



# Breaks in the Cloud

Schutz vor Risiken im Zusammenhang mit LLM Modellen

# Cloud-Sicherheit ist Pflicht

- + Es geht nicht mehr um die Frage "**Sollen wir in die Cloud wechseln**", sondern "**Wie gehen wir in die Cloud,**
- + Datensicherheit bleibt das wichtigste Anliegen, wenn es um Cloud-Sicherheit geht



# KI hat die Cloud erobert

- + Large language models (LLMs)
- + Copilots
- + Threat hunting
- + Threat detection
- + Investigation
- + Remediation
- + Und vieles mehr...



Aber die Einführung von  
KI hat auch eine neue  
**Komplexitätsebene** für  
die **Datensicherheit**  
geschaffen

# Der Blast Radius wächst unaufhaltsam.

Jährliche Datenwachstumsrate **23%**

Wachstumsrate für hoch sensible Datensätze

Unternehmen mit exponierten Cloud-Daten **47%**

Unternehmen, die keine Cloud-Datenbanken und Backups haben

Ungenutzte Cloud-Berechtigungen **99%**

Ungenutzte autorisierte Cloud-Integrationen und Pipelines

# AI Data Protection Challenges



Schutz der Privatsphäre der Daten bei gleichzeitiger Beibehaltung der KI-Funktionalität

---



Neue Bedrohungen und sich entwickelnde Angriffstechniken

---



Umsetzung von Sicherheitsmaßnahmen, ohne Innovation zu bremsen

---



Einhaltung sich ändernder gesetzlicher Vorschriften

---



Unzureichende Fähigkeiten und Ressourcen für die Entwicklung und Wartung von Sicherheitsschichten



# 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.



Search...



# Seller Home

Good morning, Andrew Cook. Let's get selling!

## Close Deals

Opportunities owned by me and closing this quarter

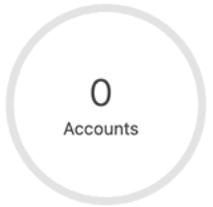


- \$0 Open
- \$0 Won
- \$0 Lost

[View Opportunities](#)

## Plan My Accounts

Accounts owned by me



- 0 Upcoming Activity
- 0 Past Activity
- 0 No Activity

[View Accounts](#)

## Build Pipeline

Leads owned by me and created in the last 30 days



- 0 Upcoming Activity
- 0 Past Activity

## My Goals

Set personal weekly or monthly goals for emails, calls, and meetings.



## Grow

Contacts

Einstein ⓘ

Let's chat!

Hi! I'm Einstein, an AI assistant. I can do things like search for information, summarize records, and draft and revise emails. What can I help you with?

Describe your task or ask a question...



# OWASP Top 10 for LLM Applications

LLM01

## Prompt Injection

This manipulates a large language model (LLM) through crafty inputs, causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.

LLM02

## Insecure Output Handling

This vulnerability occurs when an LLM output is accepted without scrutiny, exposing backend systems. Misuse may lead to severe consequences like XSS, CSRF, SSRF, privilege escalation, or remote code execution.

LLM03

## Training Data Poisoning

This occurs when LLM training data is tampered, introducing vulnerabilities or biases that compromise security, effectiveness, or ethical behavior. Sources include Common Crawl, WebText, OpenWebText, & books.

LLM04

## Model Denial of Service

Attackers cause resource-heavy operations on LLMs, leading to service degradation or high costs. The vulnerability is magnified due to the resource-intensive nature of LLMs and unpredictability of user inputs.

LLM05

## Supply Chain Vulnerabilities

LLM application lifecycle can be compromised by vulnerable components or services, leading to security attacks. Using third-party datasets, pre-trained models, and plugins can add vulnerabilities.

LLM06

## Sensitive Information Disclosure

LLMs may inadvertently reveal confidential data in its responses, leading to unauthorized data access, privacy violations, and security breaches. It's crucial to implement data sanitization and strict user policies to mitigate this.

LLM07

## Insecure Plugin Design

LLM plugins can have insecure inputs and insufficient access control. This lack of application control makes them easier to exploit and can result in consequences like remote code execution.

LLM08

## Excessive Agency

LLM-based systems may undertake actions leading to unintended consequences. The issue arises from excessive functionality, permissions, or autonomy granted to the LLM-based systems.

LLM09

## Overreliance

Systems or people overly depending on LLMs without oversight may face misinformation, miscommunication, legal issues, and security vulnerabilities due to incorrect or inappropriate content generated by LLMs.

