VARONIS

2025

# STATE OF DATA SECURITY REPORT

## *QUANTIFYING AI'S IMPACT ON DATA RISK*

Insights from 1,000 real-world IT environments

# KEY FINDINGS

## AI adoption is outpacing security measures.

While AI can drive progress, it can accelerate risk. In our review of 1,000 organizations, we uncovered alarming data security issues:

### 90% of organizations have exposed sensitive cloud data.

Critical data that's not locked down can be surfaced by AI. Exposed AI training data is vulnerable to breaches and model poisoning.

### 88% have stale but enabled ghost users.

These accounts remain enabled, providing access to applications and data. Ghost users can allow attackers to conduct reconnaissance or exfiltrate data without tripping alarms.

### 98% have unverified apps, including unsanctioned AI.

Shadow AI increases the risk of exposure and data breaches. Attackers can use unverified apps to exfiltrate data.

### 99% of organizations have sensitive data dangerously exposed to AI tools.

# ABOUT THIS REPORT

## Methodology

Varonis analyzed data security risk across a wide range of industries and geographies:

- 1,000 organizations
- Nearly 10 billion cloud resources (objects, files, reports, attachments, etc.)
- More than 20 petabytes of data — approximately 20 terabytes per organization

IaaS and SaaS applications and services reviewed:

## Countries included:

United States, Canada, Germany, France, Belgium, Netherlands, Sweden, Switzerland, U.K., Spain, Italy, Brazil, Australia, India, and more

## Firmographics

This report includes data analysis across industries:

- Healthcare/biotech/pharma
- Finance
- Government/public sector
- Insurance and professional services
- Manufacturing
- Education
- Technology
- Consumer/retail
- And others...

# TABLE OF CONTENTS

# SHADOW AI:

## *A HIDDEN THREAT TO DATA SECURITY*

Shadow AI — unsanctioned gen AI applications — poses a significant threat to data security. These tools can bypass corporate governance and IT oversight, leading to potential data leaks.

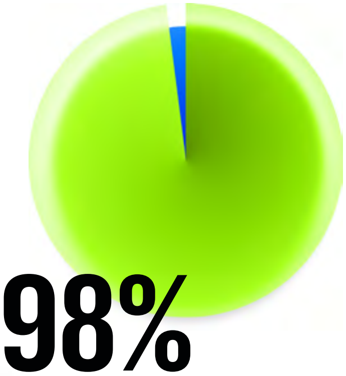
Employees can accidentally leak sensitive or confidential data using shadow AI. If these apps fail to comply with GDPR, HIPAA, and other regulations, companies could be fined.

Stale apps can remain dangerous after a user's last login. Stale OAuth applications, which haven't been used or accessed for weeks or months, still have permission to access sensitive data.

### *In 2025, millions downloaded DeepSeek.*

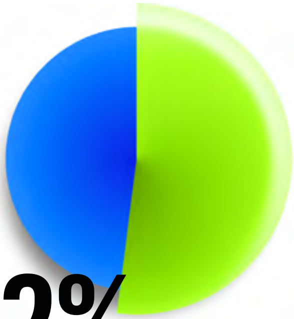
Employees who downloaded DeepSeek put their company's data at risk — an **unsecured DeepSeek database** exposed a million lines of log streams containing chat history, secret keys, backend details, and other highly sensitive information, highlighting the critical need for organizations to scrutinize and secure all AI apps.

**DESPITE THE RISKS, OUR ANALYSIS FOUND:**



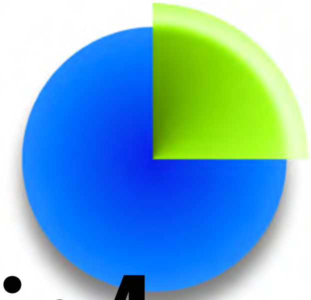
**98%**

of companies have employees using unsanctioned apps, including shadow AI. Each company has **1,200** unofficial apps on average



**52%**

of employees use high-risk OAuth apps



**1 in 4**

unverified OAuth apps (200 out of 800) in the average organization are high-risk

# *SPOTLIGHT:* MICROSOFT 365 COPILOT DATA RISK

While unsanctioned apps can heighten the risk of data breaches, even sanctioned apps can threaten sensitive data.

