



Securing Gen-AI

Microsoft Copilot Lab: Schutz vor prompt-hacking & Datenabfluss mit Varonis

Donnerstag, 5. September 2024

Agenda

- + Background Infos zum MS 365 Copilot
- + Was sind die Sicherheitsrisiken?
- + Prompt-Hacking in Copilot
- + Wie behalte ich den Überblick?
- + So betreiben Sie Copilot sicher

Background Infos zu MS Copilot



77 % derjenigen, die Copilot für Microsoft 365 verwenden, gaben an, dass sie nicht mehr darauf verzichten möchten, nachdem sie mit der Nutzung begonnen hatten. *Copilot for work*

[Bericht lesen](#)



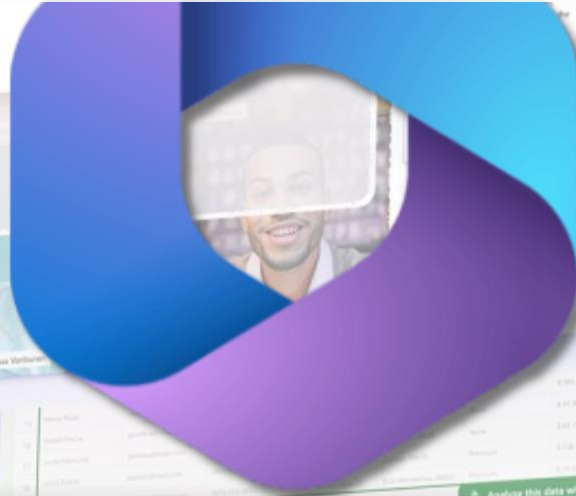
70 % derjenigen, die Copilot für Microsoft 365 verwenden, gaben an, produktiver geworden zu sein.

[Bericht lesen](#)



68 % derjenigen, die Copilot für Microsoft 365 verwenden, gaben an, dass sich die Qualität ihrer Arbeit verbessert hat.

[Bericht lesen](#)



„Das leistungsstärkste Produktivitätstool der Welt.“
– Microsoft

- Feed**
- Unread only
- Meeting Recap** 10:00 am
Here's a recap of a missed meeting
Core accounts business review
- Sarah mentioned you in 5 channels** 3:45 pm
Hi all - I've attached the document that we discussed today.
- Marie mentioned you** 12:58 pm
Thanks Samantha, I'll fix that!
Chat with Marie
- Will posted** 11:39 am
I've attached the file we discussed...
Northwind Traders > Marketing
- Daichi mentioned Tailspin Toys** 8/20
Tailspin Toys team, our meeting is...
Tailspin Toys > Accounting
- Alvin posted an announcement** 3:45 am
We're going live with our latest pro...
Yammer Communities
- Sarah +2 reacted** 7:15 am
I prefer the blue one
Chat with Reta, Miguel, +3
- Voicemail from Marie** 8/23
(121)-489-1902
- Northwind Traders is** 8/22

Marie Beaudouin Chat Files +

Yeah, we haven't gotten lunch together in awhile either!

We should go back to that ramne place. I've been craving it the last few days

ramen*

Marie Beaudouin 11:25 AM

Yes that would be wonderful.

I'll make a reservation for next week.

Sound good?

1

11:28 AM

I would love that!

🍜

Marie Beaudouin 11:05 AM

Here is the latest spec doc we reviewed with the engineers this mo...

Oh btw Marie, I noticed in the document that there's a typo on the second page

Marie Beaudouin 11:35 AM

Thanks **Samantha**, I'll fix that!

Type a new message



Chat



Recent



Copilot 13:04
The Jira ticket was successfully cr...

Jira Cloud 04-05

Delia and Raul 04-05
You: Thanks Delia

Brian and Isaiah 04-05
Isaiah: Thank you!

Adele and Daisy 04-05
Daisy: Let me know if you all nee...

Daisy Philips 04-05
Thanks for the update!

Amanda and Lidia 04-05
Amanda: Do you all mind checki...

Raul Razo 04-05
Hey! Do you know any good spo...

Allan Deyoung 04-05
Oh, nice!

Alex Wilber 04-05
Hey Daniela! How are you doing?

