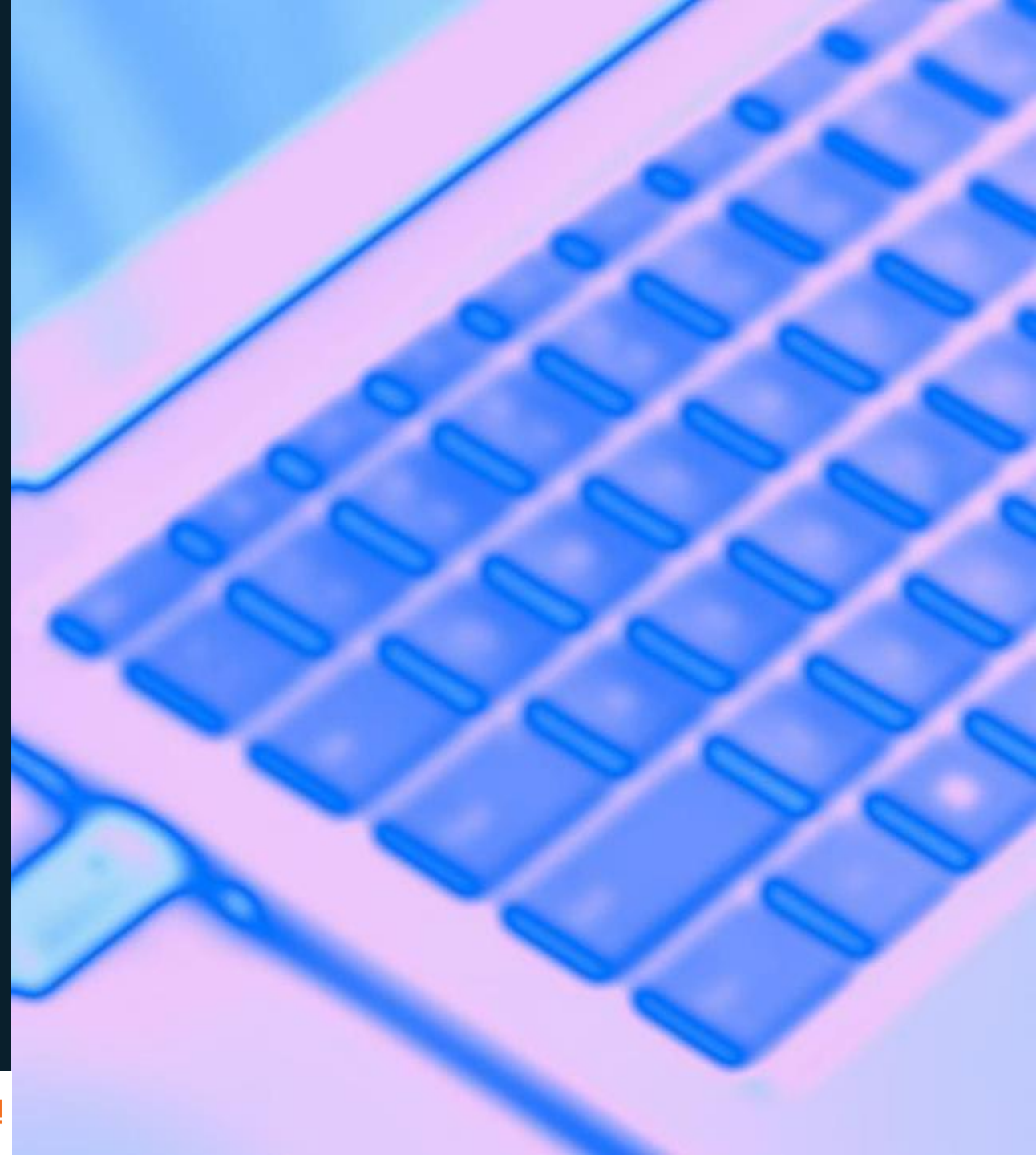


Secure data in the age of AI

Luděk Suk
Microsoft



Generative AI is reshaping the world but there are associated data security risks..

User creates document without proper access controls making it easy for other users to reference it in Copilot



Data oversharing by negligent insider



User asks generative AI to find information on a secret project and leaks it to the press for personal gain



Data leak by disgruntled insider



User negligently shares sensitive data in consumer generative AI apps



Data leak by negligent insider



Microsoft Purview

Integrated solutions to secure & govern your entire data estate

DATA SECURITY

Secure data across its lifecycle,
wherever it lives

Information Protection
Data Loss Prevention
Insider Risk Management
Data Security Posture Management

DATA GOVERNANCE

Responsibly unlock value
creation from data

Data Discovery
Data Quality
Data Curation
Data Estate Insights

DATA COMPLIANCE

Manage critical risks and
regulatory requirements

Compliance Manager
eDiscovery and Audit
Communication Compliance
Data Lifecycle Management
Records Management

Unstructured & Structured data

Traditional and AI generated data

Microsoft 365 and Multi-cloud

Shared platform capabilities

Data Map, Data Classification, Data Labels, Audit, Data Connectors

Microsoft Purview

Integrated solutions to secure & govern your entire data estate

DATA SECURITY

Secure data across its lifecycle,
wherever it lives

Information Protection
Data Loss Prevention
Insider Risk Management
Data Security Posture Management

DATA GOVERNANCE

Responsibly unlock value
creation from data

Data Discovery
Data Quality
Data Curation
Data Estate Insights

DATA COMPLIANCE

Manage critical risks and
regulatory requirements

Compliance Manager
eDiscovery and Audit
Communication Compliance
Data Lifecycle Management
Records Management

Unstructured & Structured data

Traditional and AI generated data

Microsoft 365 and Multi-cloud

Shared platform capabilities

Data Map, Data Classification, Data Labels, Audit, Data Connectors

Actions you can take to secure and govern AI apps

Discover

AI risks associated with data, access, and AI apps

Protect

AI and sensitive data with end-to-end security

Govern

AI usage to comply with regulatory and code-of-conduct policies

Actions you can take to secure and govern AI apps

Discover

AI risks associated with data, access, and AI apps

Protect

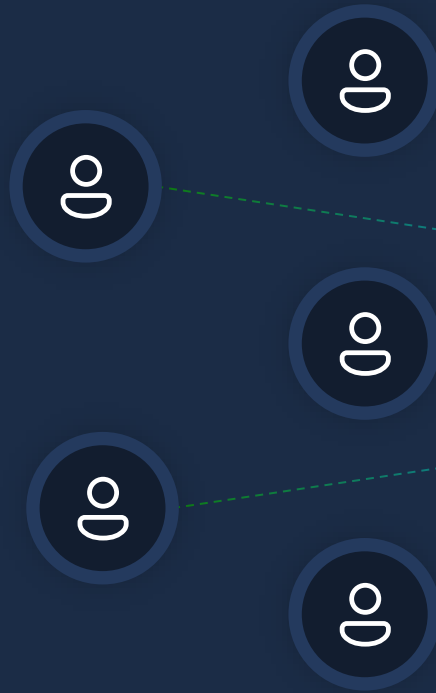
AI and sensitive data with end-to-end security

Govern

AI usage to comply with regulatory and code-of-conduct policies

Discover AI risks associated with data, access, and AI apps

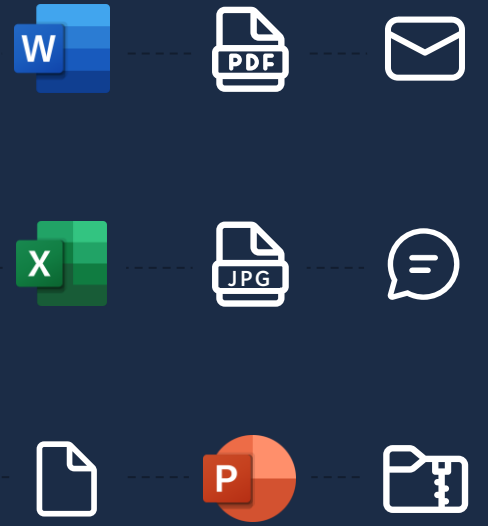
Who's using GenAI in your organization and how is access being governed?



What type of sensitive data is flowing in



Are outputs and interactions protected and well-governed?



What types of GenAI apps are being used in your organizations?



Data Security Posture Management for AI

Data security and compliance for the era of AI

- Home
- Solutions
 - DSPM for AI**
- Learn
 - Overview
 - Recommendations
 - Reports
 - Policies
- Settings
 - Activity explorer
 - Data assessments Preview

Data Security Posture Management for AI

Discover and secure all AI activity in Microsoft Copilot and other AI apps. Keep your data safe and stay on track with industry regulations. [Learn more about DSPM for AI](#)

Get started

	Activate Microsoft Purview Audit Get insights into user interactions with Microsoft Copilot experiences.	Required	7 Minutes
	Install Microsoft Purview browser extension Detect risky user activity and get insights into user interactions with other AI apps.	Required	1 Hour
	Onboard devices to Microsoft Purview Protect sensitive data from leaking to other AI apps.	Required	1 Hour
	Extend your insights for data discovery Discover sensitive data in user interactions with other AI apps.	Required	10 Minutes

Recommendations

[View all recommendations →](#)

New AI regulations

Get guided assistance to AI regulations

Stay on track with newly established industry regulations for AI, such as ISO 42001 and NIST AI RMF. To ensure safe AI interactions, we've identified the key actions associated with these regulations.

