IDC

SPOTLIGHT

Sponsored by: Cyberhaven

Sensitive and confidential data are essential to business growth and advancement, particularly with AI. Understanding where this data lives and how it is being used provides important context for how to protect it for AI use.

Rethinking Data Security and Insider Risk for Trusted AI Adoption

April 2026

Written by: Jennifer Glenn, Research Director, Information and Data Security

Introduction

Organizational data is the foundation of digital business. It provides the direction for business strategies and innovation, as well as insights derived from customer feedback, and fuels employee productivity. As artificial intelligence (AI) becomes embedded in business operations, data assumes an even more central role. AI systems depend on access to large volumes of data to deliver operational efficiency, automation, and improved business processes.

At the same time, the growth of AI introduces new challenges for security and IT leaders regarding data security and governance.

Two of the biggest factors impacting data risks in the AI era are volume and sprawl. Most organizations are storing, processing, and using massive amounts of data. This data exists in structured and unstructured environments spread across endpoints, applications, files, servers, and databases. While storing all this data can be quite expensive, it also costs the organization time and money to manage. In addition, too much data can limit visibility into important risks, which can incur costs in terms of brand reputation, AI trustworthiness, and compliance or privacy fines. Why?

- **Hidden security liabilities:** As collaboration tools and AI applications expand, the likelihood of unintentional exposure increases exponentially. Insider risk includes both malicious actors and well-intentioned employees. When sensitive data is not clearly mapped and monitored, trusted users may inadvertently expose information through misdirected sharing, misconfiguration, or inappropriate AI prompts.

  While unintentional exposure by trusted users remains common, disorganized data environments also create opportunities for malicious insiders or compromised accounts. Sensitive data scattered across endpoints and other repositories without consistent monitoring may present attractive targets. External attackers may gain access through vulnerable systems and can stay hidden among the chaos.

AT A GLANCE

KEY STATS

- Nearly half of organizational data is considered sensitive or confidential.
- However, 32% of respondents to IDC's survey had more than 75% of sensitive data mapped and monitored.

KEY TAKEAWAYS

- Sensitive data is challenging to identify and classify, putting the business at risk from insider threats and AI exposure.
- Centralized visibility and the context of organizational data enable organizations to effectively secure important data without hindering AI innovation.

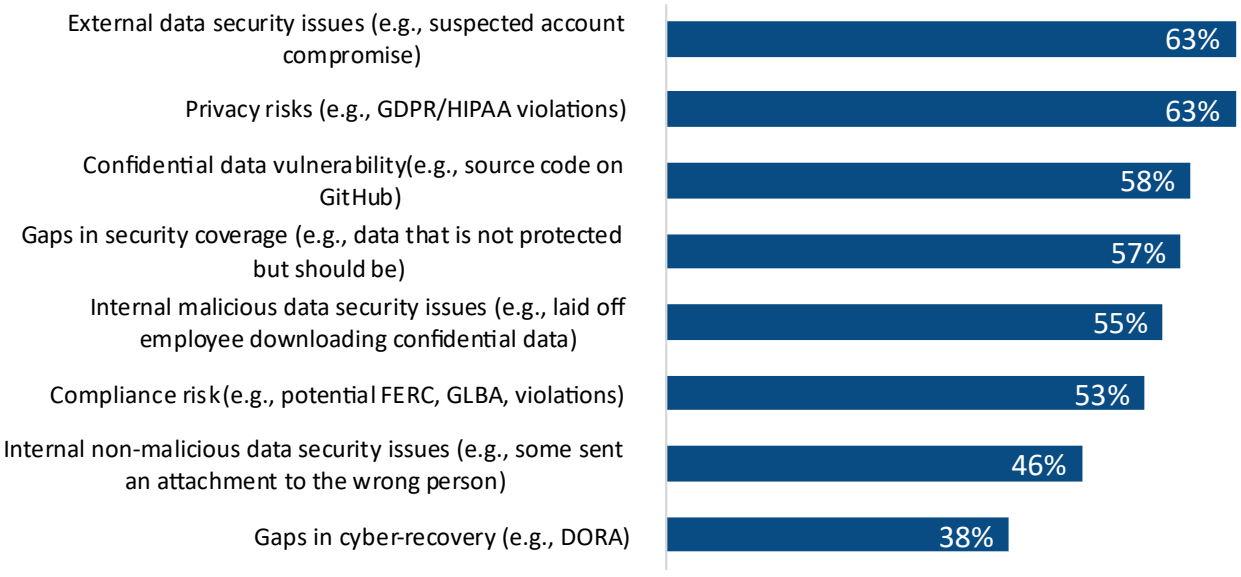
- **Compliance and privacy risks:** In many organizations, this distributed landscape has evolved faster than governance and security practices. Regulatory frameworks and data residency mandates require organizations to know where sensitive information resides and how it is being used. Without this visibility, compliance becomes reactive and resource intensive.