Microsoft 365 Copilot integrates deeply with an organization's data to boost productivity. But it also introduces security risks. Copilot can surface all accessible data, potentially exposing critical information.

Exposed data is just one Copilot prompt away. Let's consider a hypothetical insurance company with 2,000 employees: if each employee enters 20 prompts every day, five days a week, the company has over 200,000 chances for sensitive data to be exposed every week.

**WE FOUND:**

**90%**

of organizations have sensitive files exposed to all employees via M365 Copilot

**25,000+**

sensitive folders are exposed to all employees on average

**6%**

of organizations have sensitive files open to the internet

Labeling files — ensuring data is accurately categorized, managed, and protected from AI misuse — is essential for data governance.

Labeling helps enforce controls like data loss prevention and encryption. It also supports regulatory compliance requirements by showing that data is handled according to legal standards and policies. To succeed, labeling should be automated and continuous.

*Despite the importance of labeling, only 1 out of 10 companies had labeled files.*

*SPOTLIGHT:*

# SALESFORCE AGENTFORCE DATA RISK

Salesforce orgs contain a lot of sensitive data — PII, PCI, financial information, and more. Admins typically oversee the CRM, handling everything from user management to security settings. However, this often means IT and security teams aren't always in the loop. This can lead to security gaps.

Salesforce Agentforce can widen the gap if data is exposed. Agents can surface unprotected sensitive information through natural language prompting, leading to unauthorized data access and misuse.

We found an alarming number of users and service accounts that can easily access all Salesforce data through an Agentforce agent and export that data. Giving any user the permission needed to download all Salesforce data is a disaster waiting to happen.

**100%**

of companies have at least one account that can export all data

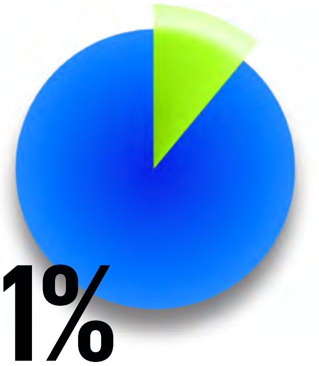
**1 in 10** accounts can freely export all Salesforce data

Our analysis revealed an unexpectedly high number of users with powerful administrative permissions. Allowing many users to install third-party apps, set field-level security, and manipulate permissions increases the risk of exposing critical data in Salesforce and raises the potential for lateral movement.

A typical mid-size enterprise company with 1,000 employees has 110 users that have permission to create and grant permissions and customize applications. If an APT compromises just one Salesforce user account, they can grant access to other attackers or sell highly privileged credentials on the dark web.

Many users also have permission to create public links. Open links can inadvertently enable AI applications, like ChatGPT, to crawl your organization's internal data to answer prompts and give unauthorized individuals access to sensitive company and customer data.

Sharing public links is easy when users have permission. If one of those users becomes compromised, a threat actor can create and use public links to steal sensitive data.



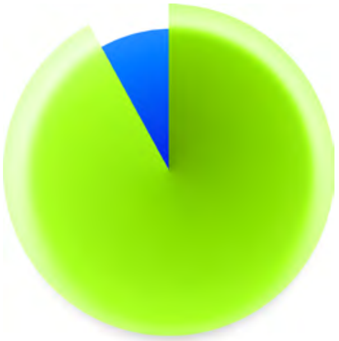
**11%**

of users can grant permissions and install custom applications

**WE FOUND THAT AN ALARMING NUMBER OF USERS CAN CREATE SHARING LINKS:**

**92%**

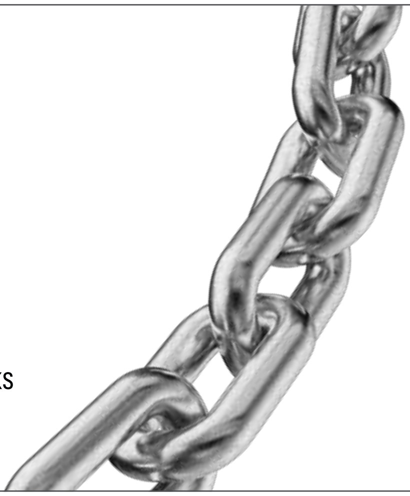
of companies allow users to create public links



of those companies,

**3,689**

users can create public links



# MODEL POISONING AND RISKS TO AI TRAINING DATA

As more organizations develop their own AI processes and products, the data used to train them is placed at risk from breaches and attacks. With training data sitting in multiple clouds or IaaS, it can be challenging to manage permissions uniformly across all environments.