Interactions with sensitive data
Last 30 days

509

[View details](#)

Data Security Investigations

Protect sensitive data referenced in Copilot responses

In the last 30 days, 0 unprotected files were referenced in Copilot responses. Start a data investigation or take steps to avoid potential oversharing of sensitive data.

Unlabeled files in Copilot responses
Last 30 days

Unlabeled files
0

SharePoint Sites with unlabeled files
0

[View details](#)

Reports

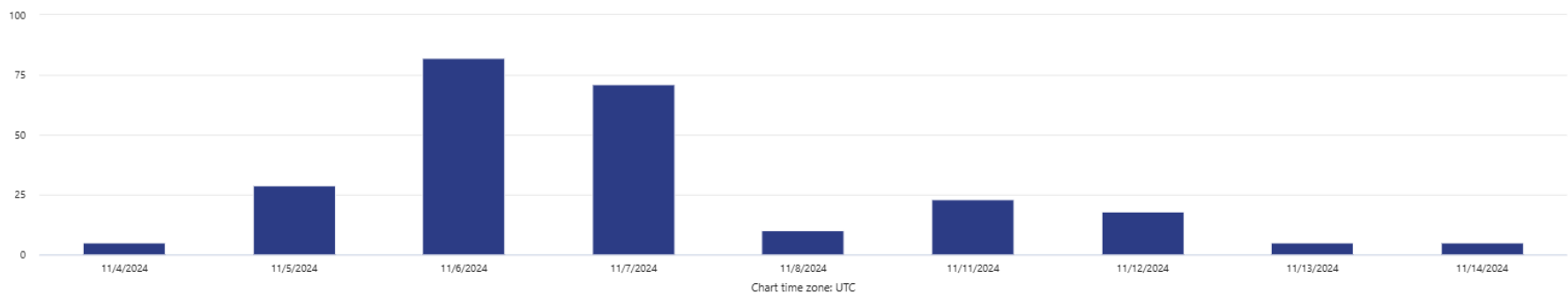
[View all reports →](#)

Activity explorer

Review AI activity including AI interactions (prompts and response), activity with sensitive info types, and more.

Filters: Timestamp: 10/16/2024-11/14/2024 Activity type: AI Interaction AI app category: Microsoft Copilot Experiences App: Any App accessed in: Any User: Any User risk level: Any Sensitive info type: Any

Resources accessed: Any Sensitive files referenced: Any Scope: Any Reset all



Export

248 items

<input type="checkbox"/>	Activity type	User	User risk level	Timestamp (UTC)	AI app category	App	App accessed in	Sensitive info type	Resources access...	Sensitive files ref...
<input type="checkbox"/>	AI Interaction	ericat@fabrikampurviewignite2024.o...		Nov 14, 2024 8:48 PM	Microsoft Copilot Experie...		Word		No	No
<input type="checkbox"/>	AI Interaction	ericat@fabrikampurviewignite2024.o...		Nov 14, 2024 8:48 PM	Microsoft Copilot Experie...		Word		No	No
<input type="checkbox"/>	AI Interaction	ericat@fabrikampurviewignite2024.o...		Nov 14, 2024 8:47 PM	Microsoft Copilot Experie...		Office		Yes	Yes
<input type="checkbox"/>	AI Interaction	ericat@fabrikampurviewignite2024.o...		Nov 14, 2024 8:46 PM	Microsoft Copilot Experie...		Office		Yes	Yes
<input type="checkbox"/>	AI Interaction	ericat@fabrikampurviewignite2024.o...		Nov 14, 2024 8:42 PM	Microsoft Copilot Experie...		Office		Yes	Yes
<input type="checkbox"/>	AI Interaction	lisab@fabrikampurviewignite2024.o...		Nov 13, 2024 8:56 PM	Microsoft Copilot Experie...		Office		Yes	No
<input type="checkbox"/>	AI Interaction	lisab@fabrikampurviewignite2024.o...		Nov 13, 2024 7:43 PM	Microsoft Copilot Experie...		Office		Yes	No
<input type="checkbox"/>	AI Interaction	andrewm@fabrikampurviewignite20...	Medium	Nov 13, 2024 7:42 PM	Microsoft Copilot Experie...		Word		No	No

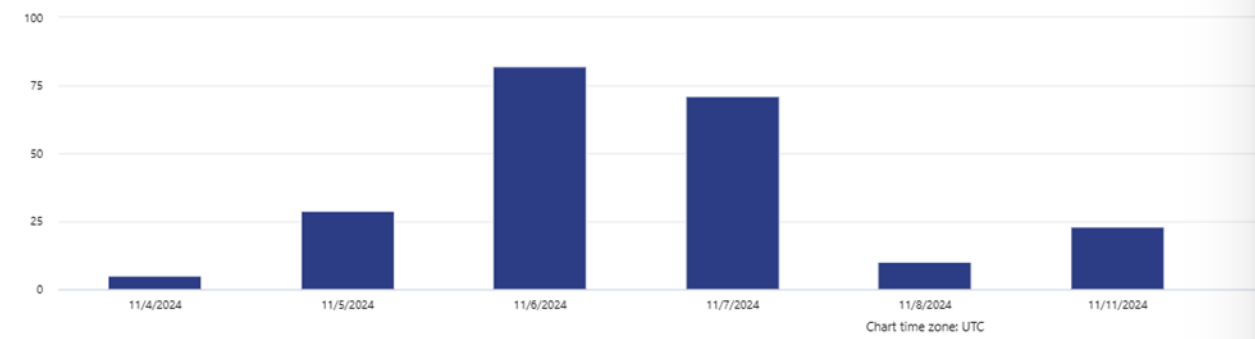
- Home
- Solutions
 - DSPM for AI
- Learn
 - Overview
 - Recommendations
 - Reports
 - Policies
- Settings
 - Activity explorer
 - Data assessments Preview

Activity explorer

Review AI activity including AI interactions (prompts and response), activity with sensitive info types, and more.

Filters: Timestamp: 10/16/2024-11/14/2024 Activity type: AI Interaction AI app category: Microsoft Copilot Experiences App: Any App accessed in: Any User: Any

Resources accessed: Any Sensitive files referenced: Any Scope: Any Reset all



Export Selected Items

<input type="checkbox"/>	Activity type	User	User risk level	Timestamp (UTC)	AI app category	App
<input checked="" type="checkbox"/>	AI Interaction	ericat@fabrikampurviewignite2024.o...		Nov 14, 2024 8:48 PM	Microsoft Copilot Experi...	
<input type="checkbox"/>	AI Interaction	ericat@fabrikampurviewignite2024.o...		Nov 14, 2024 8:48 PM	Microsoft Copilot Experi...	
<input type="checkbox"/>	AI Interaction	ericat@fabrikampurviewignite2024.o...		Nov 14, 2024 8:47 PM	Microsoft Copilot Experi...	
<input type="checkbox"/>	AI Interaction	ericat@fabrikampurviewignite2024.o...		Nov 14, 2024 8:46 PM	Microsoft Copilot Experi...	
<input type="checkbox"/>	AI Interaction	ericat@fabrikampurviewignite2024.o...		Nov 14, 2024 8:42 PM	Microsoft Copilot Experi...	
<input type="checkbox"/>	AI Interaction	lisab@fabrikampurviewignite2024.o...		Nov 13, 2024 8:56 PM	Microsoft Copilot Experi...	
<input type="checkbox"/>	AI Interaction	lisab@fabrikampurviewignite2024.o...		Nov 13, 2024 7:43 PM	Microsoft Copilot Experi...	
<input type="checkbox"/>	AI Interaction	andrewm@fabrikampurviewignite20...	Medium	Nov 13, 2024 7:42 PM	Microsoft Copilot Experi...	