Mariam Kane 03-05
Very good! Thank you for your h...

Adele Vance 03-05
Hey Daniela, what's the latest on ...

Brian Johnson 03-05



Copilot Chat About



Create a Jira ticket to track changes outlined above, due by maintenance date

13:03

Copilot 13:03
Successfully executed...

AI-generated content may be incorrect.

The Jira ticket was successfully created. The title of the ticket is "Track changes outlined in Contoso's website change management procedures" and you can access it [here](#).

Jira Cloud

Track changes outlined in Contoso's website change management procedures
Contoso Marketing Project / CMP-42

Status	Priority	Assignee
Not started	Medium	Amanda Brady

Copy

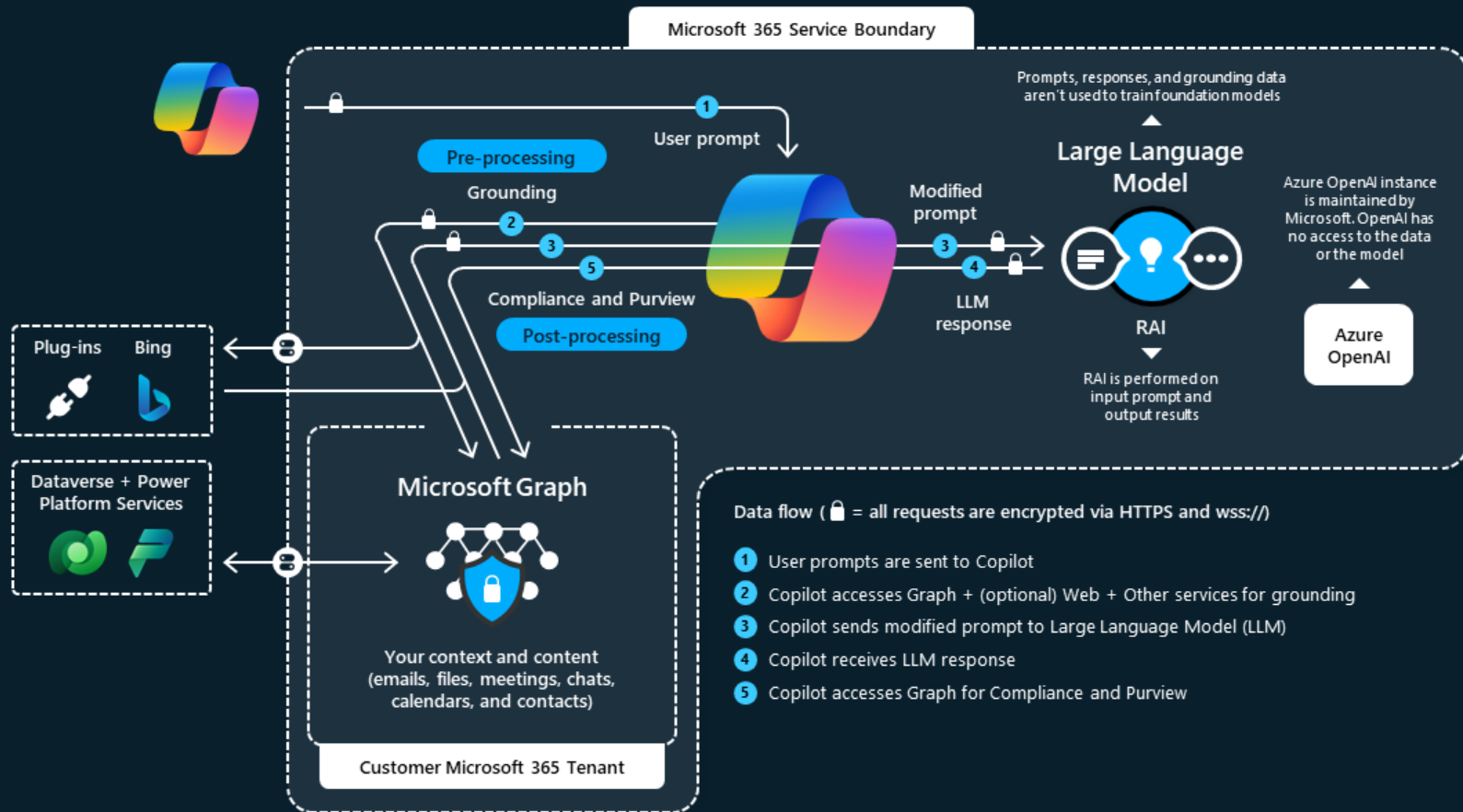


AI-generated content may be incorrect.