LLM10

## Model Theft

This involves unauthorized access, copying, or exfiltration of proprietary LLM models. The impact includes economic losses, compromised competitive advantage, and potential access to sensitive information.

# Wie kann man sich vor KI- Datensicherheitsrisiken schützen?



# AI Data Security



## Sichtbarkeit

Vollständige Scans auf großen Datenspeichern.  
Keine blinden Flecken.



## Kontinuierliche Detektion

Sind sensitive Daten offengelegt, gefährdet oder angegriffen?



## Automatisierte Bereinigung

Erfahren Sie, was geändert und erstellt wurde, damit die Transparenz immer auf dem neuesten Stand ist.

The background is a solid blue color with several diagonal stripes of varying shades of blue running from the top-left towards the bottom-right. In each of the four corners, there is a white dashed-line bracket shape, resembling a corner of a square frame.

# **KI sicher aktivieren mit Varonis**

FORRESTER®

WAVE  
LEADER 2023

Data Security Platforms

“Varonis ist die **erste Wahl** für Unternehmen, die Wert auf umfassende Datentransparenz, Klassifizierungsfunktionen und automatisierte Behebung risikobehafteter Datenzugriffe legen.”

Forrester Wave™: Data Security Platforms, Q1 2023

# Vollständig automatisierte Datensicherheit



## Sichtbarkeit in Echtzeit

Verstehen Sie Ihre wahre Datensicherheitslage in Echtzeit.



## Automatisierte Prävention

Reduzieren Sie kontinuierlich Ihren Explosionsradius, wenden Sie Labels an und setzen Sie Richtlinien durch.



## Proaktive Erkennung

Always-on, datenzentrierte UEBA + das Varonis IR-Team.

# Überwachen der Copilot-Aktivität

## + Sammeln und Anreichern von Copilot-Audit-Quellen

+ Compliance-Überwachungsprotokoll

+ Content-Suche

+ eDiscovery

+ Compliance in der Kommunikation

## + Überwachung von Unterhaltungen

+ Sehen Sie sich den genauen Austausch zwischen einem Benutzer und Copilot an

The screenshot displays the Copilot audit interface. On the left, a sidebar lists various filters such as Event, Country, Data source type, Event type, Status, Acting Account, Account Names, Affiliation, Domain, Source Device, External IPs, Affected Resource, Data source, Object type, and Affected Account. The main area shows a table of events with columns for Event, Account Name, Event Type, Prompt (Copilot), and Response. The table contains 32 results, with the first few rows showing interactions by John Smith. A 'Refine' section is visible below the table, and an 'Export' button is present. On the right, a 'Copilot Conversation' panel shows a detailed view of a conversation. The conversation starts with the prompt 'What was the payment agreement with Acme corp?' and the response 'The payment agreement with Acme Corp involves paying suppliers for goods and services delivered via wire transfer. The agreement specifies monthly payment terms with amounts for each supplier, such as TUA Airlines (\$15,000), Software Ltd (\$20,000), Tom & Law (\$25,000), Suppliers global (\$30,000), STU Corp (\$35,000), and Jid software Ltd (\$45,000). Invoices are to be sent by email with details like invoice number, date, description, quantity, unit price, total amount, and wire transfer instructions. This agreement was authored by vmsadm and last modified on Sunday afternoon. Additionally, you were invited by vmsadm to edit the document titled "Payment Agreement Acme Corp".' Below this, another prompt asks 'Can you summarize the key points?' and the response lists key points: Payment Method, Payment Terms, Supplier Payments, and Invoice Requirements. A third prompt asks 'What are the wire transfer instructions?' and the response details the wire transfer instructions for Acme Corp.