AI Interaction

Network ID
a04bb747-25ae-47a1-8ad0-6e3d0cbe37f5

Client IP
2600:8800:7089:f00:69bb:c58:2694:2f12

About this user

User
Erica T

[View more user details in insider risk management](#)

App details

AI app category
Microsoft Copilot Experiences

App accessed in
Word

Interaction details

You can view prompts and responses with your Microsoft Purview Content Explorer Content Viewer role. [Learn more about permissions for DSPM for AI](#)

Prompt

[Project Obsidian M&A.docx](#)
[AutoGenerated]Summarize

Sensitive info types detected [View related classification activity](#)

Response

[Project Obsidian M&A.docx](#)
[AutoGenerated]Project Obsidian is a merger between Company A and Company B to create a more competitive and innovative market entity. The merger aims to combine strengths, enhance market share, improve

11/7/2024

11/8/2024

11/11/2024

Chart time zone: UTC

Timestamp (UTC)	AI app category	App
Nov 14, 2024 8:48 PM	Microsoft Copilot Experi...	
Nov 14, 2024 8:48 PM	Microsoft Copilot Experi...	
Nov 14, 2024 8:47 PM	Microsoft Copilot Experi...	
Nov 14, 2024 8:46 PM	Microsoft Copilot Experi...	
Nov 14, 2024 8:42 PM	Microsoft Copilot Experi...	
Nov 13, 2024 8:56 PM	Microsoft Copilot Experi...	
Nov 13, 2024 7:43 PM	Microsoft Copilot Experi...	
Nov 13, 2024 7:42 PM	Microsoft Copilot Experi...	
Nov 13, 2024 7:41 PM	Microsoft Copilot Experi...	

AI app category

Microsoft Copilot Experiences

App accessed in

Word

Interaction details

i You can view prompts and responses with your Microsoft Purview Content Explorer Content Viewer role. [Learn more about permissions for DSPM for AI](#)

Prompt

[Project Obsidian M&A.docx](#)
[AutoGenerated]Summarize

i Sensitive info types detected [View related classification activity](#)

Response

[Project Obsidian M&A.docx](#)
[AutoGenerated]Project Obsidian is a merger between Company A and Company B to create a more competitive and innovative market entity. The merger aims to combine strengths, enhance market share, improve efficiencies, and drive growth. **Objectives of the Merger**: The merger

i Sensitive info types detected [View related classification activity](#)

Actions you can take to secure and govern AI apps

Discover

AI risks associated with data, access, and AI apps

Protect

AI and sensitive data with end-to-end security

Govern

AI usage to comply with regulatory and code-of-conduct policies

Protect sensitive data used by AI apps



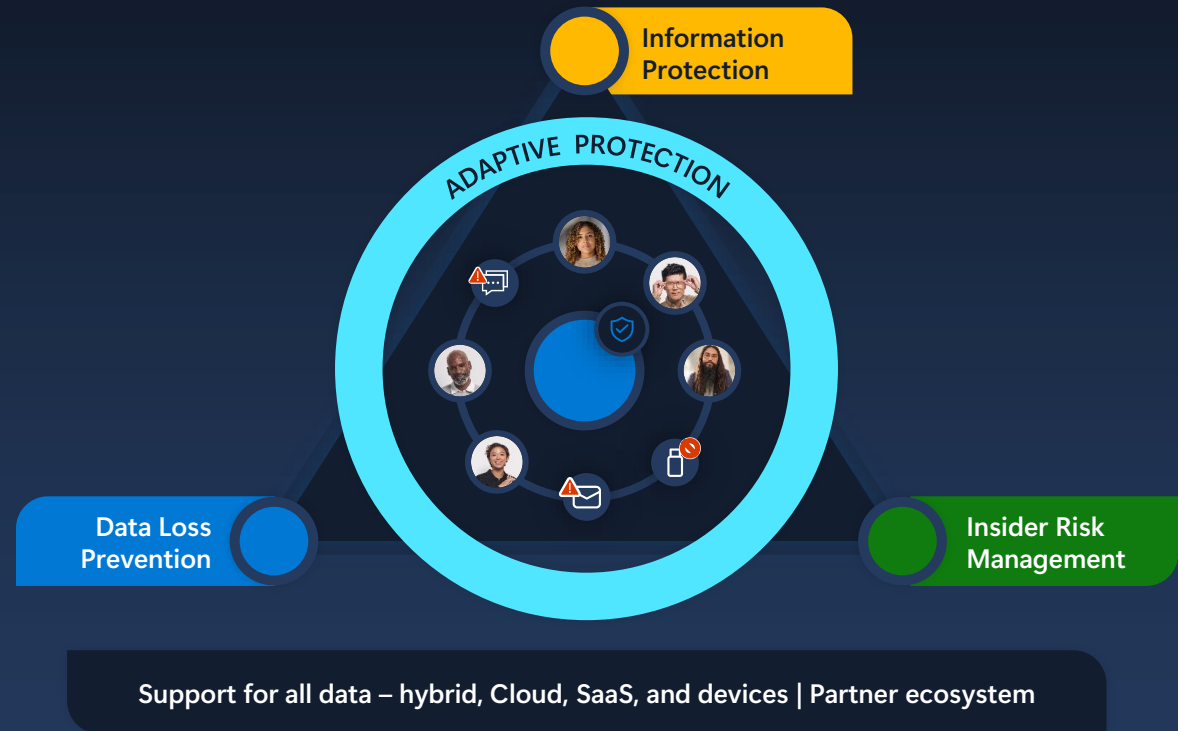
Automatically **discover, classify and label sensitive** data, and **prevent its unauthorized use** across apps, services, and devices.



Understand the **user intent and context around the use of sensitive data** to identify the most critical risks



Enable **Adaptive Protection** to assign high-risk users to appropriate DLP, Data Lifecycle Management, and Entra Conditional Access policies



Information Protection



What are the important documents for Project Obsidian

Here are some important documents for Project Obsidian:

- "Project plan for project obsidian" is a confidential document authored by **Alex Wilber** about next generation chip design. It provides overview of the project objectives, scope, deliverables, timelines, and resources. ①
- "Next generation chip" is a confidential document authored by **Joe Fromm** discusses the design of this next generation chip. It discusses the need for building dedicated chips to power future cars and compares the advantages and disadvantages of ARM and x86 chip architectures. ②
- "Obsidian" is a presentation that is authored by and last modified Tuesday evening by **Edison Gil**. It was used to kick off Project Obsidian. ③
- [2023 Obsidian Resources](#) is also related, but organization policy restricts me from discussing its content in this chat.



Copy



Share

Confidential \ Project Obsidian

Confidential data that requires protection, blocks users from copy and share of content.

3 references ^

1/30 AI-generated content may be incorrect

① **Project plan for project obsidian**

Confidential \ Project Obsidian

Word | Alex Wilber modified on 11/07/23

"Project Plan for Project Obsidian. A brief overview of the project objectives and scope. Project Objectives: The main objective of Project Obsidian is to develop..."

② **Next generation chip**

Confidential \ Project Obsidian

Word | Joe Fromm modified on 11/15/23

"A next generation chip that is more capable than any current generation in harnessing raw compute power while being able to efficiently..."

label from referenced content.



Label Inheritance

Generated artifacts such as chat or files must inherit the **most sensitive label** from referenced content.

- “Obsidian” is a presentation that is authored by and last modified Tuesday evening by Edison Gil. It was used to kick off Project Obsidian. ³
- [2023 Obsidian Resources](#) is also related, but organization policy restricts me from discussing its content in this chat.

Copy

Share

Confidential \ Project Obsidian

Confidential data that requires protection, blocks users from copy and share of content.

3 references ^



1/30

AI-generated content may be incorrect



1 Project plan for project obsidian

Confidential \ Project Obsidian

Word | Alex Wilber modified on 11/07/23

“Project Plan for Project Obsidian. A brief overview of the project objectives and scope. Project Objectives: The main objective of Project Obsidian is to develop...”

2 Next generation chip

Confidential \ Project Obsidian

Word | Joe Fromm modified on 11/15/23

“A next generation chip that is more capable than any current generation in harnessing raw compute power while being able to efficiently...”

3 Obsidian

General

Powerpoint | Edison Gil modified on 10/12/23

“Kick off for a new project...”

Who is Alex Wilbur?

What is the project objective of Project Obsidian

Timelines for Project Obsidian



Ask a work question or use / to reference people, files and more



Label Visibility

Chats referencing labeled information must display the label to the user.

Copilot generated content inherits the sensitivity label of referenced files

Generated contents inherit the sensitivity label from the source documents.

The screenshot shows a Microsoft Word document titled "AI hub A platform for data security and compliance for AI". The document is in the "Home" tab of the ribbon. A yellow banner at the top of the document area contains the following text: "SENSITIVITY Due to content created by Copilot, your organization automatically applied the sensitivity label: Confidential\Anyone (unrestricted)." A red arrow points from this banner to the main content of the document. The main content includes a title "AI hub: A new platform for data security and compliance for AI", a subtitle "A brief overview of the new features and benefits of AI hub", and three sections: "Introduction", "Key features of AI hub", and a list of features. The "Introduction" section states: "AI is transforming the way organizations operate, innovate, and compete. However, AI also brings new challenges and risks for data security and compliance, especially in the context of data privacy regulations and ethical standards. How can organizations ensure that their data and AI activities are secure, compliant, and trustworthy?". The "Key features of AI hub" section states: "AI hub offers a range of features that help organizations address the challenges and risks of AI, such as data breaches and leaks, unauthorized access and misuse, non-compliance and fines, reputational damage and loss of trust. Some of the key features of AI hub are:". The list of features includes: "Data discovery and classification: AI hub automatically scans and identifies the data sources and types that are used for AI, and assigns them a classification level". At the bottom of the document, there is a "Summarize How t..." button, a "1 of 1" indicator, and a "AI-generated content may be incorrect" warning. A blue box highlights the "Keep it" button and a text input field containing "For example, 'Make it more engaging'".