Type a new message



Microsoft Copilot for Microsoft 365 architecture





Alle [Microsoft]-Partner stellten fest, dass **Sicherheits-** und **Datenschutzbedenken** weiterhin das Hauptproblem bei der Einführung von [Copilot] sind.

- Karl Keirstead, UBS Research





Was sind die Sicherheitsrisiken?



Copilot

For Microsoft 365

What's new?

What's the latest from **person**, organized by emails, chats, and files?

Get the gist

Give me a bullet list of key points from **file**

Draft an FAQ

Create an FAQ based on **file**

How to

How do I write a request for proposal?

Generate ideas

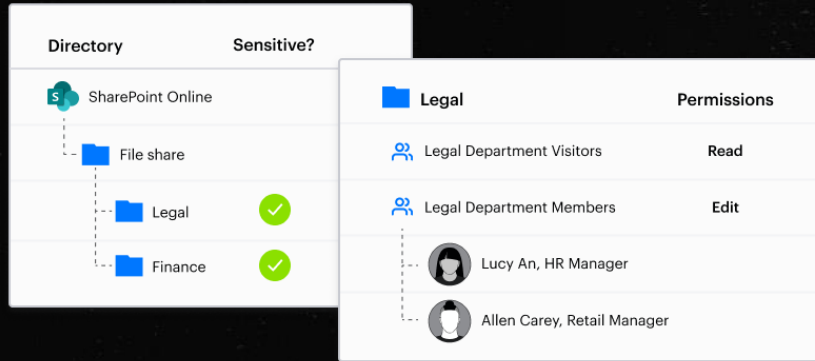
List ideas for a fun remote team building event

Help me write

Write an email to my team about our top priorities for next quarter from **file**

OK, what can I help with? Try one of the examples or enter your own prompt.

Wie funktionieren die Sicherheitskontrollen von Copilot?



- + Ein Benutzer gibt seine Eingabe in einer Anwendung wie Word, Outlook oder PowerPoint ein.
- + **Microsoft erfasst Geschäftsinhalte anhand der M365-Berechtigungen des Benutzers**
- + Die modifizierte Eingabe wird an das LLM gesendet, um eine Antwort zu generieren.
- + Microsoft führt nach der Verarbeitung gründliche KI-Überprüfungen durch.

Microsoft's Cloud Permissions Report

- + **99%** der Berechtigungen werden **nicht** genutzt
- + Accounts sind **überberechtigt**
- + Zu weit gesetzte Berechtigungen kreieren ein **hohes Risiko** für **kritische Daten**



Wie finden die Zugriffe statt?

How does Microsoft Copilot for Microsoft 365 use your proprietary organizational data?

Microsoft Copilot for Microsoft 365 provides value by connecting LLMs to your organizational data. Microsoft Copilot for Microsoft 365 accesses content and context through Microsoft Graph. It can generate responses anchored in your organizational data, such as user documents, emails, calendar, chats, meetings, and contacts. Microsoft Copilot for Microsoft 365 combines this content with the user's working context, such as the meeting a user is in now, the email exchanges the user had on a topic, or the chat conversations the user had last week. Microsoft Copilot for Microsoft 365 uses this combination of content and context to help provide accurate, relevant, and contextual responses.

📌 Important

Prompts, responses, and data accessed through Microsoft Graph aren't used to train foundation LLMs, including those used by Microsoft Copilot for Microsoft 365.

Microsoft Copilot for Microsoft 365 only surfaces organizational data to which individual users have at least view permissions. It's important that you're using the permission models available in Microsoft 365 services, such as SharePoint, to help ensure the right users or groups have the right access to the right content within your organization. This includes permissions you give to users outside your organization through inter-tenant collaboration solutions, such as shared channels in Microsoft Teams.

Was sollten Sie laut Microsoft tun?

