

The Security Leader's Guide to Generative AI & Data Security

A comprehensive resource for stopping data leaks
as GenAI changes the landscape


Generative AI (GenAI) is making your business more efficient and more effective. It's solving a decade-long resource crisis. But it creates an entirely new data security landscape that security leaders are charged with addressing—fast. Inputting information into GenAI applications often means this data is used to train the models further, potentially exposing sensitive data like source code, personally identifiable information (PII), passwords, roadmaps, presentations and other trade secrets. **Gartner's Forecast: Information Security and Risk Management, Worldwide, 2021-2027, 4Q23 Update** report predicts that in 2027, 17% of the total cyberattack/data leaks will involve GenAI.

High-profile cases of GenAI data leaks:




Microsoft

Microsoft AI researchers accidentally leaked 38 terabytes of private data, including employee passwords and internal messages, when sharing training data on GitHub. The culprit: Insecure access settings for their cloud storage, highlighting the new data security challenges of large-scale AI development.



SAMSUNG

Samsung engineers trying to leverage ChatGPT for work accidentally leaked confidential data like source code and meeting notes through the platform, highlighting the risk of sensitive information being retained and used for unintended purposes.



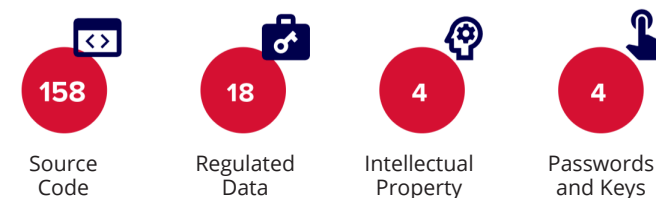
ChatGPT

A glitch allowed users to see the titles of other users' conversations, which also suggested **OpenAI** has access to user chats. OpenAI CEO Sam Altman acknowledged the "significant" error, which has now been fixed. Although users were assured that actual chats weren't accessible and the chatbot was briefly disabled to address the issue, many remained worried about privacy.

Top 5 types of identifiable files uploaded over a 30-day period into GenAI platforms:



Top 4 types of sensitive data that are frequently posted by employees in ChatGPT:



*Frequency of users sharing sensitive data per 10,000 enterprise users monthly

Our 2024 Data Exposure Report discovered that **86% of cybersecurity leaders worry that employees may input sensitive data into GenAI tools which competitors could then access. Or, confidential data entered into these platforms would be added to that language model, potentially exposing it to untrusted users generating malicious outputs – creating even more risks to compliance and intellectual property.**

- 1.** Loss of competitive advantage: Imagine your competitor using your own company roadmap, generated by a sophisticated GenAI platform trained on snippets of your confidential presentations. Leaks of source code, product plans and marketing strategies can cripple your competitive edge, allowing rivals to replicate your innovations or launch similar products before you can reach the market.
- 2.** Litigation nightmares: GenAI models trained on your customer data create a significant privacy risk. Regulations like GDPR, HIPAA and CCPA mandate robust data security measures. A single insider-related data breach involving PII such as names, addresses and social security numbers can trigger a wave of lawsuits and hefty fines under those regulations.
- 3.** Loss of reputation: When customers provide their data, they expect businesses to protect it with proper security measures. However, GenAI platforms might inadvertently receive and use this sensitive information to train models, potentially exposing it to millions of users. A data breach involving customer data, especially financial records or proprietary data, can shatter that trust. News of such leaks can spread like wildfire on social media, leading to a public relations nightmare and a significant drop in customer confidence.
- 4.** Openings for bad actors: GenAI platforms trained on leaked login credentials can launch hyper-realistic phishing attacks that can bypass traditional detection methods. These can mimic real login pages and create 'personalized' emails with uncanny accuracy, tricking unsuspecting employees into revealing even more sensitive information. Some can even generate dynamic content that adapts in real-time based on user interaction.

A KPMG survey indicates that US and global CEOs feel unprepared for the rapid adoption of GenAI, and foresee difficulties in staying ahead:

- 60%** expect it will take one to two more years to introduce their first GenAI solution.
- 67%** are either in the preliminary stages or have not yet begun evaluating risks and mitigation strategies, facing issues such as inaccuracy, cybersecurity and data privacy.
- 68%** have yet to assign a leader or team to coordinate generative AI efforts, often due to the absence of key enablers such as talent and governance.




Balancing Act: Encouraging Innovation Without Compromising Data Security

A holistic approach to insider-related data leaks with GenAI on the rise

While those approaches may seem like simple solutions, they can backfire. Employees eager to leverage the latest advancements may resort to shadow IT solutions – circumventing restrictions by using unauthorized tools and platforms outside of company oversight. This creates a blind spot for security teams and increases the risk of inadvertent data leaks. So, how can companies strike a balance?

