



Security, Governance & Trust in AI

AI Summit 2025

Leadership & Outcomes

Joe Steiner

Solutions Architect

jsteiner@concurrency.com

Security, Governance & Trust – Complimentary Session

- **30 minutes with our experts** to explore security, governance, and trust best practices.
- Get **practical insights and answers** to your specific questions.
- **Schedule a follow-up** — we'll drop the bookings link in the chat during the session.

Importance of Trust

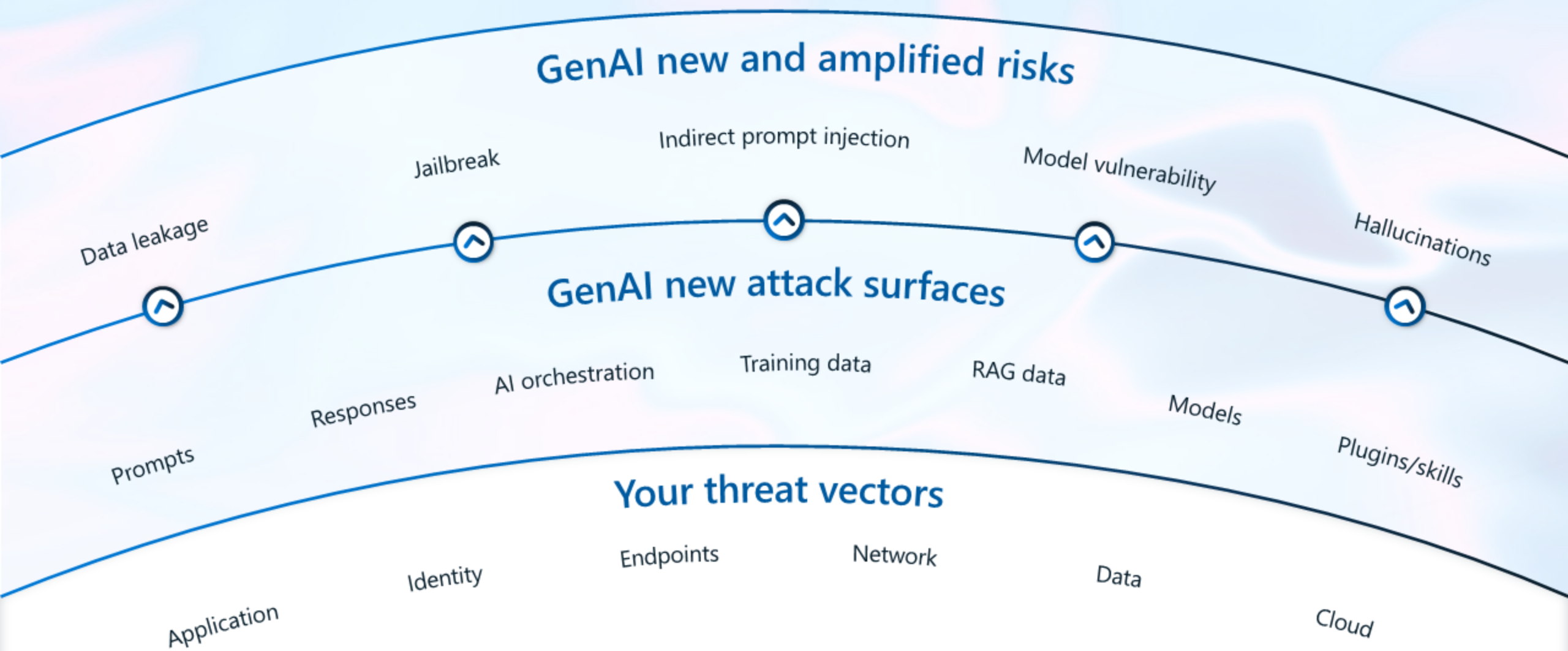
ALL New Technologies Bring Both **RISK** & **REWARD**.

- Reducing **RISK** → Increased **TRUST**
- Increased **TRUST** → Increased **ADOPTION**
- Increased **ADOPTION** → Increased **REWARD**
- Increased **REWARD** with Reduced **RISK** =
Greater **NET BENEFIT**

SECURITY & GOVERNANCE can reduce **RISK** & increase **TRUST**



GenAI attack surfaces introduce new and amplified risks



Zero Trust Security Model

“Never trust, always verify.”

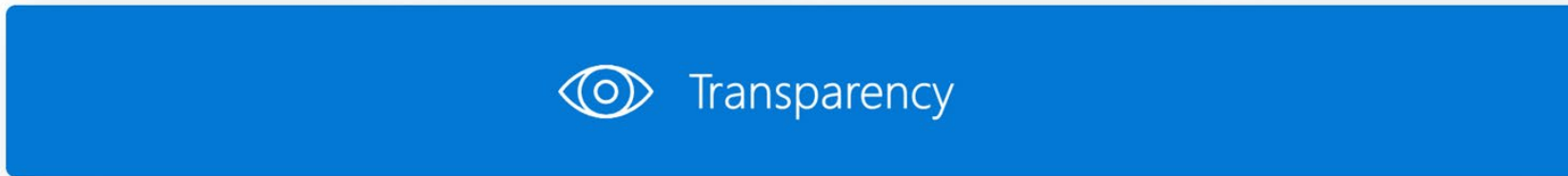
Security strategy that assumes breach and verifies each request as though it originated from an uncontrolled network.