Event	Account Name (Acting...)	Event Type	Prompt (Copilot)	Response
(1)	John Smith	Interacted with Copilot	What are the wire transfer instructions?	The w...
(1)	John Smith	Interacted with Copilot	Can you summarize the key points?	The key...
(1)	John Smith	Interacted with Copilot	What was the payment agreement with Acme corp?	The pay...
(1)	John Smith	Interacted with Copilot	can you look in our sharepoint	I found...
(1)	John Smith	Interacted with Copilot	What was the payment agreement with Acme corp?	The pag...
(1)	John Smith	Interacted with Copilot	What was the payment agreement with Acme corp?	The pag...
(1)	John Smith	Interacted with Copilot	Payment Agreement Acme Corp.docx(AutoGenerated...	The pag...
(1)	John Smith	Interacted with Copilot	What was the payment agreement with Acme corp?	The pag...
(1)	John Smith	Interacted with Copilot	What was the payment agreement with Acme corp?	The pag...
(1)	John Smith	Interacted with Copilot	Can you summarize the key points?	Certain...
(1)	John Smith	Interacted with Copilot	What was the payment agreement with Acme corp?	The pag...
(1)	John Smith	Interacted with Copilot	Can you summarize the key points?	The key...
(1)	John Smith	Interacted with Copilot	Can you summarize the key points?	The key...
(1)	John Smith	Interacted with Copilot	What was the payment agreement with Acme corp?	The pag...
(1)	Sarah Davis	Interacted with Copilot	What other promotions were mentioned in the docu...	The doc...
(1)	Carah Pivote	Interacted with Copilot	Tell me more about the promotion letter	The or...

# Warnung bei riskantem Copilot-Verhalten

- + Unangebracht
- + Böswillige Absicht
- + Sensitive Daten offengelegt
- + Ungewöhnlich

**Suspicious M365 Copilot activity: Attempt to access sensitive financial data related to payment inf...** High Actions

Intrusion | 08/01/2024 9:11 AM | Status: New | Alert ID: 4E94BA11...

### What happened

varonistest237\John Smith, Software Engineer, initiated a query using M365 Copilot, requesting information related to payment financial information

This threat model is designed to detect when a user uses Microsoft 365 Copilot to access and locate sensitive financial data related to payment information. By analyzing both the prompt issued by the user and the response provided by Copilot, the model identifies patterns and indications suggesting the user's intent to retrieve sensitive and potentially confidential financial information.

[View alerted events](#)

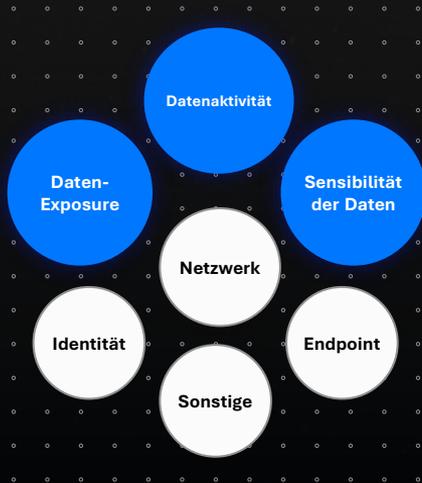
**Users (1)** **Data (1)**

**varonistest237\John Smith**  
Software Engineer

# Varonis MDDR: Stop Data Breaches 24x7x365

## Datenzentrierte Sichtbarkeit

Tiefgreifende Analysen, um Datenbedrohungen aufzudecken, die anderen Lösungen entgehen.



## AI-Powered Threat Detection

Wir verstehen, welcher Datenzugriff für jeden Benutzer und jedes Gerät normal oder abnormal ist.



KI / MASCHINELLES LERNEN

Verhaltensbasierte Bedrohungsmodelle

Alerts in Echtzeit

Automatisierte Analyse und Reaktion

## 24/7/365 MDDR

Unsere Experten überwachen und reagieren auf Warnungen

- ✓ Dedizierter Sicherheitsexperte
- ✓ Industry-best SLA
- ✓ Kommunikation rund um die Uhr
- ✓ Proaktive Bedrohungsjagd
- ✓ Incident Response Team
- ✓ Threat Intelligence
- ✓ Security Posture Assessments

# LLM Bereitschaft in 10 Tagen

Varonis reduzierte das Copilot-Exposure eines Finanzinstituts innerhalb von 10 Tagen um 99,8 %.  
Keine Auswirkungen auf das Unternehmen mit 20 Automatisierungsrichtlinien, um ~1 Million Dateien zu sperren.

**992K**  
Dateien

## Ausgangspunkt

~1 Mio. Dateien in SharePoint, Teams und OneDrive der Organisation. Varonis klassifizierte schnell 120K sensitive Dateien.

**57K**  
Offengelegte  
Dateien

## Identifizieren des Copilot- Risikos

Verwendet Varonis, um exponierte PHI, PII und Anmeldeinformationen zu identifizieren, bei denen das Risiko besteht, dass sie über Copilot offengelegt werden.