Models trained on sensitive data can inadvertently expose confidential information. Exposed training data can lead to unauthorized access and misuse, compromising the integrity and security of AI systems.

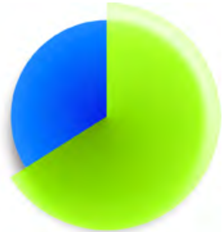
With vast volumes of sensitive information and scores of users to manage, cloud data security can be challenging at scale. Our analysis showed that cloud data, including unmasked data and exposed buckets, is largely overexposed and under-protected.

WE FOUND:

**9 of 10**

organizations have exposed sensitive data in the cloud

**66%**



of companies have cloud data exposed to anonymous users

Data encryption protects the data used to train an LLM by converting it into a secure format that can only be accessed by authorized parties with a decryption key. This ensures that sensitive information remains confidential and is not exposed to unauthorized users. Encryption also helps prevent data breaches and leaks during the training process, from data collection to preprocessing to model training.

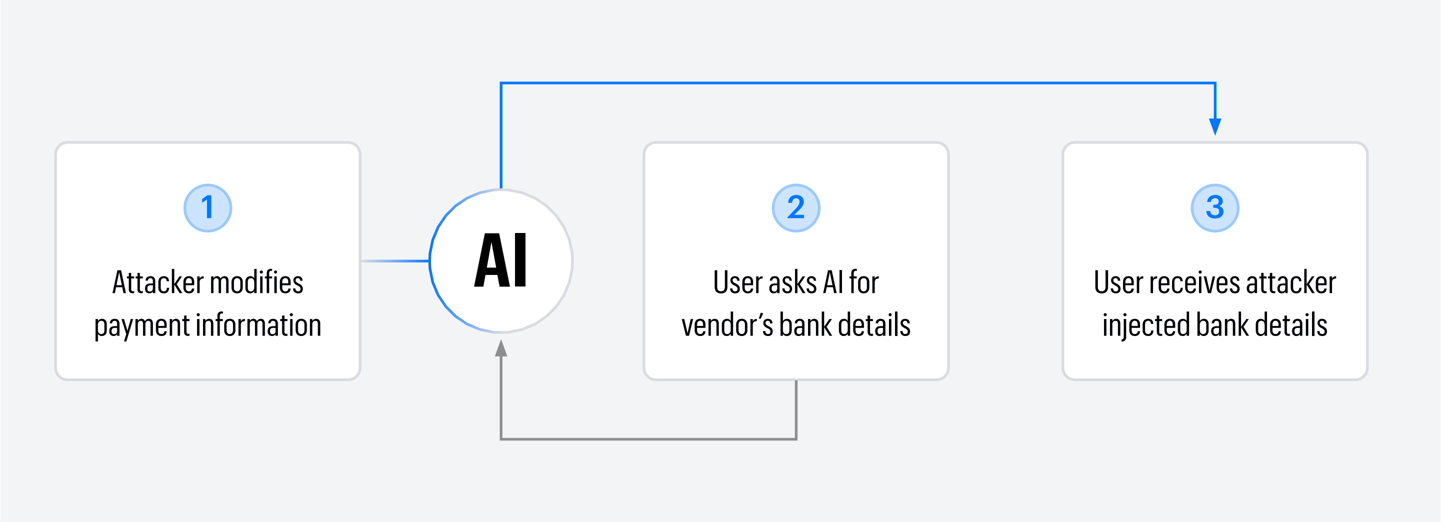
**WE FOUND:**

**2,000** unencrypted object stores in the average organization

**1,500** unencrypted databases in the average organization

Another major risk is model poisoning, where attackers manipulate training data to corrupt an AI model's performance. This happens when a malicious user gains access to the model's cloud resources, such as containers, storage accounts, and databases, and can write to or modify those resources without triggering alarms.

Model poisoning can lead to dangerous outcomes. Imagine an attacker modifying payment information details used in a model. Unaware, the company deploys the model. When a user asks for the vendor's bank details, they are provided with the bank details that the attacker injected.


