user as malware or phish	2681395	Agent, Credential Phish	Low	Phishing Triage Agent

Incident Details:

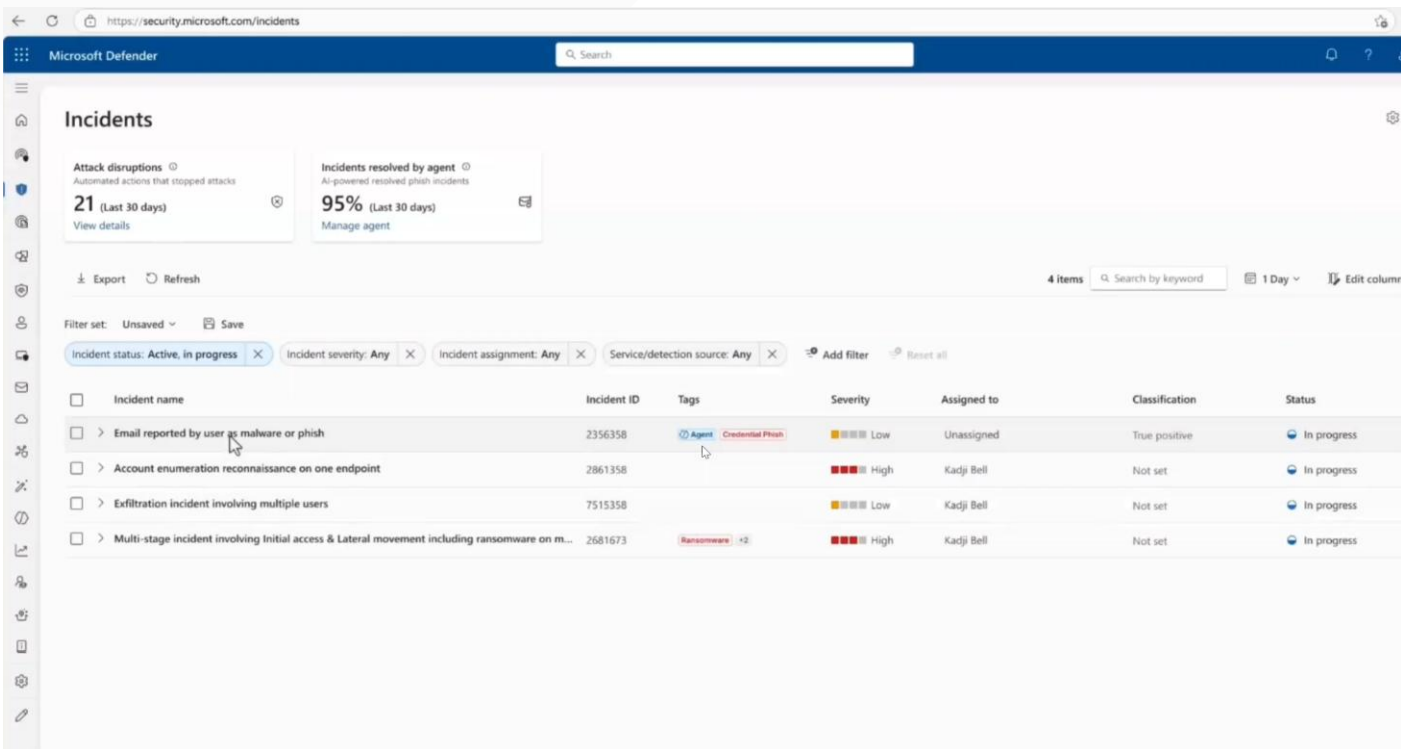
- Incident ID:** 23575642
- Assigned to:** Phishing Triage Agent
- Classification:** False positive
- First activity:** 3/18/25 at 12:21 AM
- Last activity:** 3/18/25 at 12:35 AM
- Impacted assets:** Users (1), Mailboxes (1)

Latest agent activity:

- Completed:** Classify as 'False positive' and resolve alert. The user submitted an unexpected email reminding them to register for an upcoming professional conference. The message was unsolicited and prompted the recipient to take immediate action to reserve their spot, resembling common phis...