File Home Insert Layout References Review View Help

Calibri Light (H... 28 B I U ...

SENSITIVITY Due to content created by Copilot, your organization automatically applied the sensitivity label: Confidential\Anyone (unrestricted).

AI hub: A new platform for data security and compliance for AI

A brief overview of the new features and benefits of AI hub

Introduction

AI is transforming the way organizations operate, innovate, and compete. However, AI also brings new challenges and risks for data security and compliance, especially in the context of data privacy regulations and ethical standards. How can organizations ensure that their data and AI activities are secure, compliant, and trustworthy?

AI hub is a new platform that helps organizations manage, monitor, and secure their AI activities. AI hub provides a comprehensive solution for data discovery and classification, AI activity tracking and auditing, data protection and encryption, and compliance and governance policies. AI hub enables organizations to drive data security and compliance controls for AI, and to gain visibility and insights into their AI activities.

Key features of AI hub

AI hub offers a range of features that help organizations address the challenges and risks of AI, such as data breaches and leaks, unauthorized access and misuse, non-compliance and fines, reputational damage and loss of trust. Some of the key features of AI hub are:

- Data discovery and classification: AI hub automatically scans and identifies the data sources and types that are used for AI, and assigns them a classification level

Summarize How t... < 1 of 1 > AI-generated content may be incorrect

Keep it For example, "Make it more engaging"

Page 1 of 1 214 words English (U.S.) Editor Suggestions: Showing Confidential\Anyone (unrestricted) 90% Give Feedback to Microsoft

Data Loss Prevention



- Home
- Solutions
 - Data Loss Prevention
 - Overview
 - Policies**
 - Alerts
 - Activity explorer
 - Classifiers
 - Explorers
- Learn
- Settings
- Data Loss Prevention
 - Information Protection
 - Insider Risk Management

Policies

Use data loss prevention (DLP) policies to help identify and protect your organization's sensitive info. For example you can set up policies to help make sure information in email and docs isn't shared with the wrong people. [Learn more about DLP](#)

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


[View role groups](#)










+ Create policy ↓ Export ↻ Refresh 3 items Search [Customize columns](#)

<input type="checkbox"/>	Name		Priority	Last modified	Status
<input type="checkbox"/>	Microsoft AI hub - Discover sensitive prompts in AI assistants	⋮	0	Oct 29, 2024 1:45 PM	On
<input type="checkbox"/>	Microsoft AI hub - Adaptive Protection in AI assistants	⋮	1	Oct 29, 2024 1:56 PM	On
<input type="checkbox"/>	Default policy for Teams	⋮	2	Oct 29, 2024 6:50 PM	On

Template or custom policy Name Admin units **Locations** Policy settings Policy mode Finish

We'll apply the policy to data that's stored in the locations you choose.

 Protecting sensitive info in on-premises repositories (SharePoint sites and file shares) is now in preview. Note that there are prerequisite steps needed to support this new capability. [Learn more about the prerequisites](#)

Location	Scope	Actions
 Exchange email	Turn on location to scope	
 SharePoint sites	Turn on location to scope	
 OneDrive accounts	Turn on location to scope	
 Teams chat and channel messages	Turn on location to scope	
 Devices	Turn on location to scope	
 Instances	Turn on location to scope	
 On-premises repositories	Turn on location to scope	
 Fabric and Power BI workspaces	Turn on location to scope	
<input checked="" type="checkbox"/>  Microsoft 365 Copilot (preview)	All users & groups	Edit

Back

Next

Cancel

Template or custom policy Name Admin units Locations **Policy settings** Advanced DLP rules Policy mode Finish

Create rule

Content contains

Group name *

Default


Group operator

Any of these

Sensitivity labels

Obsidian Merger

Add

 Create group

+ Add condition


Actions

Use actions to protect content when the conditions are met.

Prevent Copilot from processing content

Content that matches your conditions won't be used by Copilot to generate responses.

[Learn more about this action](#)

 Currently, this action is supported only for labeled files in SharePoint and OneDrive that are processed in Microsoft 365 Chat experiences.

 Exclude content from processing in Copilot

Save

Cancel

Your status message is showing in chat and channels when people message or @mention you.

- Activity
- Chat
- Teams
- Calendar
- Calls
- OneDrive
- Viva Learning
- Apps

Copilot Using Project Specificati...

New chat

June 26, 2024 at 10:41 AM

Using [Project Specification for Project Obsidian.docx](#) generate a summary on the partners involved and the kind of tasks they will perform

[Project Specification for Project Obsidian.docx](#) is related, but organization policy restricts me from discussing its content in this chat.

View prompts

Message Copilot...

+ Add people, files, and more for better results

0 / 2000

📎 @ 📎 📎 ▶

Insider Risk Management



- Home
- Insider Risk Management
- Overview
- Recommended actions
- Alerts
- Cases
- Policies
- Users
- Reports
- Forensic evidence
- Adaptive protection (preview)
- Notice templates
- Audit log
- Solutions

Insider risk management > Alerts

Alerts

9 Spotlight | 4 Departing employee - 7 day | 84 Data Leaks: High Severity | 0 Assigned to me

Export | 1,292 alerts | Search | Customize columns

Filter set: Unsaved | Save | Severity: All values | Status: All values | Time detected (UTC): All values | Add filter | Reset all

ID	User	Policy	Status	Spotlight	Alert severity	Time detected	Assigned to	Case
31ac5f2b	AnonyFB12P@contoso.com	Risky AI Usage	New		High	4 hours ago		
21ac5f2b	AnonyEF34B@contoso.com	Data Leaks - Org wide	New		Medium	4 hours ago		
31ac5fda	AnonyAS34T@contoso.com	Risky AI Usage	New		Low	4 hours ago		
43ac5f2b	AnonyZS34T@contoso.com	Data Theft Departing User - Org wide	New	Spotlight	Medium	4 hours ago		
93ac5f2b	AnonyAZ31A@contoso.com	Data Leaks - Advanced Detections	New		High	4 hours ago		
31nd5f2b	AnonyAS89U@contoso.com	Data Leaks - Org wide	New		Medium	4 hours ago		
31ac31bb	AnonyL548T@contoso.com	Risky AI Usage	New		Low	4 hours ago		
31ac5f2b	AnonyBS04Y@contoso.com	Risky AI Usage	New		Medium	4 hours ago		
51ac5f2b	AnonyAE64T@contoso.com	Data Leaks - Advanced Detections	New		High	4 hours ago		
61ac5f2b	AnonyAS34L@contoso.com	Data Leaks - Org wide	New	Spotlight	Medium	4 hours ago		
34dc5f2b	AnonyAS14Q@contoso.com	Data Theft Departing User - Org wide	New		Low	4 hours ago		
11ac5f2b	AnonyYO34Q@contoso.com	Risky AI Usage	New	Spotlight	Medium	4 hours ago		
81ac5f2b	AnonyAS34T@contoso.com	Data Leaks - Advanced Detections	New		High	4 hours ago		Lorem Case Name Ipsum

(31ac5f2b) Risky AI Usage

Assign Needs review **Confirm all alerts & create case** Dismiss alert

All risk factors Activity explorer User activity Forensic evidence

Filters: Show: All scored activity for this user Risk category: Any Activity type: Any Reset all