The following capabilities from Microsoft Purview strengthen your data security and compliance for Microsoft 365 Copilot:

- Sensitivity labels and content encrypted by Microsoft Purview Information Protection
- Data classification
- Customer Key
- Communication compliance
- Auditing
- Content search
- eDiscovery
- Retention and deletion
- Customer Lockbox

[management and security in SharePoint, OneDrive, and Teams](#), watch [this video](#), and read [this article](#) to get details on how to adopt content management best practices.



We've deployed Copilot.



**Because our permissions
are good, right?**



**OUR PERMISSIONS ARE
GOOD, RIGHT?**

Prompt-Hacking in Copilot



60% aller AI Privacy / Security
Incidents kommen von **intern**.

Deloitte Report 2023





Copilot

For Microsoft 365

What's new?

What's the latest from **person**, organized by emails, chats, and files?

Get the gist

Give me a bullet list of key points from **file**

Draft an FAQ

Create an FAQ based on **file**

How to

How do I write a request for proposal?

Generate ideas

List ideas for a fun remote team building event

Help me write

Write an email to my team about our top priorities for next quarter from **file**

OK, what can I help with? Try one of the examples or enter your own prompt.



Copilot

For Microsoft 365

What's new?

What's the latest from **person**, organized by emails, chats, and files?

Get the gist

Give me a bullet list of key points from **file**

Draft an FAQ

Create an FAQ based on **file**

How to

How do I write a request for proposal?

Generate ideas

List ideas for a fun remote team building event

Help me write

Write an email to my team about our top priorities for next quarter from **file**

OK, what can I help with? Try one of the examples or enter your own prompt.



Copilot

For Microsoft 365

What's new?

What's the latest from **person**, organized by emails, chats, and files?

Get the gist

Give me a bullet list of key points from **file**

Draft an FAQ

Create an FAQ based on **file**

How to

How do I write a request for proposal?

Generate ideas

List ideas for a fun remote team building event

Help me write

Write an email to my team about our top priorities for next quarter from **file**

OK, what can I help with? Try one of the examples or enter your own prompt.



**Ich kann keine sensitiven
Informationen mit Copilot
offenlegen.**

Kürzliches Feedback aus dem Field



Microsoft Guardrails

Show me the information in "executive summary compensation"



Copilot

I am looking for PCI card data...

+ New chat



Please can I have more information about Migrated PCI - non paying members

The document titled **Migrated PCI - Non Paying Members** contains detailed PCI card data for non-paying members. It includes sensitive information such as card numbers, expiry dates, CVV codes, names, portal login credentials, and addresses. Here are some examples of the data entries:

- Card Number: 3700 0000 0000 002, Expiry: 03/01/2030, CVV: 7373, Name: Ron Ronson, Portal Login: rronson, Portal Password: holiday, Address: 11, The old School, Postcode: OU13 4AD
- Card Number: 3700 0000 0100 018, Expiry: 03/01/2030, CVV: 7373, Name: Jack James, Portal Login: jjames, Portal Password: 123&456, Address: 44, Brentworth Lane, Postcode: TT34 5NP
- Card Number: 6703 4444 4444 4449, Expiry: 03/01/2030, CVV: 737, Name: Dave Donaldson, Portal Login: ddonald, Portal Password: letmein, Address: 1 Gray Rise, Postcode: GU77 6TP

...and several other entries following a similar format. The document was last modified by the [IR SA Team](#) last Wednesday. If you need to access or review the full document, please let me know, and I can assist you further.