  The advancement of AI agents introduces additional considerations. AI agents operate based on granted permissions and available data access, without human intuition or contextual judgment. If access controls are overly broad or misaligned, AI systems may retrieve and process sensitive data beyond the intended use. In environments with fragmented visibility, these exposures may go undetected. According to IDC's April 2025 *Data Security and Privacy Survey,* privacy, compliance, and insider risk are among the top data security and privacy concerns for business leaders (see Figure 1).

- **Poor-quality output and AI distrust:** Finally, while not directly tied to security risks, poor data management can impact AI quality and, in turn, the company's reputation. As AI becomes embedded in business workflows, stakeholders increasingly expect transparency regarding how data is sourced and used. Excessive amounts of data not only hide liabilities but also keep valuable information from being used correctly. For example, poorly governed data may include duplicate information, misclassified data, and outdated assets, which can produce biased outputs or incorrect information, undermining confidence in AI tools.

Figure 1: Business leaders concerned with multiple data security issues

Q *What data security and privacy risks are your security team responsible for reporting to business leaders?*



n = 618

Note: Multiple responses were allowed.

Source: IDC's Data Security and Privacy Survey, April 2025

The shift toward data-centric security

Since AI is reliant on data for its success, it makes sense to center security controls around the data itself. In IDC's April 2025 *Data Security and Privacy Survey*, respondents highlighted the use of multiple data security tools, such as encryption, data loss prevention (DLP), and data access governance (DAG).

While enforcement solutions, such as DLP, are important, capabilities such as data discovery, classification, and data security posture management (DSPM) are rapidly becoming foundational components of modern data security programs. These tools are designed to provide answers to fundamental questions about data assets, including:

- What types of data exist within the organization?
- Where does that data reside?
- Who or what has access to it?
- Is it protected appropriately, given its sensitivity and regulatory requirements?

Answering these questions enables organizations to organize, analyze, and manage data more effectively. By identifying misconfigurations, excessive permissions, and concentrations of sensitive data, these solutions offer valuable context to improve the accuracy and effectiveness of DLP and DAG tools.

Technology effectiveness also depends on integration and coordination. Discovery, classification, monitoring, and enforcement mechanisms must work together to provide meaningful context. Siloed tools may generate fragmented views, making it difficult to prioritize remediation efforts or align controls with business objectives.

For business and security leaders, trust, transparency, and unified data visibility are increasingly interconnected. Data-centric security platforms that provide contextual insight into data usage and movement can help bridge this gap.

Benefits of data-centric security platforms

The contextual understanding of organizational information that comes from a data security platform supports policy enforcement around data movement, exfiltration prevention, access control, and retention management. Furthermore:

- **From a technology perspective:** Centralized visibility establishes a foundation for data hygiene. Knowing where data resides and how it is used enables faster and more accurate risk assessments. Risk prioritization becomes achievable. Rather than attempting to remediate every potential issue simultaneously, security teams can focus on high-impact exposures. Improved policy accuracy also supports operational efficiency, reducing unnecessary friction while maintaining appropriate safeguards.

  Enhanced visibility contributes to more efficient responses to privacy and compliance audits. Automated classification, when supported by AI-driven analytics, can accelerate the identification of sensitive information and improve consistent enforcement across all environments. This enables organizations to respond more quickly to regulatory inquiries and proactively identify policy violations.

- **From a business perspective:** These capabilities support confident AI innovation. When organizations understand their data landscape, they can address risks before deploying AI applications. This reduces the likelihood of retroactive remediation and strengthens stakeholder trust.

  AI initiatives also benefit from improved data readiness. Well-classified, organized data can be more readily used to train models or power AI-driven applications. Automation and visibility reduce the time and effort required to prepare data sets, accelerating time to value.