Sort by: Date occurred

User activity scatter plot 6 Months 3 Months 1 Month

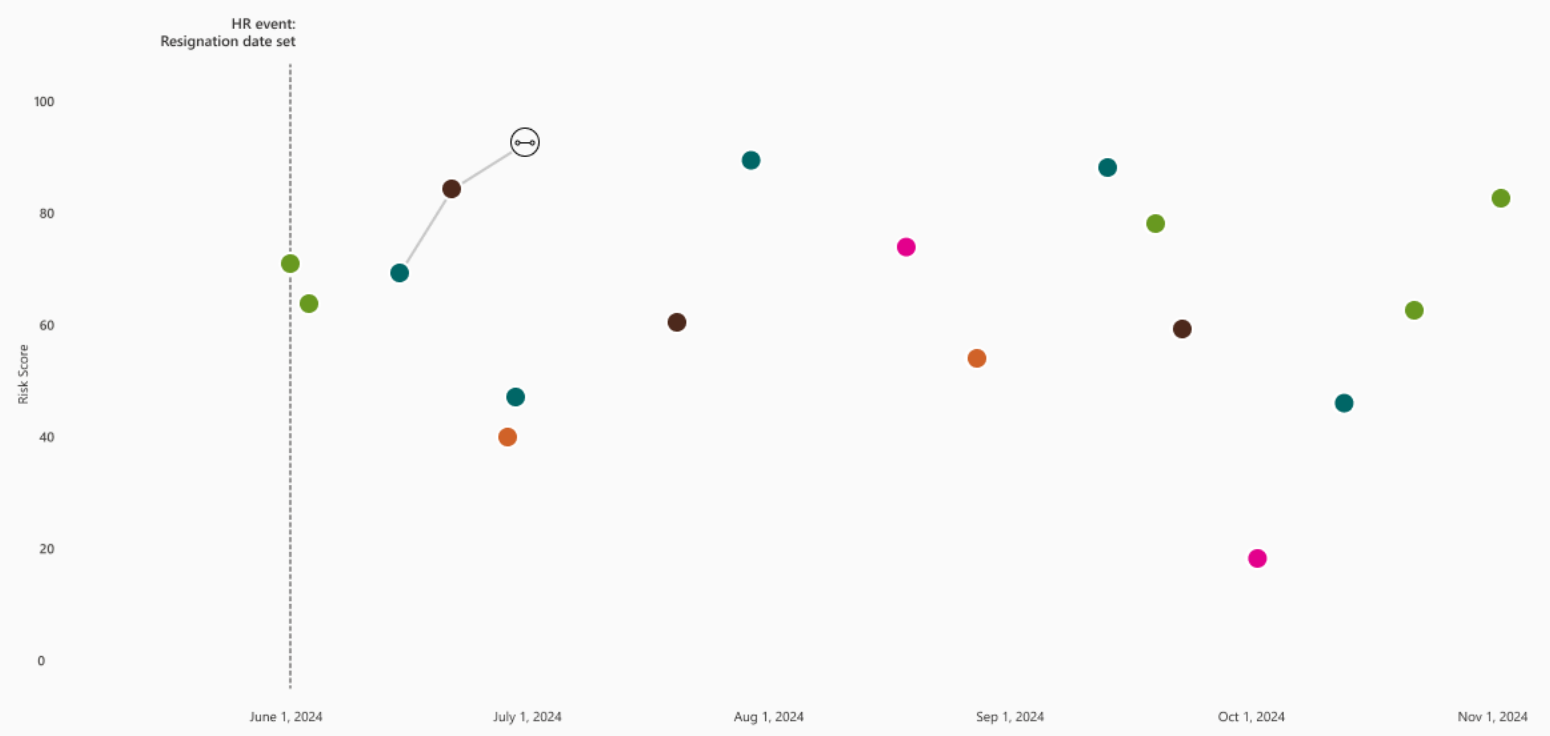
- AI Usage: Prompt attacks entered in Copilot**

June 1, 2023 (UTC) | Risk score: 55/100
1 events: Policy: Obsidian Merger
- AI usage: Sensitive response received from M365 Copilot**

June 2, 2024 (UTC) | Risk score: 85/100
10 events: Sensitive response received from Copilot
10 events: Response from sensitive site, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
10 events: Responses that have labels applied including: General
5 events: Accessed files containing trainable classifiers including: M&A Files
- (2) SEQUENCE: Label on file downgraded or removed then file accessed by Copilot**

June 16 - June 20, 2024 (UTC) | Risk score: 95/100
2 events: Sequence: File label downgraded, then file accessed by Copilot
2 events: Files that have labels applied, including: Highly Confidential
2 events: Files containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
- Exfiltration: File copied to removable media**

June 18, 2024 (UTC) | Risk score: 75/100
5 events: File copied to removable media.
1 events: Containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
5 events: file containing labels applied, including: General



Access Clean up Collection Exfiltration Infiltration Obfuscation Risky AI usage Sequence Cumulative exfiltration

(31ac5f2b) Risky AI Usage

Assign Needs review **Confirm all alerts & create case** Dismiss alert

All risk factors Activity explorer User activity Forensic evidence

Filters: Show: All scored activity for this user Risk category: Any Activity type: Any Reset all

Sort by: Date occurred

User activity scatter plot 6 Months 3 Months 1 Month

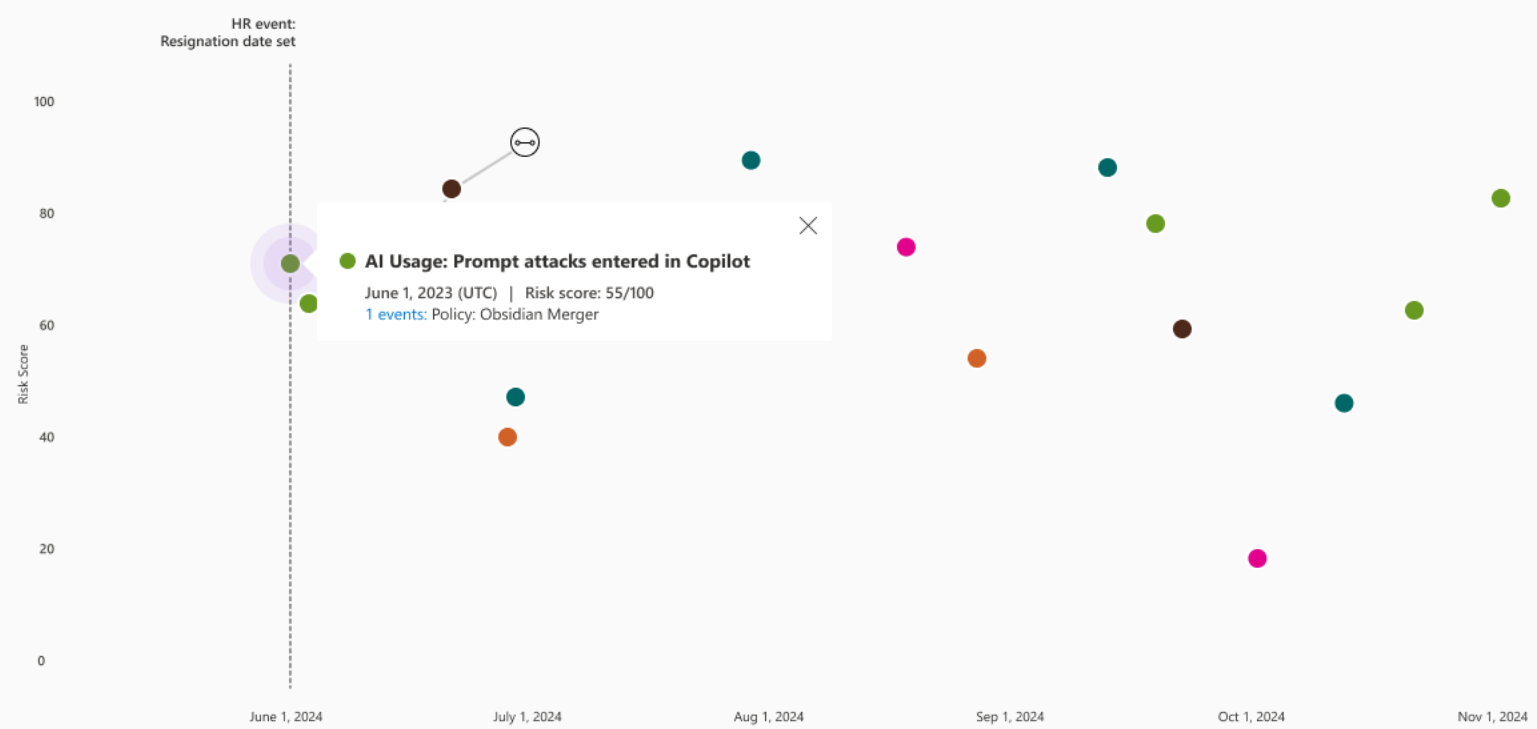
- AI Usage: Prompt attacks entered in Copilot**

June 1, 2023 (UTC) | Risk score: 55/100
1 events: Policy: Obsidian Merger
- AI usage: Sensitive response received from M365 Copilot**

June 2, 2024 (UTC) | Risk score: 85/100
10 events: Sensitive response received from Copilot
10 events: Response from sensitive site, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
10 events: Responses that have labels applied including: General
5 events: Accessed files containing trainable classifiers including: M&A Files
- (2) SEQUENCE: Label on file downgraded or removed then file accessed by Copilot**

June 16 - June 20, 2024 (UTC) | Risk score: 95/100
2 events: Sequence: File label downgraded, then file accessed by Copilot
2 events: Files that have labels applied, including: Highly Confidential
2 events: Files containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
- Exfiltration: File copied to removable media**

June 18, 2024 (UTC) | Risk score: 75/100
5 events: File copied to removable media.
1 events: Containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
5 events: file containing labels applied, including: General



User activity Forensic evidence

Activity for this user Risk category: Any Activity type: Any Reset all

User activity scatter plot

6 Months 3 Months 1 Month

Entered in Copilot

5/100

Received from M365

5/100

Received from Copilot
Info, including: Obsidian, Credit
number (SSN)
Policies applied including: General
Trainable classifiers including:

Downgraded or

by Copilot

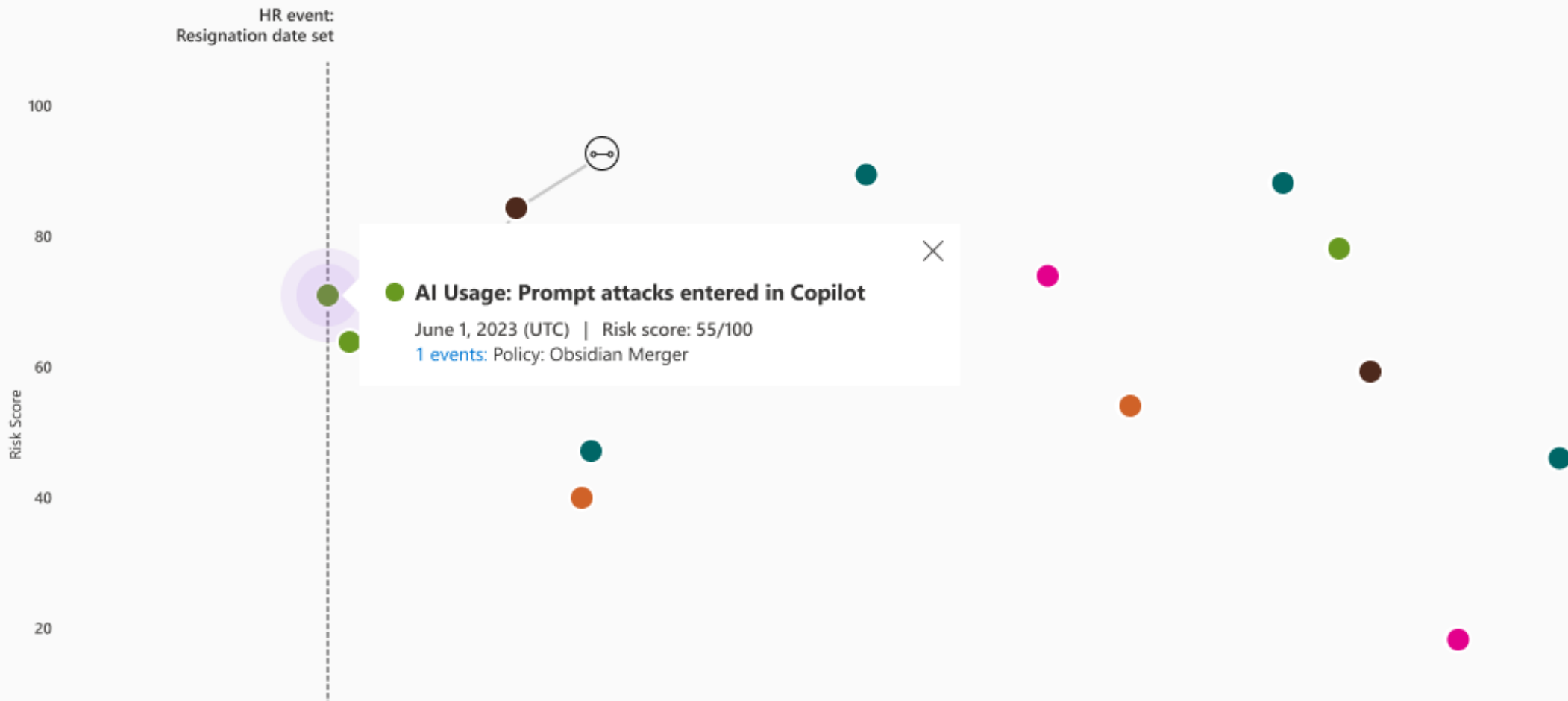
Risk score: 95/100
Downgraded, then file accessed by
Info, including: Highly

Info, including: Obsidian, Credit
number (SSN)

Removable media

75/100

media.



(31ac5f2b) Risky AI Usage

Assign Needs review **Confirm all alerts & create case** Dismiss alert

All risk factors Activity explorer User activity Forensic evidence

Filters: Show: All scored activity for this user Risk category: Any Activity type: Any Reset all

Sort by: Date occurred

User activity scatter plot 6 Months 3 Months 1 Month

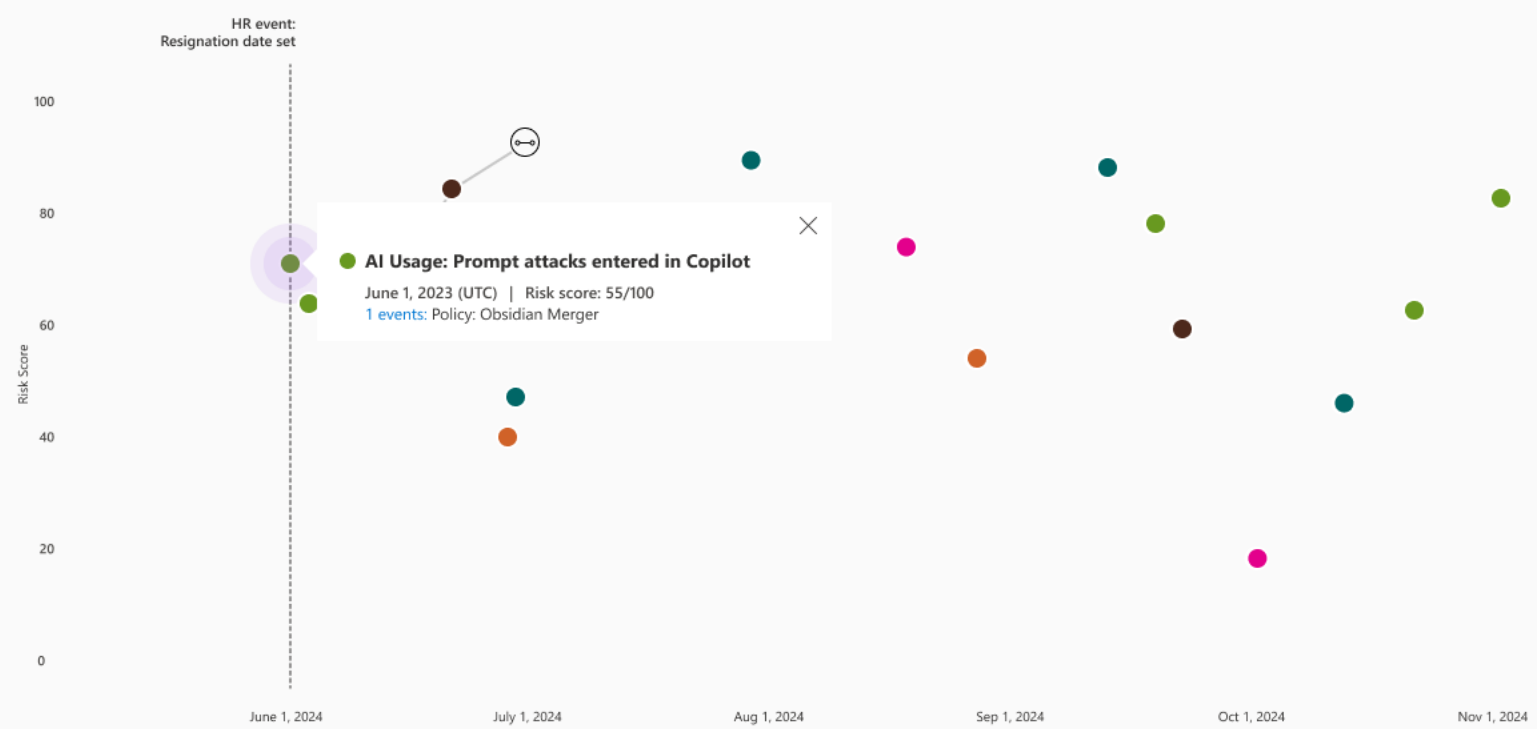
- AI Usage: Prompt attacks entered in Copilot**

June 1, 2023 (UTC) | Risk score: 55/100
1 events: Policy: Obsidian Merger
- AI usage: Sensitive response received from M365 Copilot**

June 2, 2024 (UTC) | Risk score: 85/100
10 events: Sensitive response received from Copilot
10 events: Response from sensitive site, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
10 events: Responses that have labels applied including: General
5 events: Accessed files containing trainable classifiers including: M&A Files
- (2) SEQUENCE: Label on file downgraded or removed then file accessed by Copilot**

June 16 - June 20, 2024 (UTC) | Risk score: 95/100
2 events: Sequence: File label downgraded, then file accessed by Copilot
2 events: Files that have labels applied, including: Highly Confidential
2 events: Files containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
- Exfiltration: File copied to removable media**

June 18, 2024 (UTC) | Risk score: 75/100
5 events: File copied to removable media.
1 events: Containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
5 events: file containing labels applied, including: General



Access Clean up Collection Exfiltration Infiltration Obfuscation Risky AI usage Sequence Cumulative exfiltration

(31ac5f2b) Risky AI Usage

Assign Needs review **Confirm all alerts & create case** Dismiss alert

All risk factors Activity explorer User activity Forensic evidence

Filters: Show: All scored activity for this user Risk category: Any Activity type: Any Reset all