Strongest security strategy for an Enterprise Cloud, AI & Automation Foundation

Core principles:

- Verify Explicitly
- Least Privilege Access
- Assume Breach

Microsoft Responsible AI Principles

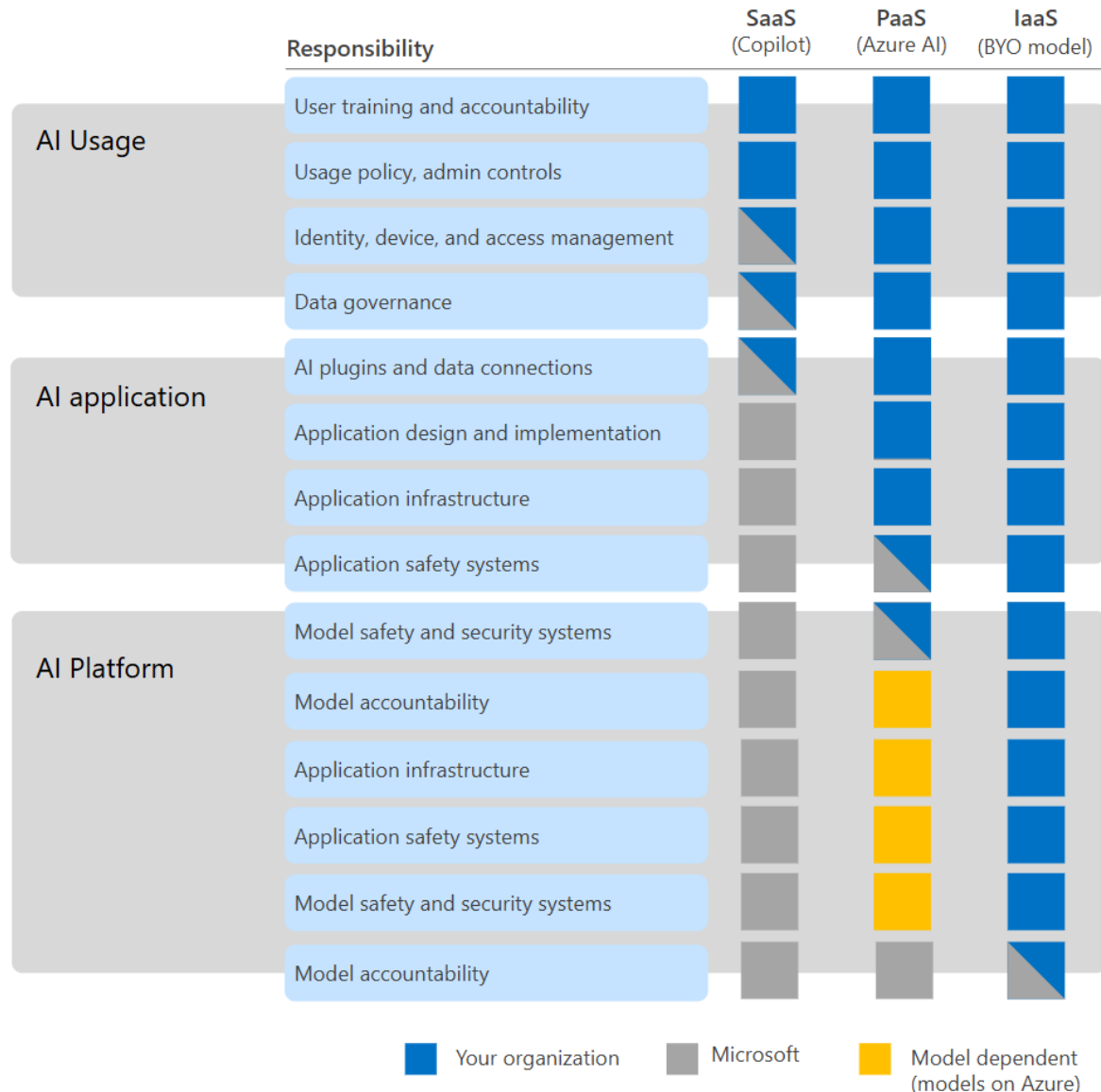


Trust of AI
extends beyond
Security concerns

Microsoft's
Responsible AI
Framework &
Dashboard assist
with this



Shared Responsibility Model



Different models of AI provide for varying degrees of responsibility

In all models the org is responsible for:

- User Behavior
- Identity & Access Mgmt
- Data Governance

Securing & Governing AI



Begin with crafting an AI Policy
Educate your Employees
Enforce across these areas



Identity
Management



Application &
Access Control



Data Protection



Ongoing Monitoring
& Governance



Responsible AI
Model Management

USER

- Identity in Entra ID
- Conditional Access rules for restricting or enabling use of different apps or agents in different scenarios

AI AGENT

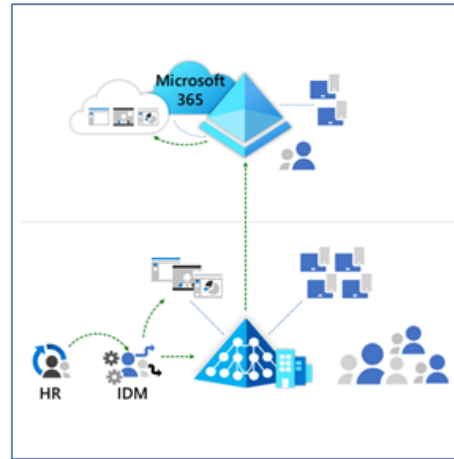
- Identity in Entra ID
- Control Permissions for who can use the agent
- Control Permissions for what systems the agent can interface with