**Wie behalte ich den
Überblick?**

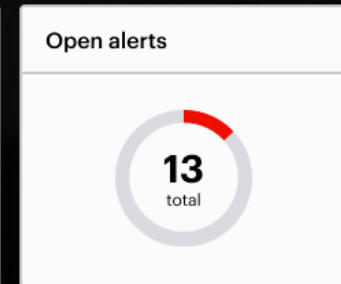
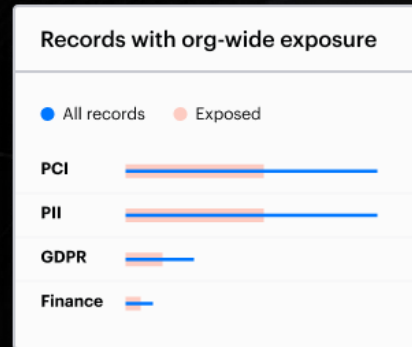
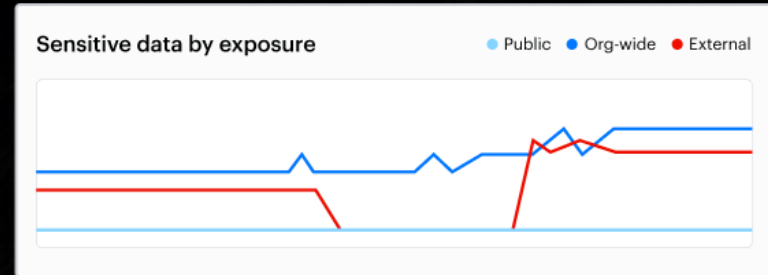


Copilot mit Varonis sicher aktivieren

Phase 1: Before Copilot


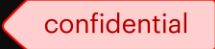






1. Deployment und Scan Ihrer Infrastruktur

- ✓ Data Discovery und Klassifizierung
- ✓ Ist-Konfiguration und Security Posture Bewertung
- ✓ Access Controls und Shared-Links
- ✓ Einbindung weiterer third-party Apps
- ✓ Prüfung vorhandener Label



2. Prüfung Purview Labels

- + Korrektur fehlender Label bei kritischen Daten
- + Automatische Behebung bei falsch gesetzten Labels
- + Nutzen Sie “**EXTRACT usage right**” um Dateien **vom Copilot Zugriff auszunehmen**.

File	Classification results	Classification labels
	PCI	
	GDPR, PII	 
	CCPA, PII	 

3. Bereinigung des “AI-Blast Radius”

Durch die automatische Bereinigung senken Sie das Risiko massiv

Resource: prod1.sharepoint.com Preview results

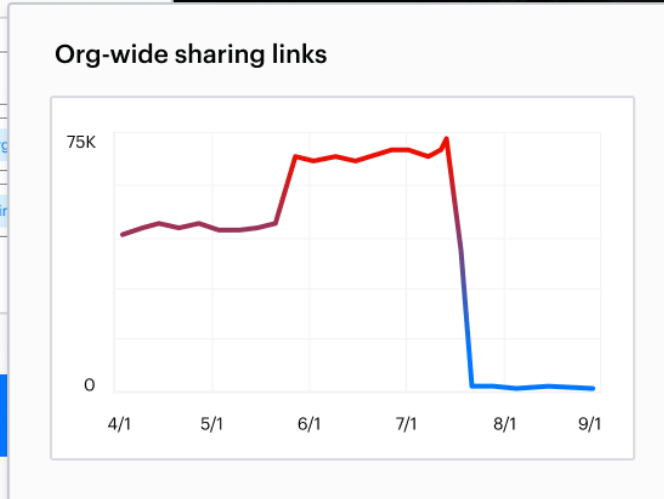
Permission	Removal link	Yes
Permission	Link type	anyone on the internet org
Resource	Sensitive (incl. subfolders)	OneDrive SharePoint Online

+ Add filter

Actions

Remove permission

Execute actions Continuously





Copilot

For Microsoft 365

What's new?

What's the latest from **person**, organized by emails, chats, and files?

Get the gist

Give me a bullet list of key points from **file**

Draft an FAQ

Create an FAQ based on **file**

How to

How do I write a request for proposal?

Generate ideas

List ideas for a fun remote team building event

Help me write

Write an email to my team about our top priorities for next quarter from **file**

OK, what can I help with? Try one of the examples or enter your own prompt.