Sort by: Date occurred

- AI Usage: Prompt attacks entered in Copilot**

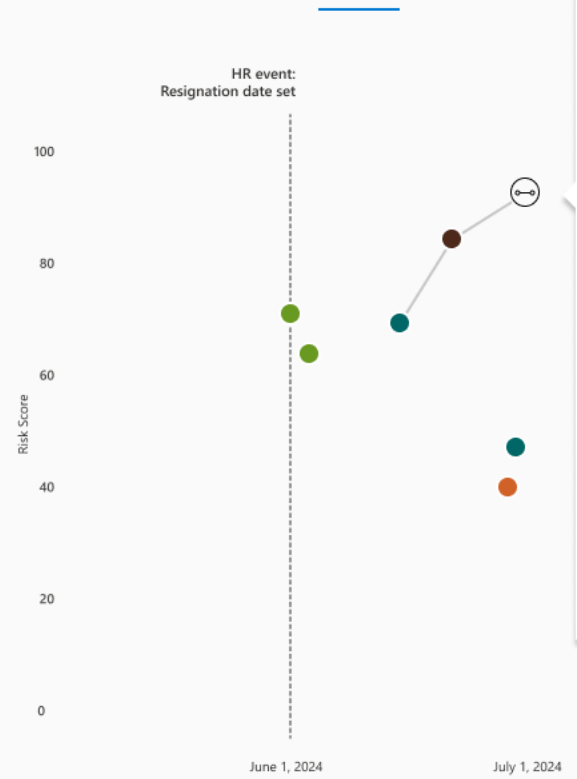
June 1, 2023 (UTC) | Risk score: 55/100
1 events: Policy: Obsidian Merger
- AI usage: Sensitive response received from M365 Copilot**

June 2, 2024 (UTC) | Risk score: 85/100
10 events: Sensitive response received from Copilot
10 events: Response from sensitive site, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
10 events: Responses that have labels applied including: General
5 events: Accessed files containing trainable classifiers including: M&A Files
- (2) SEQUENCE: Label on file downgraded or removed then file accessed by Copilot**

June 16 - June 20, 2024 (UTC) | Risk score: 95/100
2 events: Sequence: File label downgraded, then file accessed by Copilot
2 events: Files that have labels applied, including: Highly Confidential
2 events: Files containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
- Exfiltration: File copied to removable media**

June 18, 2024 (UTC) | Risk score: 75/100
5 events: File copied to removable media.
1 events: Containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
5 events: file containing labels applied, including: General

User activity scatter plot



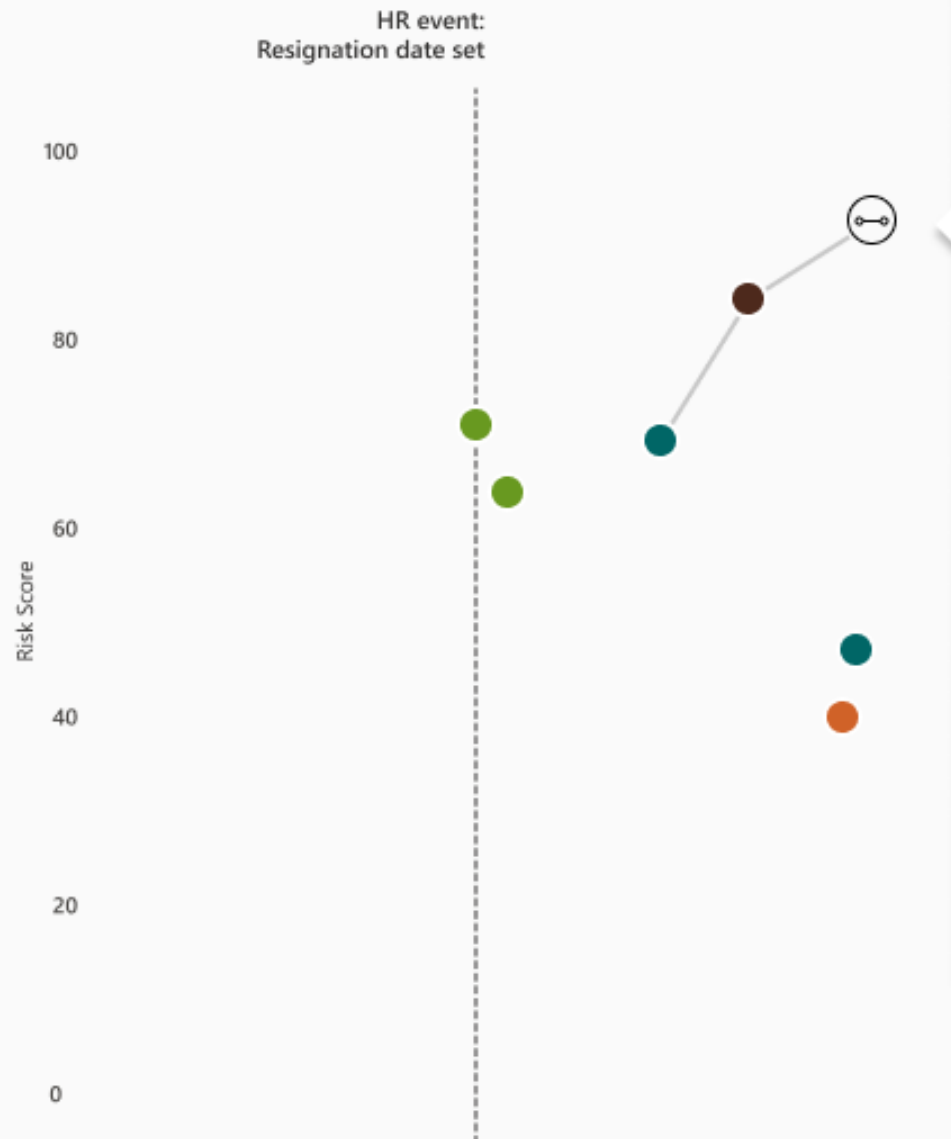
(4) SEQUENCE: Label on file downgraded or removed then file accessed by Copilot
 June 16 - June 20, 2024 (UTC) | Risk score: 95/100
 2 events: Sequence: File label downgraded, then file accessed by Copilot
 2 events: Files that have labels applied, including: Highly Confidential
 2 events: Files containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)

AI usage: Sensitive response received from M365 Copilot
 June 18, 2024 (UTC) | Risk score: 75/100
 10 events: Sensitive response received from Copilot
 10 events: Response from sensitive site, including: Obsidian
 10 events: Responses that have labels applied including: General
 5 events: Accessed files containing trainable classifiers including: M&A Files

Obfuscation: Labels of sensitive files removed on SharePoint
 June 18, 2024 (UTC) | Risk score: 55/100
 1 events: 1 Labels of SharePoint files labels removed
 1 events: file that have labels applied including: Highly Confidential
 1 events: Containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)

User activity scatter plot

6 Months 3 Months 1 M



✓ (4) SEQUENCE: Label on file downgraded or removed then file accessed by Copilot

June 16 - June 20, 2024 (UTC) | Risk score: 95/100

2 events: Sequence: File label downgraded, then file accessed by Copilot

2 events: Files that have labels applied, including: Highly Confidential

2 events: Files containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)

● AI usage: Sensitive response received from M365 Copilot

June 18, 2024 (UTC) | Risk score: 75/100

10 events: Sensitive response received from Copilot

10 events: Response from sensitive site, including: Obsidian

10 events: Responses that have labels applied including: General

5 events: Accessed files containing trainable classifiers including: M&A Files

● Obfuscation: Labels of sensitive files removed on SharePoint

June 18, 2024 (UTC) | Risk score: 55/100

1 events: 1 Labels of SharePoint files labels removed

1 events: file that have labels applied including: Highly Confidential

1 events: Containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)

(31ac5f2b) Risky AI Usage

Assign Needs review **Confirm all alerts & create case** Dismiss alert

All risk factors Activity explorer User activity Forensic evidence

Filters: Show: All scored activity for this user Risk category: Any Activity type: Any Reset all

Sort by: Date occurred

- AI Usage: Prompt attacks entered in Copilot**

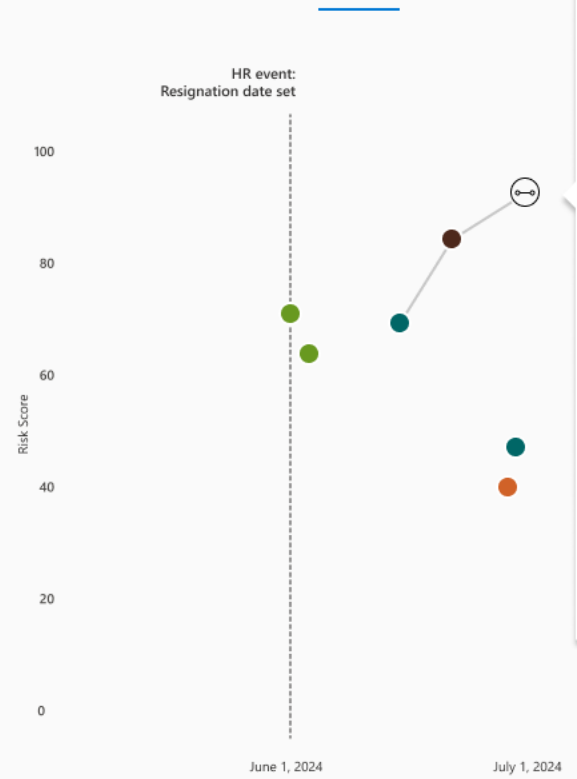
June 1, 2023 (UTC) | Risk score: 55/100
1 events: Policy: Obsidian Merger
- AI usage: Sensitive response received from M365 Copilot**