Cloud Entra & Intune Architecture Progression

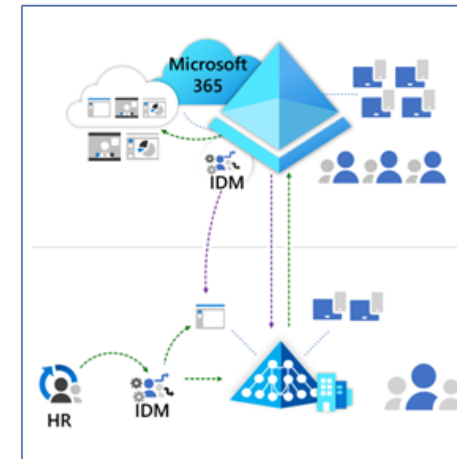
Cloud attached



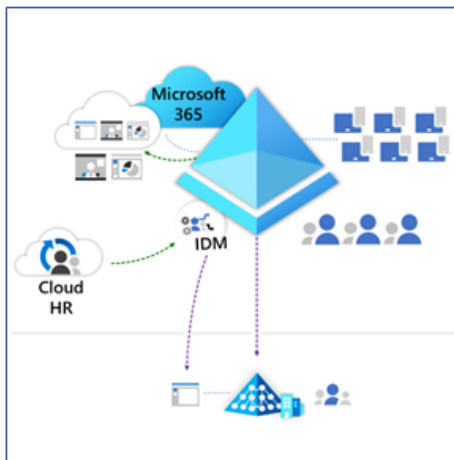
Hybrid



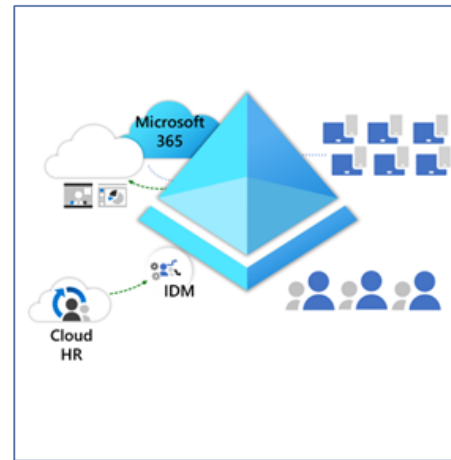
Cloud first



AD minimized



100% cloud



1. Cloud Attached
2. Hybrid
3. **Cloud First**
4. **AD Minimized**
5. **100% Cloud**



USER

- Restrict access to only Trusted apps or agents
- Prevent company data from entering untrusted apps or agents
- Restrict access to AI agents that are beyond the Users scope

AI AGENT

- Restrict access to only Trusted & Permissioned users
- Restrict access to only Trusted systems
- Understand who and what the agent can interface with

Restricting Access to Trusted AI Apps only

Microsoft Defender

Cloud Discovery







Created on Oct 13, 2023, 8:16 AM

Dashboard Discovered apps Discovered resources IP addresses Users

Queries: Select a query ☐ Advanced filter

Apps: App tag: ☐ ☐ None Risk score: 0 Compliance risk factor: Select factors Security risk factor: Select factors

☐ Bulk selection 1 - 6 of 6 discovered apps

App	Risk ...	Tags	Traffic	Upload	Transact...	Users	IP addre...	Last see...	Actions
 Microsoft Bing Chat Generative AI	<div><div></div></div> 10		20 MB	6 MB	167	142	120	Sep 14, 2023	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="button" value="⋮"/>
 Google Bard Generative AI	<div><div></div></div> 9		2 MB	447 KB	13	13	11	Sep 14, 2023	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="button" value="⋮"/>
 OpenAI ChatGPT Generative AI	<div><div></div></div> 8		16 MB	4 MB	134	121	89	Sep 14, 2023	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="button" value="⋮"/>
 Soundful Generative AI	<div><div></div></div> 5		3 MB	962 KB	28	28	21	Sep 14, 2023	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="button" value="⋮"/>
 Nichesss Generative AI	<div><div></div></div> 4		2 MB	550 KB	16	16	9	Sep 14, 2023	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="button" value="⋮"/>
 Framework Generative AI	<div><div></div></div> 3 UNSANCTIONED		2 MB	481 KB	14	14	9	Sep 14, 2023	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="button" value="⋮"/>

Preventing Data from Entering Untrusted AI

The screenshot shows a Microsoft Word window with the title bar 'Document1 - Word' and a 'Confidential' sensitivity label. The ribbon is set to 'Home'. A yellow banner at the top of the document area reads: 'POLICY TIP Your organization automatically applied the sensitivity: Confidential/Project Obsidian. OK'. The document content is titled 'FAQs for Project Obsidian' and includes the following sections:

- What is Project Obsidian?**

Project Obsidian is a software product that helps you analyze and visualize data from various sources. It allows you to create interactive dashboards, charts, graphs, and reports that can be shared with others.
- Who can use Project Obsidian?**

Project Obsidian is designed for anyone who needs to work with data, whether you are a student, a researcher, a business analyst, or a data enthusiast. You do not need any coding skills or prior experience to use Project Obsidian.
- How can I get Project Obsidian?**

Project Obsidian is currently in beta testing and will be available for public release soon. You can sign up for the beta program at <https://www.projectobsidian.com/beta> and get early access to the product.
- What are the features of Project Obsidian?**

Project Obsidian has many features that make data analysis and visualization easy and fun. Some of the features are:

 - Data import: You can import data from various sources, such as CSV, Excel, JSON, SQL, and more.
 - Data transformation: You can clean, filter, sort, group, and aggregate your data using a simple drag-and-drop interface.

The status bar at the bottom indicates 'Page 1 of 1', '331 words', and 'Confidential/Project Obsidian'. A large 'CONFIDENTIAL' watermark is visible diagonally across the page.