  Finally, operational efficiencies emerge from reduced manual investigation and remediation efforts. Automated discovery and monitoring reduce the administrative burden on security and compliance teams, enabling more strategic resource allocation.

Considering Cyberhaven for AI data security and insider risk management

Cyberhaven, headquartered in Mountain View, California, is a vendor of data security solutions that aim to help organizations achieve data visibility and control. In November 2025, Cyberhaven added a data security posture management solution to its platform. The Cyberhaven DSPM capability is designed to provide organizations with unified visibility, classification, and control over sensitive data across cloud, on-premises, endpoint, and AI environments, with the intent to enhance the effectiveness of DLP enforcement. Key features of Cyberhaven's data security platform are:

- **Automated discovery and classification:** This feature uses AI analytics to identify sensitive and confidential information, regardless of where it resides within the organization.
- **Real-time monitoring and risk assessment:** This capability enables security teams to detect anomalous behavior, unauthorized access, and potential data exfiltration across all environments.
- **Granular policy enforcement:** This function allows organizations to tailor controls based on data sensitivity, user roles, and regulatory requirements.
- **Integration with existing tools:** This feature works with existing security tools and workflows, supporting unified management and reducing operational complexity for security teams.
- **Compliance with industry and privacy regulations:** This capability provides detailed reporting, audit trails, and automated policy enforcement for both industry regulations (e.g., GLBA) and privacy laws.
- **Data lineage:** This functionality includes information about how and where data moves, offering transparency and trust in AI outputs.
- **Insider risk management:** This feature addresses insider risk through behavioral analytics and continuous monitoring, supporting early detection and mitigation of both malicious and non-malicious threats.

Challenges

The complexities of securing organizational data in the AI era will continue to increase. Data is continuously created, modified, and shared as it moves between new systems, applications, and AI tools. Achieving visibility will require ongoing effort to not only locate sensitive data but also identify which users and agents can access and use it. Providing context and analysis of the data will become critically important to maintaining security while implementing AI throughout the organization.

It is not uncommon for enterprise organizations to use multiple security tools for data discovery, classification, and protection. Multiple data security tools can result in siloed policy creation and enforcement, as well as an administrative burden on the practitioners who need to monitor and maintain these solutions. A platform that connects these tools offers benefits in terms of consistency, management, and visibility.

Integrating security tools into a platform, such as Cyberhaven's, is one approach to creating the broad visibility and context necessary to tackle AI data security. However, integrating diverse systems can be a challenge and may require specific customization or support services to be effective.

Conclusion

AI adoption is forcing organizations to rethink how they organize, analyze, and control data. As AI becomes further embedded in enterprise workflows, data governance and security can no longer be treated as secondary considerations. The scale, distribution, and dynamic movement of data require that organizations shift their focus to visibility, context, and accountability.

Organizations must continuously review where sensitive information resides, how it is accessed, and whether the security controls around that data are adequate to meet regulatory, operational, and ethical requirements. Organizations that deploy advanced data security tools, such as discovery, classification, DSPM, and DLP, from a unified platform will be in a better position for successful AI implementations.

Trusted AI outcomes require trusted data foundations. Organizations that prioritize unified visibility, contextual analysis, and insider risk management can use AI to innovate confidently while reducing exposure, maintaining compliance readiness, and building stakeholder trust in a data-driven world.

Trusted AI outcomes require trusted data foundations. Unified visibility, contextual analysis, and insider risk management are essential for building AI-ready data.

About the analyst

Jennifer Glenn, *Research Director, Information and Data Security*

Jennifer Glenn is research director for the IDC Security and Trust Group and is responsible for the information and data security practice. Ms. Glenn's core coverage includes a broad range of technologies including messaging security, sensitive data management, encryption, tokenization, rights management, key management, and certificates.

MESSAGE FROM THE SPONSOR

Cyberhaven protects sensitive data wherever it lives and goes. Built for the AI era, Cyberhaven's unified data security platform combines DSPM, data loss prevention, insider risk management, and AI security with deep data lineage and agentic AI. Cyberhaven helps organizations stop data loss, reduce insider risk, and enable AI adoption securely, without slowing their business.

For more information, visit: www.cyberhaven.com.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
One Beacon Street
Suite 33100
Boston, MA 02108, USA
T 508.872.8200
F 508.935.4015
blogs.idc.com
www.idc.com