June 2, 2024 (UTC) | Risk score: 85/100
10 events: Sensitive response received from Copilot
10 events: Response from sensitive site, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
10 events: Responses that have labels applied including: General
5 events: Accessed files containing trainable classifiers including: M&A Files
- (2) SEQUENCE: Label on file downgraded or removed then file accessed by Copilot**

June 16 - June 20, 2024 (UTC) | Risk score: 95/100
2 events: Sequence: File label downgraded, then file accessed by Copilot
2 events: Files that have labels applied, including: Highly Confidential
2 events: Files containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
- Exfiltration: File copied to removable media**

June 18, 2024 (UTC) | Risk score: 75/100
5 events: File copied to removable media.
1 events: Containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)
5 events: file containing labels applied, including: General

User activity scatter plot



(4) SEQUENCE: Label on file downgraded or removed then file accessed by Copilot
 June 16 - June 20, 2024 (UTC) | Risk score: 95/100
 2 events: Sequence: File label downgraded, then file accessed by Copilot
 2 events: Files that have labels applied, including: Highly Confidential
 2 events: Files containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)

AI usage: Sensitive response received from M365 Copilot
 June 18, 2024 (UTC) | Risk score: 75/100
 10 events: Sensitive response received from Copilot
 10 events: Response from sensitive site, including: Obsidian
 10 events: Responses that have labels applied including: General
 5 events: Accessed files containing trainable classifiers including: M&A Files

Obfuscation: Labels of sensitive files removed on SharePoint
 June 18, 2024 (UTC) | Risk score: 55/100
 1 events: 1 Labels of SharePoint files labels removed
 1 events: file that have labels applied including: Highly Confidential
 1 events: Containing sensitive info, including: Obsidian, Credit Card Number, U.S Social Security Number (SSN)

Actions you can take to secure and govern AI apps

Discover

AI risks associated with data, access, and AI apps

Protect

AI and sensitive data with end-to-end security

Govern

AI usage to comply with regulatory and code-of-conduct policies

Govern AI usage to comply with regulatory and code-of-conduct policies

Assess, implement, and strength compliance against AI regulations and standards – EU AI Act, NIST AI RMF, ISO 42001, and more



Detect regulatory or code of conduct violations in AI – Violence, hate, self-harm, sexual, unauthorized disclosure, regulatory collusion, gift & entertainment, etc.



Manage data lifecycle of AI interactions and investigate potential incidents – Configurable retention policy for AI interactions, activities logs, eDiscovery



John Doe project obsidian | 🔍

Compliance Manager



- Home
- Solutions
- Learn
- Settings
- Compliance Manager
- eDiscovery
- Data Lifecycle Managem...
- Insider Risk Managem...
- AI Hub (preview)
- Compliance Manager
 - Overview
 - Improvement actions
 - Solutions
 - Assessments
 - Regulations
 - Policies
 - Alerts
- Related solutions
 - Data Lifecycle Management
 - Data Loss Prevention

Regulations

Review the list of regulations available to your organization. you can create assessments for specific regulations to track your compliance against them. [Learn more about regulations](#)

Free regulation licenses used **1/3** | Purchased regulation licenses used **0/0**

[View details](#)

4 items Group

Filter

Service: **Any** | Role type: **Any** | Overarching regulation: **Any** | Created by: **Any** | Activation: **Any** | Availability: **Premium AI**

Regulation	Availability	Created by	Last updated date	Overarching regulation	Created date	Activation	Services
Premium AI templates (4)							
EU Artificial Intelligence Act	Premium AI	Microsoft	4/17/2024	EU Artificial Intelligence Act	4/17/2024	Inactive	Universal, Microsoft 365
ISO/IEC 23894:2023	Premium AI	Microsoft	4/25/2024	ISO/IEC 23894:2023	4/25/2024	Inactive	Microsoft 365, Universal
ISO/IEC 42001:2023	Premium AI	Microsoft	3/29/2024	ISO/IEC 42001:2023	3/29/2024	Inactive	Microsoft 365, Universal
NIST AI Risk Management Framework (RMF...)	Premium AI	Microsoft	3/27/2024	NIST AI Risk Management Fra...	3/27/2024	Inactive	Microsoft 365, Universal

eDiscovery Premium



- Home
- Solutions
- Learn
- Settings
- eDiscovery
- Insider Risk Management
- Information Protection
- Audit
- AI Hub (preview)

eDiscovery (Premium) > Cases > Copilot data spillage > 11.09.2023 Alex

Saved filter queries

Filters Undo filter query Redo filter query

AND

Select a filter

+ Add filter Add subgroup

1 of 41 selected

#	Subject/Title	Status	Tag Status	Date (UTC)	Sender/Author
1	Microsoft 365 Chat...	Ready	No Tag		
2	Create an FAQ bas...	Ready	No Tag	Nov 8, 2023 9:39 PM	Alex Wilber <Alex...
<input checked="" type="checkbox"/> 3	what are the latest ...	Ready	No Tag	Nov 7, 2023 12:11 ...	Alex Wilber <Alex...
4	What are the impo...	Ready	No Tag	Nov 9, 2023 11:59 ...	Alex Wilber <Alex...
5	Summarize Ignite T...	Ready	No Tag	Nov 8, 2023 7:14 PM	Microsoft 365 Chat...
6	Microsoft 365 Chat...	Ready	No Tag		
7	summarize Ignite T...	Ready	No Tag	Nov 8, 2023 3:51 PM	Alex Wilber <Alex...
8	What are the impo...	Ready	No Tag	Nov 8, 2023 5:22 PM	Alex Wilber <Alex...

what are the latest files on P

Source Plain text Annotate Metadata

Show pinned metadata

Alex Wilber <AlexW@MODERNCOMMS382604.OnMicrosoft.com> 11/7/2023 8:11 PM

what are the latest files on Project Obsidian? **Prompt**

Microsoft 365 Chat 11/7/2023 8:11 PM

Here are the latest files on Project Obsidian:

Tag Group by families (2) Group by conversations (2)

- Home
- Solutions
- Learn
- Settings
- eDiscovery
- Insider Risk Management
- Information Protection
- Audit
- AI Hub (preview)

eDiscovery (Premium) > Cases > Copilot data spillage > 11.09.2023 Alex

Saved filter queries

Filters Undo filter query Redo filter query

AND

Select a filter

+ Add filter Add subgroup

1 of 41 selected

#	Subject/Title	Status	Tag Status	Date (UTC)	Sender/Author
1	Microsoft 365 Chat...	Ready	No Tag		
2	Create an FAQ bas...	Ready	No Tag	Nov 8, 2023 9:39 PM	Alex Wilber <Alex...
<input checked="" type="checkbox"/> 3	what are the latest ...	Ready	No Tag	Nov 7, 2023 12:11 ...	Alex Wilber <Alex...
4	What are the impo...	Ready	No Tag	Nov 9, 2023 11:59 ...	Alex Wilber <Alex...
5	Summarize Ignite T...	Ready	No Tag	Nov 8, 2023 7:14 PM	Microsoft 365 Chat...
6	Microsoft 365 Chat...	Ready	No Tag		
7	summarize Ignite T...	Ready	No Tag	Nov 8, 2023 3:51 PM	Alex Wilber <Alex...
8	What are the impo...	Ready	No Tag	Nov 8, 2023 5:22 PM	Alex Wilber <Alex...

what are the latest files on P

Source Plain text Annotate Metadata

Show pinned metadata

Response

- "kickoff.pptx" by **MOD Administrator** was last modified yesterday afternoon. It is a presentation about the project kickoff for Project Obsidian, outlining the project timeline, budget, and fund allocation. [1](#)
- "Design update.docx" by **Alex Fromm** and **MOD Administrator** was last modified yesterday evening. It is a document about the updated engine chip design for the automated car team, discussing the need to redesign the AI500 chip to improve throughput and reduce overheating. [2](#)
- "Next generation chip.docx" by **MOD Administrator** was last modified yesterday evening. It is a document discussing the need for dedicated chips to power future cars, and compares the advantages and disadvantages of ARM and x86 chip architectures. [3](#)

Tag Group by families (2) Group by conversations (2)



Microsoft Partner Security Day

Praha, 11. 2. 2025