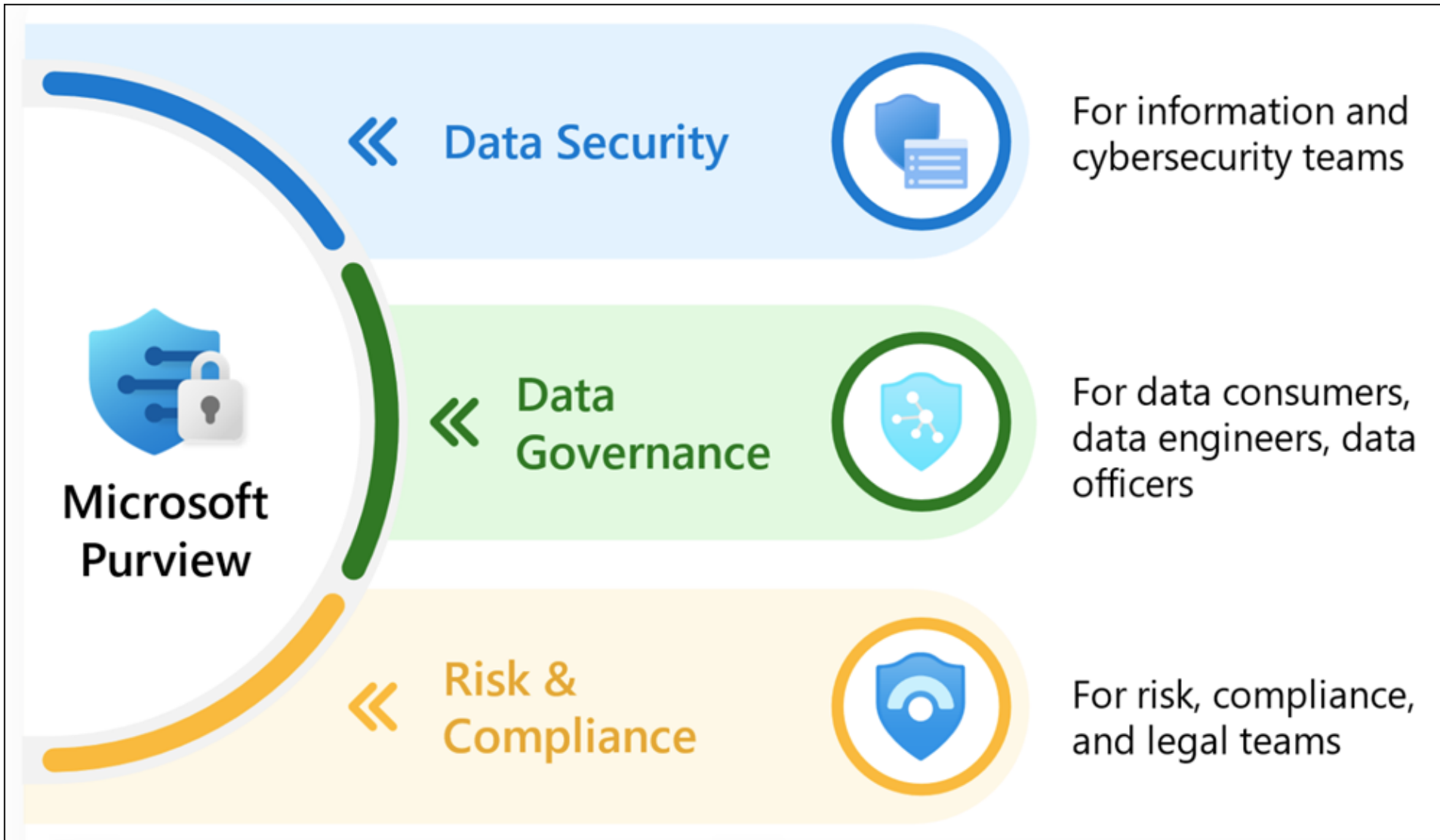
USER

- Establish & teach a data protection standard
- Restrict sensitive data to only those who should have access
- Use Data Loss Prevention (DLP) to protect

AI AGENT

- Ensure data protection standard is honored
 - Ensure agent only uses/shows data the user has access to
 - Protect response/output in alignment with data protection standards
 - Restrict agent from making changes to data it should not

Microsoft Purview



Purview covers multiple offerings:

- Unstructured data via documents (incl in M365 E3 or E5)
- Structured data via databases (pay as you go in Azure)



Can you summarize an FAQ about Project Obsidian and reference important documents for the project so I can send it out in an email?



Work content and chats cannot be seen outside your organization

Here is a summary of the FAQ about Project Obsidian and some important documents for the project:

- **Obsidian FAQ** is a document authored by **Alex Wilber** that provides a guide to the new project management tool. It explains that Obsidian is a web-based project management tool designed to be simple, intuitive, and flexible. It is free for up to 5 users and 10 projects, with the option to upgrade to a premium plan for more features and unlimited users and projects. To get started, users need to create an account, sign in, and create their first project. Tasks can be added and assigned to users, with the ability to set due dates, add labels, and attach files ¹
- Some other important documents for Project Obsidian include:
 - **"Project Plan for Project Obsidian"** is a document authored by **Alex Wilber** that provides an overview of the project objectives, scope, deliverables, timeline, and resources ²
 - **"Onbsidia"** is a PowerPoint presentation titled "Onbsidia" that provides an overview of the Project Obsidian kickoff ²
 - **"Design update"** is a document authored by **MOD Administrator** that discusses the need for building dedicated chips to power future cars and compares the advantages and disadvantages of ARM and x86 chip architectures ³

You can use this information to create an email summarizing the FAQ and important documents for Project Obsidian. Let me know if you need any further assistance.

AI-generated content may be incorrect

3 References ^

1

Obsidian FAQ

General\All Employees (unrestricted)

Word Alex Wilber modified on 11/07/23

"Obsidian FAQ A guide to the new project management tool What is Obsidian? Obsidian is a project management tool that helps you organize your tasks, track your progress, and ..."