Phishing Triage Agent in the Defender XDR Portal



Incident name	Incident ID	Tags	Severity	Assigned to	Classification	Status
> Email reported by user as malware or phishing	2356358	Agent, Credential Phish	Low	Unassigned	True positive	In progress
> Account enumeration reconnaissance on one endpoint	2861358		High	Kadji Bell	Not set	In progress
> Exfiltration incident involving multiple users	7515358		Low	Kadji Bell	Not set	In progress
> Multi-stage incident involving initial access & lateral movement including ransomware on m...	2681673	Ransomware, +2	High	Kadji Bell	Not set	In progress

Autonomously triages phishing alerts

Leveraging generative AI, the agent perform advanced analyses to identify even the most sophisticated phishing attempts with high accuracy

Learn from feedback

The agent learns and evolves its behavior through feedback provided by analysts in natural language, ensuring it becomes increasingly accurate and tailored to the organization's specific context

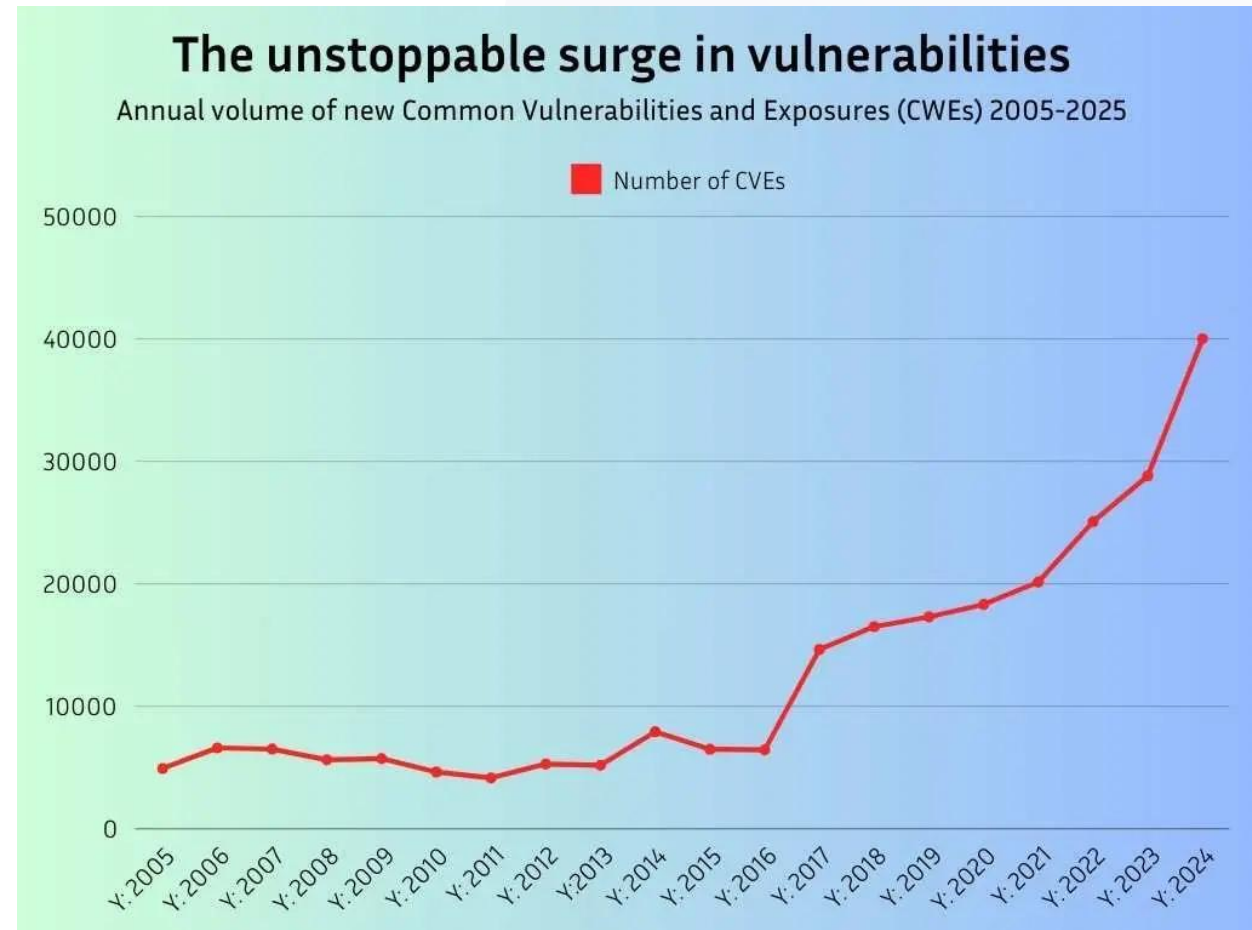
Transparency for verdicts

The agent offers full transparency in its decision-making, providing explanations for its verdicts in clear, natural language, as well as a flow diagram that visually outlines the steps taken to reach a conclusion