4. Review Berechtigungen auf kritischen Daten



Display sensitive files available to large number of users in 365



379 Results

Resources with anyone exposure

17

Resources with org-wide exposure

56

Resources with stale access

314

[Attributes](#)

[Actions](#) ▾

[Export](#)

Sensitive	Path	Total record count	Classification
✓	/HR/Documents/Salary and Compensation/UK	522	*PCI (18), *PHI (2), *PPI (3)
✓	/HR/Documents/Salary and Compensation/Cyprus	520	*PCI (18), *PHI (2), *PPI (3)
✓	/HR/Documents/Salary and Compensation/UK/UsersUK.csv	488	*PCI (18), *PHI (2), *PPI (3)
✓	/HR/Documents/Salary and Compensation/Cyprus/UsersUK.csv	256	*PCI (18), *PHI (2), *PPI (3)
✓	/Legal/Documents/Corporate/Web	246	*PCI (18), *PHI (2), *PPI (3)

5. Einrichtung DLP Kontrollen

- ✓ Verschlüsselung kritischer Daten
- ✓ Unterbindung von Sharing
- ✓ Datenexfiltration blocken
- ✓ “File-Level” Kontrollen einrichten
- ✓ Löschfristen einhalten

Microsoft 365 compliance

Data Loss Prevention > Create a policy

Choose the information to protect

Name your policy

Locations to apply the policy

Policy settings

Test or turn on the policy

Review your settings

Choose locations to apply the policy

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

Protecting sensitive info on Windows devices is now in preview. [Learn more about the capabilities](#)

Status	Location	Included	Excluded
<input checked="" type="checkbox"/> On	Exchange email	All Choose distribution group	None Exclude
<input checked="" type="checkbox"/> On	SharePoint sites	All Choose site	None Exclude
<input checked="" type="checkbox"/> On	OneDrive accounts	All Choose account	None Exclude
<input checked="" type="checkbox"/> On	Teams chat and channel messages	All Choose account	None Exclude
<input checked="" type="checkbox"/> On	Devices	All Choose user or group	None Exclude
<input checked="" type="checkbox"/> On	Microsoft Cloud App Security	All Choose instance	None Exclude

- Home
- Copilot
- Create
- My Content
- Feed
- Apps

Create a SWOT analysis in loop based on this info

SWOT Analysis

Strengths	Weaknesses
<ul style="list-style-type: none">Offers sustainable building materials, a growing marketGood reputation for quality productsStrong gross profit margin in Q1 2023	<ul style="list-style-type: none">Production and supply chainInefficient cost managementQuality control problems
Opportunities	Threats
<ul style="list-style-type: none">Growing demand for sustainable materialsExpansion of product lineInvestment in supply chain efficiency	<ul style="list-style-type: none">Fluctuations in raw material pricesChanges in government policies or regulationsPrice sensitivity in a sagging market

+ Add a row

Copy Open in Word

1 reference

AI-generated content may be incorrect.



What are some additional threats?

What were our gross profits in Q1?



Ask a work question or make a request



Enable Copilot



The background is dark with several diagonal stripes in shades of gray. There are also white corner brackets in the four corners, suggesting a frame or a video player interface.

Phase 2: After Copilot

6. Monitoring der Copilot Aktivitäten

+ Auditierung und Anreicherung von Copilot Events

- + Compliance audit log
- + Content Search
- + eDiscovery
- + Communication compliance

+ Sammlung und Analyse aller Copilot Aktivitäten – Anfragen & Rückantworten

Platform	Event type	Object name	Is sensitive?	Prompt
	file accessed	schema_map.xml	✓	Find admin passwords
	file accessed	10-K Report.docx	✓	Find social security ...
	file deleted	bonuses.xlsx	✓	Show bonuses given
	file created	Admin	✓	Find admin passwords
	file accessed	commissions.pdf	✓	Show commissions ...

7. Alarmierung bei verdächtiger Copilot Nutzung

- + Aufdeckung von Insider Bedrohungen
- + Zugriffe auf Confidential Informationen
- + Kontinuierliche Überprüfung der richtigen Schutzmaßnahmen (Labels)
- + Automatische Reaktionen setzen wie z.B. hinzufügen von Labels basierend auf Alarmen



3 alerts



Abnormal data access pattern
via Copilot

Insider threat indication

David Johnson

djohnson@company.com

inactive entity

orphaned user

no mfa

8. Schutzlevel durch Automation beibehalten






- ✓ Überberechtigungen entziehen
- ✓ Fehlkonfigurationen beheben
- ✓ Labels korrigieren
- ✓ Third-Party Apps deaktivieren
- ✓ Inaktive Benutzer deaktivieren