2

Document3

Document3

General\All Employees (unrestricted)

Word Alex Wilber modified on 11/08/23

"Project Obsidian: - **Project Plan for Project Obsidian** is a document authored by **project objec...



New chat

Ask me anything or type / to add people, files, and more



0/4000

Secure by default with Microsoft Purview



Foundational

Start with recommended labels



Managed

Address files with highest sensitivity



Optimized

Expand to your entire M365 data estate



Strategic

Operate, expand, and retroactive actions

Activities

- Start with default labels and protection at file and site level
- Turn on data security pre-requisites and adv. analytics
- Train users on managing exceptions
- Turn on DLP for labeled content

- Manually configure priority sites default library labeling
- Autolabeling for credentials and contextual conditions
- Turn on DLP for content that is not labeled
- Turn on Adaptive Protection and data leak behavioral rules

- Auto-label sensitive files on clients (low thresholds)
- Simulate auto-labeling sensitive files at rest
- Reduce false positives with advanced classifiers
- Automate and improve M365 protection to historical and in use data

- Operational review of user labeling behaviors
- Iterate with new labeling scenarios
- Set up accountability chain and lifecycle management
- Extend protection to Azure SQL and non-M365 storage

Outcomes

M365 new/updated content protected

M365 priority content protected

M365 historical content protected

Protection beyond M365

Efforts

1 week

2 weeks

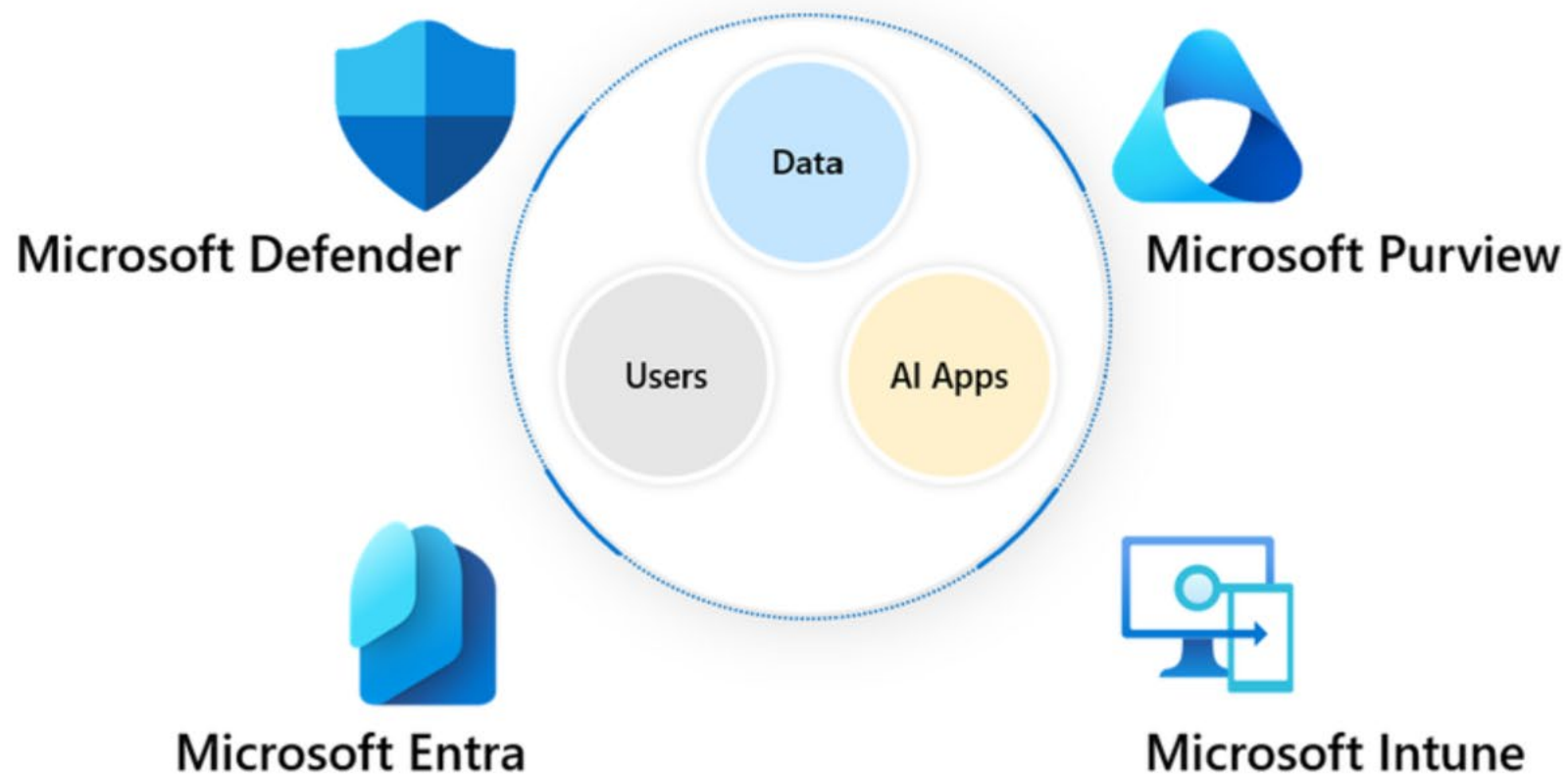
2 weeks iterations

Situational

Deployment guide available: <https://aka.ms/PurviewDeploymentModels/SecureByDefault>