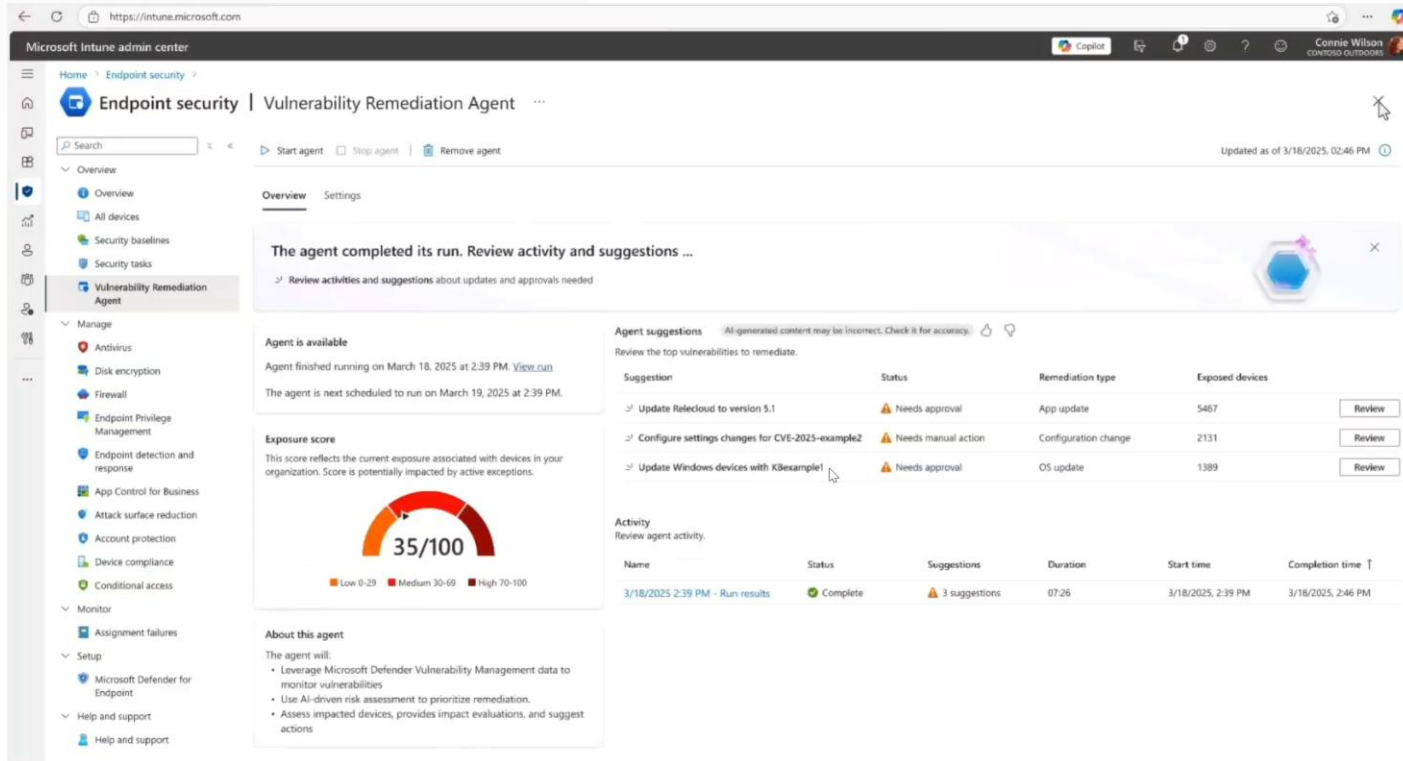


Threat actors are exploiting vulnerabilities faster than ever

- 40.000 common vulnerability and exposures were published in 2024
- Threat actors constantly exploiting vulnerabilities reviving old techniques or inventing new ones to bypass detection
- Time to exploit vulnerabilities is falling significantly



Intune Agent – Intelligent, end to end Vulnerability Remediation



Discovery of vulnerability data

The agent analyzes vulnerability data from Microsoft Defender Vulnerability Management (MDVM) to help IT pros discover the vulnerabilities in their environment

AI generated analysis to streamline remediation

The agent prioritizes the remediation of high impact vulnerabilities by leveraging MDVM and Copilot for an impact analysis, providing explanations for its recommendations in clear natural language & offering full transparency in its decision making

Step by Step remediation guide

The agent provides a step-by-step remediation guide, equipping admins with both the critical insights needed to assess high-impact vulnerabilities and the actionable steps to take in Intune to resolve them





Alert triage in Microsoft Purview

The data loss prevention agent helps security teams by evaluating alerts based on the **sensitivity risk**, **exfiltration risk**, and **policy risk**. The agent then sorts the triaged alerts into four categories. These categories are presented in the DLP solution on the Alerts page.