1
Attacker modifies payment information
AI
2
User asks AI for vendor's bank details
3
User receives attacker injected bank details

Model poisoning can also happen accidentally. Imagine an analyst at a healthcare company that inadvertently trains a model on bad data. Without accurate data, doctors and healthcare staff can make the wrong decisions for patient health. If model poisoning is not detected early, it would be tough to detect it by monitoring the model's performance.

# GHOSTS IN THE MACHINE:

## *EXPLOITING THE BLAST RADIUS*

Once inside your environment, attackers aim to land and expand. Stale user accounts, or "ghost users," are active accounts of former employees or contractors. These ghost users can remain enabled, providing access to applications and data, potentially allowing attackers to conduct reconnaissance or exfiltrate data without triggering alarms.

Managing active identities is crucial, especially with the deployment of AI agents. If just one identity is compromised, bad actors can move laterally, use AI to find sensitive information, or plant malicious software. Regular audits and robust identity management practices ensure that only needed identities remain active and that each identity has appropriate access.

**88%**

of organizations have stale but enabled ghost users, with 15,000 per organization on average

**176,000**

inactive external identities in the average organization

**10**

stale users with admin roles in the average organization

**>31,000**

stale permissions in the average organization

Insiders with legitimate access can inadvertently or maliciously expose sensitive data. Strengthening identity management protocols can help mitigate these risks by ensuring only authorized users can access critical systems and data.

WE FOUND:

**7 of 8** organizations have sensitive data exposed to ***every*** user

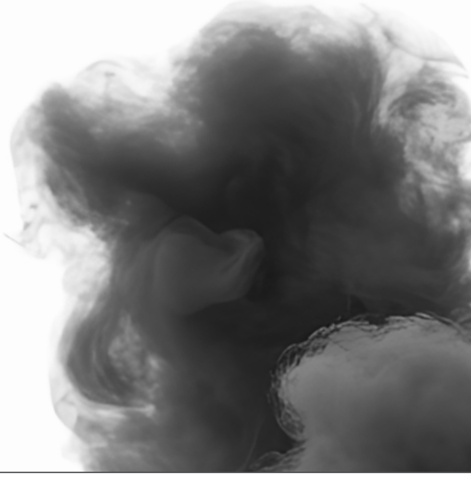
# CLOUD IDENTITIES:

## *SPRAWLING AND COMPLEX*

Groups, memberships, and roles can build up over time. A single user can gather dozens of roles and group memberships. Meanwhile, understaffed IT and security teams often struggle to revoke unused or unnecessary memberships when users change roles or leave the company.

Our analysis shows organizations have fallen behind in managing permissions and securing identities, particularly non-human identities like APIs and service accounts. Poor credential management and excessive privileges lead to unauthorized access and data breaches.

Cloud identity and entitlement management can involve thousands of permissions and roles, requiring constant work to stay ahead of risk. AWS alone has **over 18,000 possible identity and access management permissions** to manage.



**OUR ANALYSIS OF ORGANIZATIONS USING AWS SHOWED:**

**20,224** managed policies in the average AWS account

**3,087** over-permissive policies in the average AWS account

# MISSING MFA:

## *A CRITICAL SECURITY GAP*

Unenforced MFA simplifies attackers' efforts, leaving accounts vulnerable to password spraying, credential stuffing, and phishing attacks. Without MFA, unauthorized account access becomes easier, leading to potential data breaches. MFA is effective only when enabled and enforced.

The largest breach of 2024, which resulted in the **loss of 190 million patient records,** was attributed to missing MFA. As a result, **proposed changes** to HIPAA currently include MFA as a non-negotiable requirement.

**WE FOUND:**

**1 in 7**

organizations do not use or enforce MFA across their SaaS and multi-cloud environments