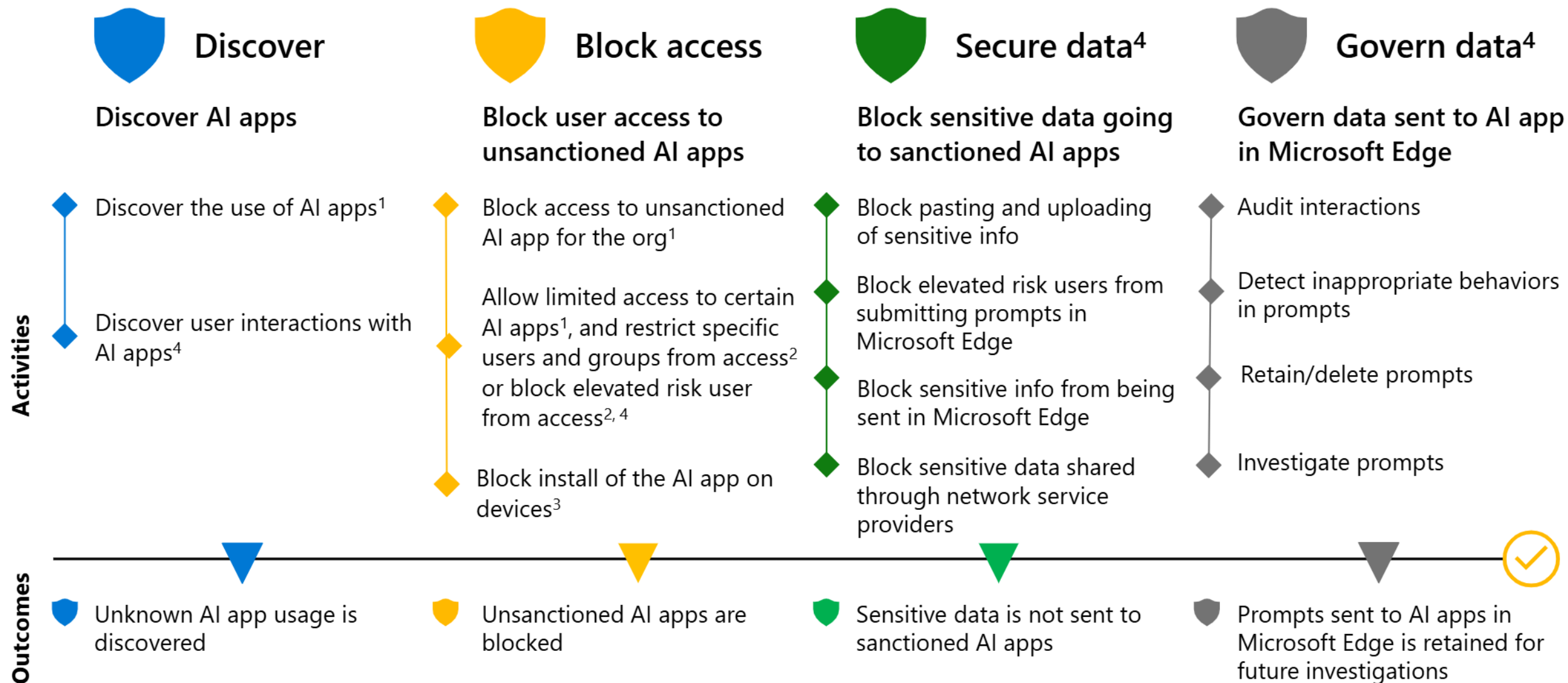
Last updated: May 4, 2025

Microsoft Security Portfolio



Prevent data leak to shadow AI with Microsoft

1. Defender for Cloud Apps
2. Entra
3. Intune
4. Purview



Ongoing Monitoring & Governance

Continue to refine AI Policy over time

Identity

- Entra ID Governance to monitor security group assignment

App & Access