**99.8%**  
Reduzierung  
des  
Exposure

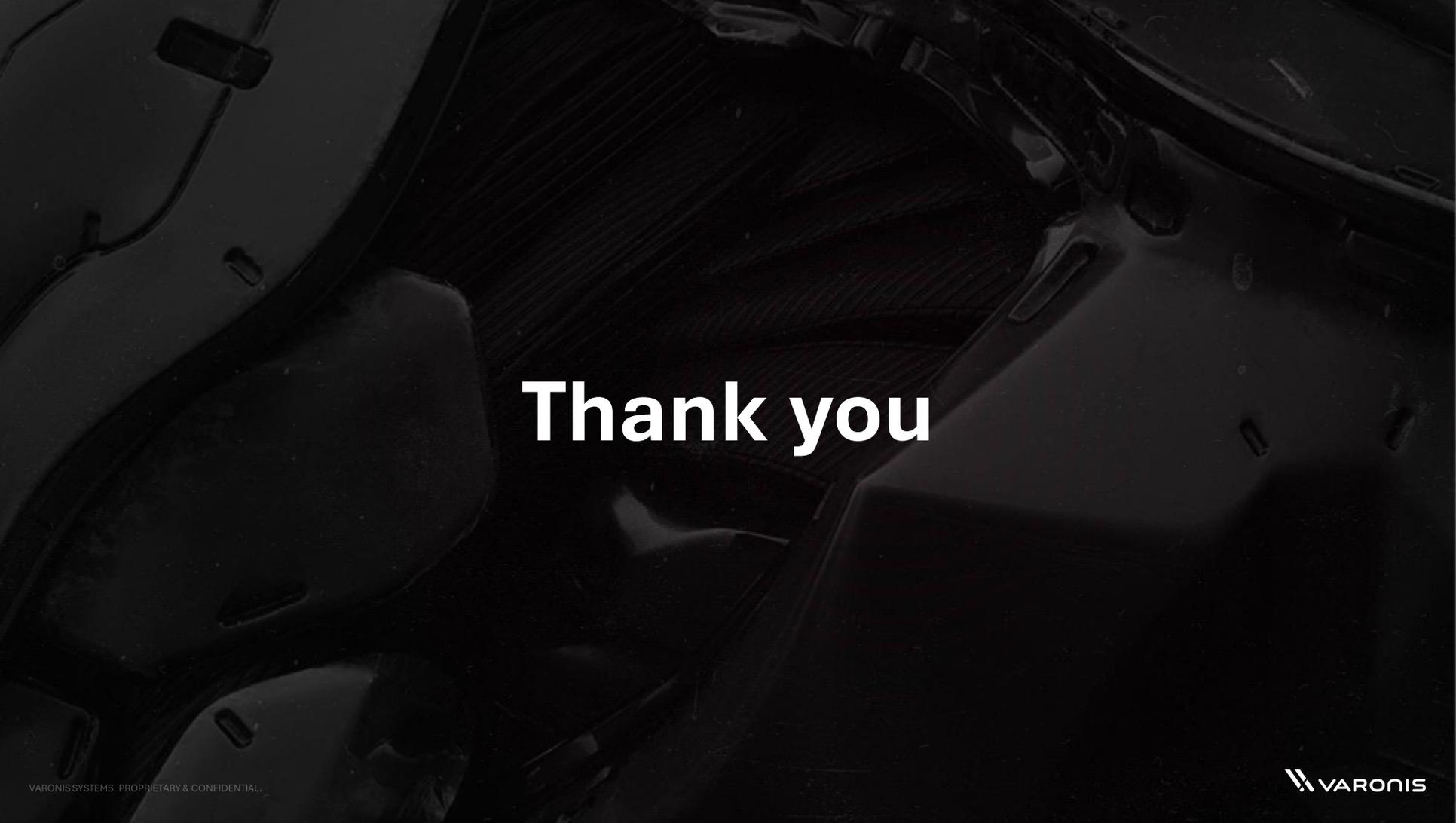
## 10-Tage Unterschied

Automatisch eliminierte ~2K Freigabelinks und Zugriff auf 57K Dateien in der richtigen Berechtigung mit Varonis-Richtlinien.

**0**  
Beschwerden  
über Tickets

## Keine Auswirkungen auf die Kunden

Benachrichtigung der M365-Administratoren proaktiv über den Behebungsplan. Das Team hörte keine Beschwerden.



**Thank you**