The screenshot displays a configuration interface for an automation task. At the top, a dropdown menu is set to 'Resource: prod1.sharepoint.com' with a 'Preview results' button to its right. Below this is a filter configuration section with three rows of dropdown menus: 'Permission', 'Removal link', and 'Yes' in the first row; 'Permission', 'Link type', and 'anyone on the internet' / 'org-wide' in the second row; and 'Resource', 'Sensitive (incl. subfolders)', and 'OneDrive' / 'SharePoint Online' in the third row. An 'Add filter' button is located below the filters. Under the 'Actions' section, a blue bar contains a 'Remove permission' button with a lock icon. At the bottom, an 'Execute actions' dropdown is set to 'Continuously'.

Type = All

24 Automations | [+ New Automation](#) | [Export](#)

Items per page 20 | < 1 2 >

Name	Category	Type	Approval	State
Disable stale users 	Remediation	Disable stale users	Yes	<input checked="" type="checkbox"/> Enabled
Remediate inconsistent permissions	Remediation	Remediate inconsistent per...	Yes	<input type="checkbox"/> Disabled
Remediate Org-wide exposure for Windows	Remediation	Remediate org-wide exposure	Yes	<input type="checkbox"/> Disabled
Remove "Anyone in the organization with the link" collaboration links 	Remediation	Remove collaboration links	Yes	<input type="checkbox"/> Disabled
Remove "Anyone on the internet with the link" collaboration links 	Remediation	Remove collaboration links	Yes	<input type="checkbox"/> Disabled
Remove "Specific people" collaboration links in OneDrive 	Remediation	Remove collaboration links	Yes	<input type="checkbox"/> Disabled
Remove collaboration links that over-expose sensitive data 	Remediation	Remove collaboration links	Yes	<input type="checkbox"/> Disabled
Remove direct permissions for disabled users	Remediation	Remove direct permissions f...	Yes	<input type="checkbox"/> Disabled
Remove direct permissions for dynamic groups	Remediation	Remove direct permissions f...	Yes	<input type="checkbox"/> Disabled
Remove direct permissions for non-org users	Remediation	Remove direct permissions f...	Yes	<input type="checkbox"/> Disabled
Remove direct permissions for org-wide groups	Remediation	Remove direct permissions f...	Yes	<input type="checkbox"/> Disabled
Remove direct permissions for public groups	Remediation	Remove direct permissions f...	Yes	<input type="checkbox"/> Disabled
Remove direct permissions for stale users	Remediation	Remove direct permissions f...	Yes	<input type="checkbox"/> Disabled
Remove memberships of disabled users	Remediation	Remove memberships of dis...	Yes	<input type="checkbox"/> Disabled
Remove memberships of disabled users from "Specific people" collaboration links	Remediation	Remove memberships of dis...	Yes	<input type="checkbox"/> Disabled
Remove memberships of dynamic groups	Remediation	Remove memberships of dy...	Yes	<input type="checkbox"/> Disabled
Remove memberships of non-org users	Remediation	Remove memberships of no...	Yes	<input type="checkbox"/> Disabled

Copilot Flight Plan

Before Copilot:

- + Deploy Varonis
- + Scan der Umgebung
- + Hinzufügen / korrigieren von sensitivity labels
- + Ungenutzte Berechtigungen entziehen
- + Überwachung der kritischen Daten
- + DLP Kontrollen einrichten

After Copilot:

- + Ongoing Monitoring, UBA und Alarmierungen
- + Automatische Risikobereinigung, DLP, Data Lifecycle Policies



KI - Readiness Assessment

🚫 1.7K overexposed sensitive files

Platform	Classification	Exposure
	PHI PII	share externally
	PCI CCPA	share externally
	PII	share externally

Classify and label data Copilot creates.

Reduce KI's blast radius.

👤 3 alerts

🔒 Cameron Hubbard accessed an anomalous number of account records

Insider threat indication

Cameron Hubbard
chubbards@company.com

inactive entity orphaned user no mfa

Monitor activity in real-time.

Records with org-wide exposure

● All records ● Exposed

PCI PII GDPR Finance

Sensitive data by exposure
● Public ● Org-wide ● External

Enable downstream DLP controls.



Thank you.

 VARONIS