- AI Hub in Purview to monitor AI usage
- Defender XDR & Risk Assessments

Data

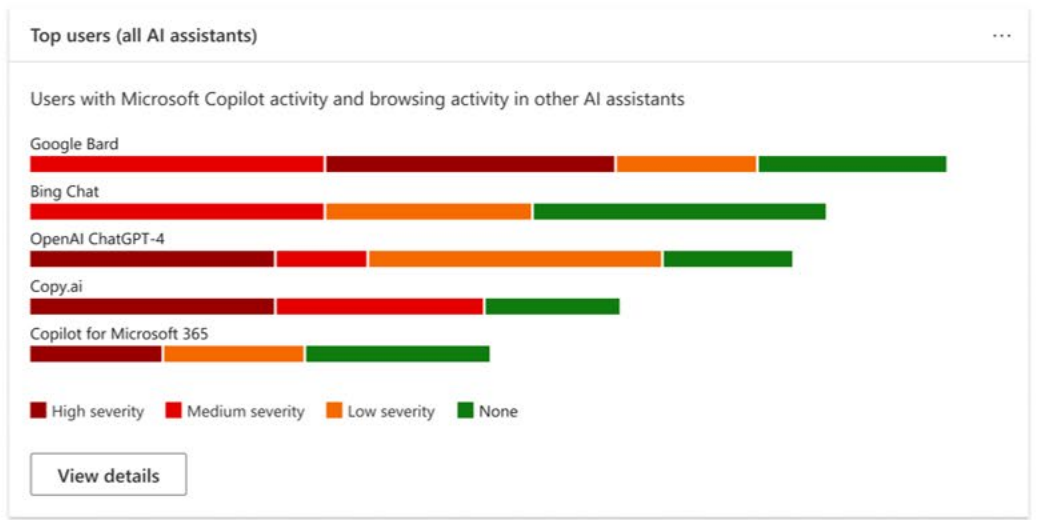
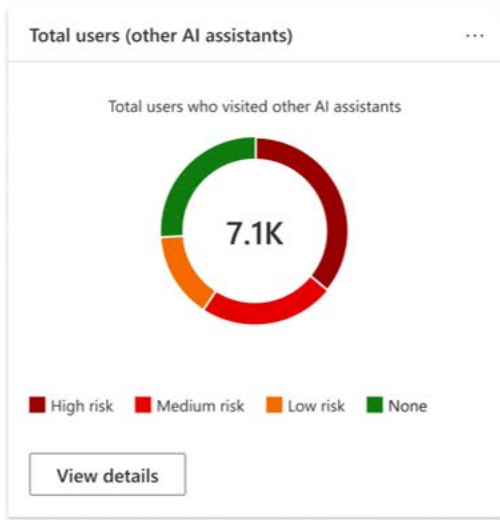
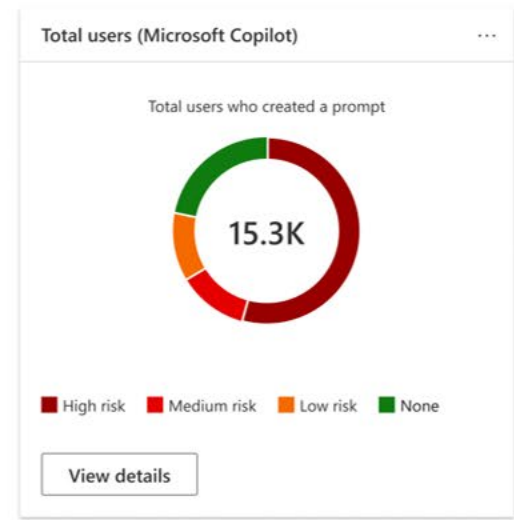
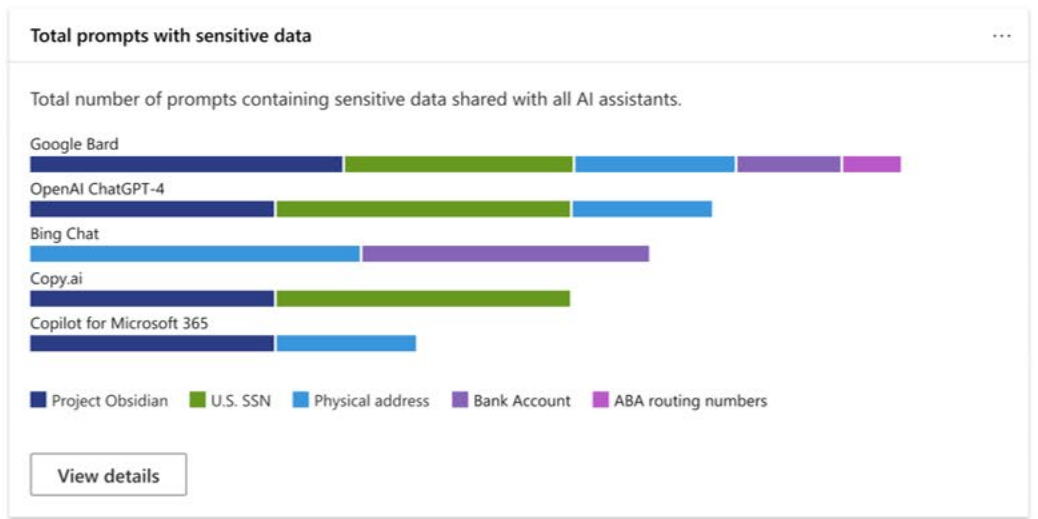
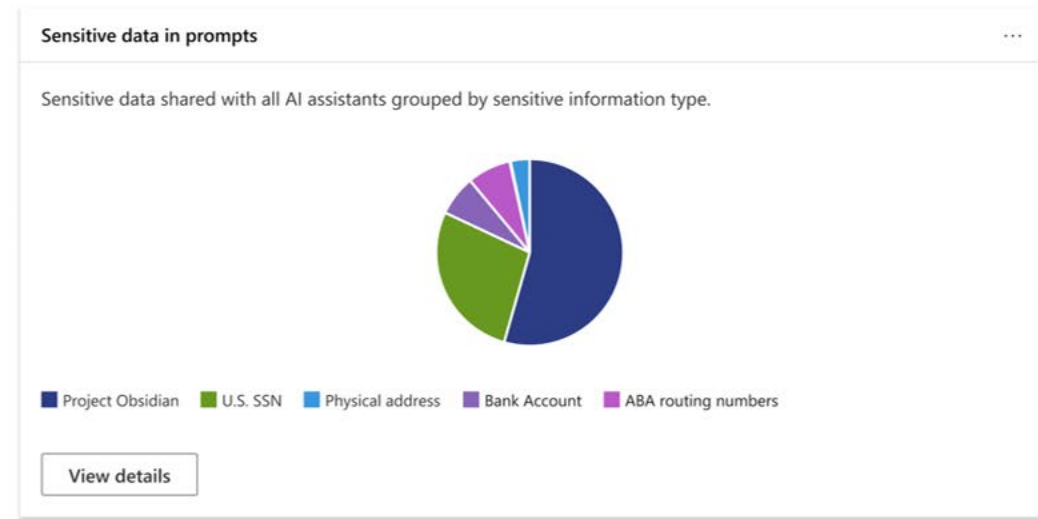
- Purview Audit, Communication Compliance, eDiscovery

AI hub (preview)