1. **Detect and mitigate risks:** Stay ahead with modern risk detection solutions that don't rely on heavy content inspection or classification. To combat insider-driven data leaks, you need to understand who is moving data, its origin and destination. Deploy a solution that identifies known risks, and surfaces unknown risks from day one with no policy setup.
2. **Educate and empower:** Train employees on GenAI security risks and empower them to report accidental or potential breaches. This can be done with controls for responding to mistakes and unacceptable activities. Implementing a solution that both blocks threats and uses pop-up education and alert-triggered video lessons to correct employee errors aids in reducing event volume and risk. Integrated micro-trainings, like **Mimecast Instructor**, ensure employees learn to avoid risky behaviors and automate responses to low-severity risk.
3. **Contain and control:** Training employees reduces most incidents, but user error is inevitable. You need a solution that gives you containment controls that will enable you to swiftly minimize the damage and then investigate risk to secure the data.
4. **Block high-risk data sharing:** A comprehensive response strategy stops high-risk users from sharing data with unsanctioned GenAI platforms. Blocking their activities ensures secure collaboration for the rest of your organization. A data protection solution with full response controls helps address this issue and fosters a security-first culture.
5. **DLP for defense:** Upgrade and invest in modern Data Loss Prevention (DLP) solutions that can monitor and restrict data movement across unauthorized channels, including GenAI platforms.

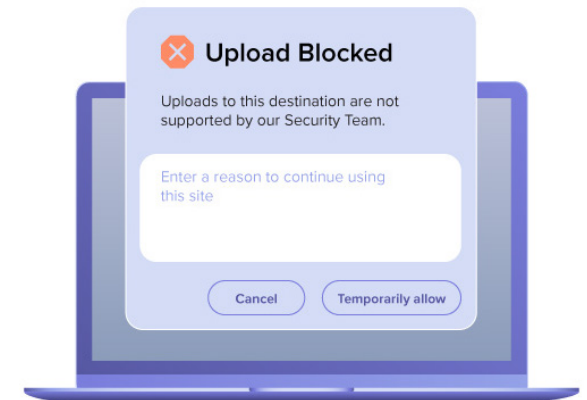
Cisco's **2024 Data Privacy Benchmark Study** found that many companies are aware of the risk of data leaks through GenAI platforms and taking measures to limit exposure:

-  **63% have set restrictions on data input.**
-  **61% limit the GenAI tools employees can use.**
-  **27 %have temporarily banned GenAI applications altogether.**

Most companies rely on traditional DLP tools, which require constant fine-tuning and drain time and resources from an already overburdened workforce. Yet, they still face numerous insider-driven data incidents as they're ill-equipped to face **modern risks stemming from AI/GenAI** and the spread of cloud applications.

A holistic response strategy with automated training to handle low-risk events and block unacceptable activity is ideal. With companies aiming to find a unified solution to quickly detect and stop threats, **Incydr offers the streamlined answer:**

1. Visibility into cloud and endpoint exfiltration in one, including Git push/pull activity, Salesforce downloads, Airdrops and cloud syncs.
2. Correct users when data is shared incorrectly with integrated security lessons that prevent risky activity from becoming the norm.
3. Validate actual file contents to know for sure how sensitive the data is.
4. Implement real-time blocking for high-risk employees working closely with intellectual property.



About Mimecast

Mimecast Incydr™ is the leader in data loss and insider threat protection. Native to the cloud, Incydr data protection rapidly detects data exposure, loss, leak, and theft and speeds incident response – all without lengthy deployments, complex policy management, or disrupting employee productivity. The solution offers a complete range of response solutions, including automated microlearning modules for accidental non-malicious risk, case management for efficient investigation collaboration, and automated blocking for the highest-risk use cases. The Incydr IRM Program Launchpad helps organizations get up and running quickly to ensure success and return on investment.

With Incydr, security professionals can protect corporate data and reduce data loss from insiders while fostering an open and collaborative culture for employees. Innovative organizations, including the fastest-growing security companies, rely on Incydr to safeguard their ideas. Incydr's data protection solution is FEDRAMP-authorized and can be configured for GDPR, HIPAA, PCI, and other compliance frameworks. Founded in 2001, the company is headquartered in Lexington, Massachusetts, and backed by Accel Partners, JMI Equity, NewView Capital, and Split Rock Partners. Incydr has played a defining role in developing a vision and requirements for the IRM category and is a founding member of the Insider Risk Community.

Incydr data protection offers tangible cost benefits for customers, as found in a commissioned report by Forrester Research. Deploying in just two weeks, the solution pays for itself in 6 months, before most DLP tools are even off the ground. The average organization can expect to see a 172% return on investment, including savings from data loss that total over \$680,000 as well as powerful team time savings – with 50% faster closing of incidents.

The Company has several offices across the United States, and its clients include the most recognizable security, technology, manufacturing, and life sciences organizations, such as CrowdStrike, Okta, Lyft, BAYADA Home Health Care, Rakuten, Sumo Logic, MacDonald-Miller, MACOM, Ping Identity, Shape Technologies, and Snowflake.