**1,800**

users have non-expiring passwords in the average company

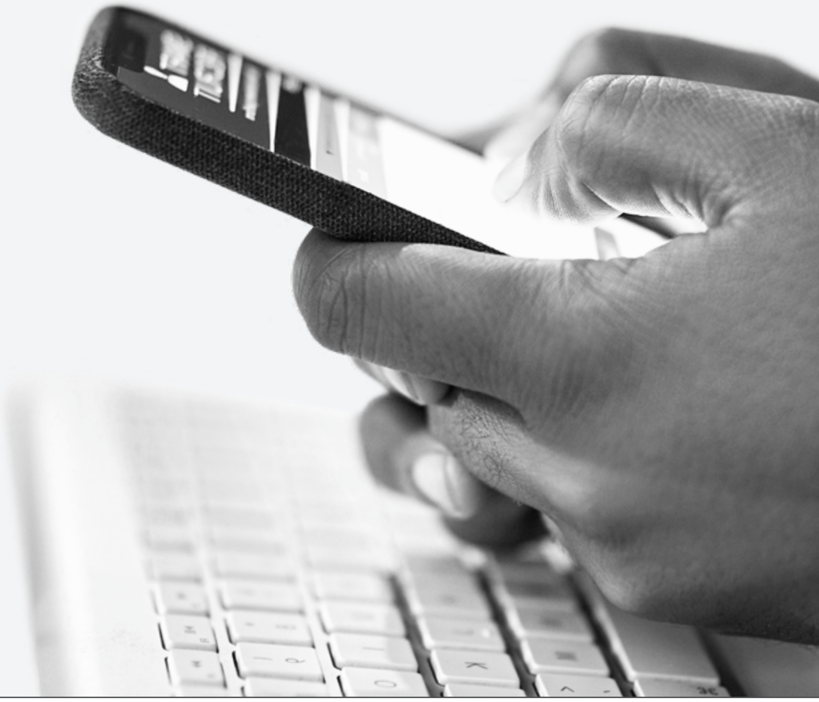
**5**

global admins have passwords that never expire in the average company

Without appropriate authentication controls, attackers can simply log in with stolen credentials and gain access to AI tools to quickly find the most valuable data in your organization.

## *The campaign targeting Snowflake in 2024 is a perfect example of why every company should enforce MFA by default.*

Hackers **used stolen credentials** and leveraged missing MFA to access multiple customer environments before exposing customer data on the dark web. Investigations showed that the attackers accessed Snowflake accounts via compromised credentials from a third-party contractor.



## *MFA alone is not a silver bullet.*

**Varonis Threat Labs** showed how attackers can bypass MFA using stolen browser cookies. By using custom-made malicious browser extensions and automation scripts, attackers can extract and reuse authentication cookies to impersonate users without needing credentials, while keeping persistence.

The proof-of-concept, called **"Cookie Bite"** grants unauthorized access to M365 apps for further recon and privilege escalation. The research shows these techniques also apply to many other cloud platforms and services.

# THE STATE OF DATA SECURITY

# AI'S THREAT LOOMS LARGE

The evidence is clear — organizations have a ticking time bomb in their data security strategy: AI.

AI presents new dangers to data, and organizations must take proactive steps to secure their critical information.

## 1. Reduce your blast radius.

Assume that breaches will occur. Proactively decrease the damage an attacker can do with just one stolen identity. Aim to minimize your blast radius by continuously monitoring data and remediating issues, locking down permissions and access to prevent identity-based attacks, and monitoring AI copilots, chatbots, and agents to prevent exploitation and misuse.

## 2. Data security is AI security. Data powers AI.

To prevent AI risks and AI-related data breaches, continuously monitor your data, automate access governance and posture management, and employ proactive threat detection. A holistic approach to data security keeps AI secure.

## 3. Use AI for good.

AI is a powerful tool for defenders. IT and security teams can harness AI and automation to:

- Accurately identify, classify, and label sensitive data across large datasets, ensuring no critical information is exposed or at risk
- Remediate vulnerabilities and serve as a frontline SOC analyst
- Catch malicious insiders and abnormal behavior that indicates an attack

UNIFIED DATA SECURITY. AUTOMATED OUTCOMES.

# PARTNER WITH THE LEADER IN DATA SECURITY.

The Varonis Data Security Platform was named a **Leader and a Customer Favorite** by a top analyst firm and was Gartner® Peer Insights™ **Customer's Choice and #1-rated DSPM.**

## Request your free Data Risk Assessment.

Find out how Varonis accelerates your progress, not your risk.

### Access to the Varonis platform

Get full access to the Varonis Data Security Platform for the length of your assessment at no cost.

### Dedicated IR analyst

Our experts will monitor your data during your assessment and call you if they see anything alarming.

### Key findings report

A detailed summary of your data security risks that is yours to keep, even if you don't become a customer.

Get your assessment