Analytics

Policies










Activity explorer





















Pending (19) Resolved (0) Exports

Filter  Save the query  Reset  Filters


Body/Subject: Any Date: Any Sender: Any Tags: Any



1 of 19 selected 

<input type="checkbox"/>		Subject	Tags	Sender	Recipients	Sentiment
<input type="checkbox"/>		Copilot in Word	...	Alex Wilber <Alex...>	Copilot	Positive
<input type="checkbox"/>		Copilot in BizChat	...	Alex Wilber <Alex...>	Copilot	Positive
<input type="checkbox"/>		Copilot in BizChat	...	Diego Siciliani <Di...>	Copilot	Positive
<input type="checkbox"/>		Copilot in BizChat	...	Diego Siciliani <Di...>	Copilot	Positive
<input type="checkbox"/>		Copilot in BizChat	...	Diego Siciliani <Di...>	Copilot	Positive
<input checked="" type="checkbox"/>		Copilot in BizChat	...	Diego Siciliani <Di...>	Copilot	Positive
<input type="checkbox"/>		Copilot in BizChat	...	Diego Siciliani <Di...>	Copilot	Positive
<input type="checkbox"/>		Copilot in BizChat	...	Diego Siciliani <Di...>	Copilot	Positive
<input type="checkbox"/>		Copilot in PowerPo...	...	Alex Wilber <Alex...>	Copilot	Neutral
<input type="checkbox"/>		Copilot in PowerPo...	...	Alex Wilber <Alex...>	Copilot	Negative
<input type="checkbox"/>		Copilot in BizChat	...	Alex Wilber <Alex...>	Copilot	Neutral
<input type="checkbox"/>		Copilot in BizChat	...	Alex Wilber <Alex...>	Copilot	Neutral
<input type="checkbox"/>		Copilot in Word	...	Alex Wilber <Alex...>	Copilot	Neutral
<input type="checkbox"/>		Copilot in Word	...	Alex Wilber <Alex...>	Copilot	Positive
<input type="checkbox"/>		Copilot in BizChat	...	Diego Siciliani <Di...>	Copilot	Positive

Copilot in BizChat



Source Plain Text User history

 Conditions detected: Gifts & entertainment (take these tickets as a gift) 

From: Diego Siciliani <DiegoS@MODERNCOMMS382604.OnMicrosoft.com>
Sent on: Thursday, February 1, 2024 6:03:40 AM
To: Copilot <>
Subject: Copilot in BizChat

can you help rephrase this sentence so it's less obvious that I am giving our free gifts to clients? "I would love for you to take these tickets as a gift as an expression of my gratitude for all of the hard work you put in"

Resolve

Notify

Tag as

Escalate

Escalate for investigation

Alerts > Suspicious interaction with Copilot for Microsoft 365 - Finance related file access

alex

Priority ▲ 1045

Microsoft Copilot for Microsoft 365



What happened



A risky user had suspicious interaction with Copilot for Microsoft 365, in which files that may be related to finance were accessed. An attacker might have compromised a user account and is using it t...

[Read more](#)

Related events

[Expand all](#)2/20/2024
11:05:01 AM

CopilotInteraction action by Alex in Microsoft Copilot for Microsoft 365

2/13/2024
1:15:13 PM

CopilotInteraction action by Alex in Microsoft Copilot for Microsoft 365

2/13/2024
2:13:59 PM

CopilotInteraction action by Alex in Microsoft Copilot for Microsoft 365



Suspicious interaction with Copilot for Microsoft 365 - Finance related file access

High Unknown Resolved

[Manage alert](#)

INSIGHT

Quickly classify this alert

Classify alerts to improve alert accuracy and get more insights about threats to your organization.

[Classify alert](#)

Alert state



Classification

Not Set

[Set Classification](#)

Assigned to

User-
assafyatziv@mcas-
test9.com

Alert details



Category

Collection

MITRE ATT&CK

Techniques

[T1119: Automated...](#)

Detection source

Service source

Responsible AI Model Management

If you have built a custom AI model then in addition to securing & governing access and use of the model you are responsible for managing the model you created in terms of:

- Privacy & Security
- Reliability & Safety
- Fairness & Inclusiveness
- Transparency
- Accountability

Microsoft's AI Content Safety & Responsible AI Dashboard can be used to assist

Model Debugging via Responsible AI dashboard

Identify



Diagnose



Mitigate



Error Analysis

Identify cohorts with high error rate versus benchmark and visualize how the error rate distributes



Fairness Assessment

Evaluate model fairness by exploring a variety of model performance metrics across sensitive groups



Model Interpretability

Interpret and debug model.



Counterfactual Analysis and What If

Generate diverse counterfactual explanations for debugging.
Perform feature perturbations



Exploratory Data Analysis

Understand dataset characteristics



Unfairness Mitigation

Mitigate fairness issues
(via Fairlearn.org)



Data Enhancements

Enhance your dataset and retrain model



Model
Comparison



Compare

Backward
Compatibility

Decision Making via Responsible AI dashboard

Understand data



Inform Actions



Exploratory-Data-Analysis
Understand dataset characteristics



Causal Inference
Understand the causal impact of
your features on real-world
outcomes



Counterfactual Analysis
Generate diverse counterfactual
explanations for informing end
users

Identity Management

- Manage AI Agents like Humans
- Verify all User requests, Use Least-Privileged Access & Secure based on Risk
- Log Everything

Application & Access Control

- Restrict use of Untrusted AI Agents & Applications
- Manage Applications on all devices

Data Protection

- Know your data
- Secure the data at the source and in use
- Track where your data goes

Ongoing Monitoring & Governance

- Monitor AI Usage
- Monitor Data Access
- Monitor AI Model Performance
- Log Everything

AI Model Management

- Continue to “Debug” your AI Model
- Continue to ensure you understand the Data used and Systems connected

Security, Governance & Trust – Complimentary Session

- **30 minutes with our experts** to explore security, governance, and trust best practices.
- Get **practical insights and answers** to your specific questions.
- **Schedule a follow-up** — we'll drop the bookings link in the chat during the session.

Thank you!



Joe Steiner
Solutions Architect



Follow Concurrency!