- **All:** Which contains all alerts the agent is working on or has fully triaged.
- **Needs attention:** These are the alerts that the agent has triaged and determined that they should be assigned top priority in your review process.
- **Less urgent:** These are the alerts that the agent has triaged and determined that they should be assigned a lower priority in your review process.
- **Not triaged:** These are the alerts that the agent has yet to triage or the agent unsuccessfully attempted to triage.



“

Demo

”



Deploy Agent

Microsoft Security Copilot

Prompts to try

Prompts

Promptbooks

Role

Plugin

Application assignments
What groups is `Appname` assigned to?

Microsoft Intune

Cloud architecture guidance
List best practices for implementing zero trust architecture in the cloud.

Microsoft Documentation

Common cloud config issues
List common cloud security misconfigurations and how to avoid them.

Microsoft Documentation

Critical Defender incidents
What are the critical Defender incidents right now?

Microsoft Defender XDR

Critical Sentinel incidents
What are the critical Sentinel incidents right now?

Microsoft Sentinel

Defender incident summary
Provide a summary of Defender incident `Incident ID`.

Incident Analysis

Device comparison
Compare this device with another device

Microsoft Intune

Device group memberships
What groups is `device name` a member of?

Microsoft Intune

Device policies
Show policies assigned to this device

Failed user log-ins
How many failed sign-ins occurred for `UPN` in the last 48 hours and

Identity management guidance
List best practices for identity and access management in Azure.

Identity policy creation
Help me create an identity policy for criteria .

How can Security Copilot help?

The logo for Microsoft Security Copilot, featuring a stylized orange 'C' with a blue eye-like shape inside.

Triage DLP Alert by Agent

Microsoft Purview

Search

Copilot

Home

Solutions

Agents

Learn

Settings

Data Loss Prevention

Overview

Policies

Alerts

Classifiers

Explorers

Diagnostics

Related solutions

Information Protection

Insider Risk Management

Alerts

Alert Triage Agent (Preview) Standard

Insider risk severity and user activity context are not available at this moment because Data sharing is turned OFF in Insider risk management settings. Please get in touch with your Insider risk admin to modify this.

If your role group permissions are restricted to a specific set of users, you'll only be able to view alerts for those users. [Learn more about role group permissions.](#) View role groups

Did you know you can now manage your DLP alerts in the Microsoft Defender portal? Alerts are automatically combined into incidents, which provide a comprehensive view into potential policy violations and advanced tools for investigation and remediation. [Learn more about incidents](#) [Go to Incidents page](#)

Export Refresh Set status

1 of 15 selected Customize columns

Filter Reset Filters

Time range: 5/17/2025-6/17/2025 User: Any Alert status: Any Alert severity: Any

<input type="checkbox"/>	Alert name	Severity ⓘ	Status	Time detected
<input checked="" type="checkbox"/>	DLP policy match for email with subject 'Nuumeri di Carta da disabilitare'	Low	Active	Jun 17, 2025 11:12 AM
<input type="checkbox"/>	DLP policy match for email with subject 'Nuumeri di Carta da disabilitare'	Low	Active	Jun 17, 2025 11:09 AM
<input type="checkbox"/>	DLP policy match for email with subject 'Nuumeri di Carta da disabilitare'	Low	Active	Jun 17, 2025 11:09 AM
<input type="checkbox"/>	DLP policy match for document 'C_Credito.docx' in OneDrive	Low	Active	Jun 17, 2025 3:28 AM
<input type="checkbox"/>	DLP policy match for document 'File CC.docx' in SharePoint	Low	Active	Jun 16, 2025 12:40 PM
<input type="checkbox"/>	DLP policy match for document 'Final_Tech_2020.pdf' in SharePoint	Low	Active	Jun 15, 2025 11:43 PM
<input type="checkbox"/>	DLP policy match for document 'TryCatch - Progetto_Office_365_Offerta (16)....	Low	Active	Jun 15, 2025 11:43 PM
<input type="checkbox"/>	DLP policy match for document 'cc.docx' in OneDrive	Low	Active	Jun 15, 2025 10:52 PM

Copilot

- 4/20271656562491 : n pede. Suspendisse dapibus lorem pellentesque magna. Integer nulla. Donec blandit feugiat ligula. 4720271656562491 5298314193078412 344135437869521 card verification card identification number cvn cid cvc2
- 5298314193078412 : e dapibus lorem pellentesque magna. Integer nulla. Donec blandit feugiat ligula. 4720271656562491 5298314193078412 344135437869521 card verification card identification number cvn cid cvc2 cvv2
- 4720271656562491 : n pede. Suspendisse dapibus lorem pellentesque magna. Integer nulla. Donec blandit feugiat ligula. 4720271656562491 5298314193078412 344135437869521 card verification card identification number cvn cid cvc2
- 5298314193078412 : e dapibus lorem pellentesque magna. Integer nulla. Donec blandit feugiat ligula. 4720271656562491 5298314193078412 344135437869521 card verification card identification number cvn cid cvc2

AI-generated content may be incorrect. Check it for accuracy.

What activity was performed on the data in this alert?

Describe the sensitive info, file labels, or data triggering this alert.

Ask specific question about a topic in Purview documentation

0 / 2000

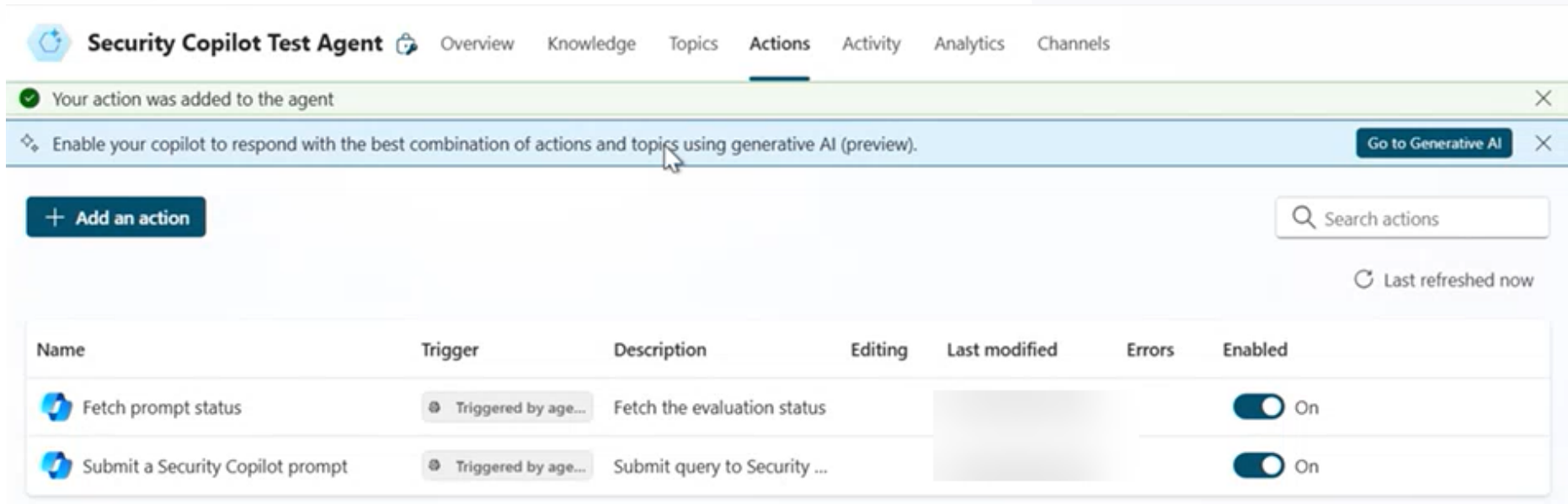


Microsoft Security Copilot - Copilot Studio connector







The Copilot Studio connector enables you to access Security Copilot

The connector exposes the following connector actions:

- **Submit a Security Copilot prompt** - Submit a natural language prompt to create a new Security Copilot investigation. After completion, the evaluation result will then be returned to your workflow.
- **Fetch a Security Copilot prompt status** - Submit a natural language prompt to pull the status of a Security Copilot evaluation. After completion, the evaluation result will then be returned to your workflow.



The screenshot displays the 'Security Copilot Test Agent' interface with the 'Actions' tab selected. A green notification bar at the top states 'Your action was added to the agent'. Below it, a blue banner reads 'Enable your copilot to respond with the best combination of actions and topics using generative AI (preview)' with a 'Go to Generative AI' button. The main area features a '+ Add an action' button and a search bar labeled 'Search actions'. A table lists two actions:

Name	Trigger	Description	Editing	Last modified	Errors	Enabled
 Fetch prompt status	 Triggered by age...	Fetch the evaluation status				 On
 Submit a Security Copilot prompt	 Triggered by age...	Submit query to Security ...				 On





Phishing Triage



Alert Triage in DLP



Alert Triage in IRM



Conditional Access Optimization



Vulnerability Remediation



Threat Intelligence Briefing

“

Grazie!

Q&A

”

