CROWDSTRIKE

White Paper

Detecting and Stopping Data Loss in the Generative AI Era

Generative AI Is a Tremendous Opportunity — with Caveats

Though it is still new to the market, many organizations see generative AI (GenAI) as a force multiplier. According to TechTarget, nearly half of organizations (48%) have (or plan to have) a budget allocated to GenAI.[1] GenAI is predicted to become a $1.3 trillion USD market by 2032.[2] But while the adoption of GenAI is accelerating, only 21% of organizations have implemented policies to regulate the use of GenAI by employees.[3]

In the modern business landscape, data resides in and moves across clouds when employees download, store and transfer it to various web applications and USB devices. The steep average cost of a breach — $4.45 million[4] — underscores the significance of data protection for modern enterprises in the GenAI era.

48%

of organizations have (or plan to have) a budget allocated for GenAI

$1.3 trillion USD

is the predicted GenAI market by 2032

Only 21%

of organizations have GenAI usage governance/policies

GenAI Risks and Data Protection Challenges

Marketing, sales, product and service development teams are the teams that most commonly use GenAI in organizations.[5] When these teams leverage AI, they risk putting important data into GenAI models, like proprietary product or technical information, design documentation, source code, customers' personally identifiable information (PII), confidential sales data and more. All GenAI models rely on data for training, and organizations should create a governance framework for their employees to prevent sensitive data from leaking through GenAI tools. The framework should go beyond employee training — it should include enforcing security controls and conducting periodic assessments of data egressing the organization. But when organizations decide to completely block GenAI usage to contain data leaks, they can hinder user productivity, innovation agility and revenue growth.

Despite cybersecurity concerns, the number of legal professionals using tools like ChatGPT or in-house equivalents every month has increased to 26%, up from 11% in July 2023.[6]

Though GenAI is a powerful productivity tool, organizations risk losing sensitive data due to accidental exposure or malicious data leakage.

Accidental exposure of sensitive information: Employees with access to GenAI tools from their endpoint may inadvertently copy and paste or upload sensitive data. With GenAI, for example, sales employees could compose emails, marketing teams could generate content and developers could generate code — all based on sensitive data they pasted into the GenAI tool. As GenAI models constantly memorize and learn, this sensitive data will become a part of the models' training data, and the models' outputs may reveal the sensitive information.

Malicious data leakage: Adversaries with access to valid credentials and malicious insiders with access to sensitive data stores contribute to data leakage via GenAI tools.

Whether it's accidental or malicious, the potential consequences are catastrophic: damage to brand reputation, compliance fines and loss of intellectual property.

[1] https://www.techtarget.com/searchenterpriseai/feature/Generative-AI-in-business-Fast-uptake-earmarked-funding
[2] https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/
[3] https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year
[4] https://www.ibm.com/reports/data-breach
[5] https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year
[6] https://www.cityam.com/over-a-quarter-of-lawyers-use-ai-tools-despite-security-fears/

GenAI Data Protection Challenges with Legacy DLP Solutions

Organizations have used various data loss prevention (DLP) tools for the past two decades. Despite this, they still face challenges in accurately identifying sensitive data movement. There are four main obstacles to implementing an effective data protection strategy and scaling it to stop data from leaking into GenAI tools:

1. **Overabundance of endpoint agents:** Deploying and managing an excessive number of endpoint agents poses a significant challenge to endpoint performance, agent updates/management and operationalizing data leak prevention.

2. **Inadequate visibility into data flows:** It is a substantial challenge to achieve immediate and consistent visibility into data flows from source to destination — including information on ***who*** (the identity of the user handling the data and the endpoint they are using), ***what*** (the sensitive information in the data), ***how*** (copy/paste, upload, nested ZIP file, mutated data, etc.), ***why*** (contextual information on the data egress operation) and ***where*** (software-as-a-service applications/USB/GenAI).

3. **Reluctance to enforce blocking controls:** Most traditional DLP implementations are in monitor-only mode because organizations are afraid that enforcing blocking controls may negatively impact user experience and productivity.

4. **Balancing GenAI adoption and usage:** Ensuring GenAI usage aligns seamlessly with the organization's data protection strategies requires careful consideration — organizations must fuel innovation without compromising sensitive data.

Increase Confidence in GenAI Adoption with CrowdStrike Falcon Data Protection

CrowdStrike Falcon® Data Protection, a module of the industry-leading CrowdStrike Falcon® platform, takes a modern approach to securing your enterprise data from adversaries. Falcon Data Protection adds the much-needed context to content, automatically connecting the dots across ***who***, ***what***, ***how***, ***why*** and ***where*** in a single console. This capability, which empowers enterprises to prevent data theft in the GenAI era, is missing from legacy DLP solutions.

Let's assume organizations adopting GenAI have governance frameworks with periodic user training on GenAI usage. These are important and nice to have, but they don't stop sensitive data from creeping into GenAI tools. Falcon Data Protection empowers organizations with the "must-haves" for safer, faster GenAI adoption:

Easy deployment: Falcon Data Protection is easy to deploy and manage, as it's built into the cloud-native Falcon platform's unified sensor. This modern software-as-a-service (SaaS)-based approach removes the need for on-premises infrastructure, and the lightweight Falcon sensor is optimized for performance, eliminating the "deployment regrets" associated with legacy DLP solutions.

Instant visibility into data flows: Even before configuring and classifying sensitive data, Falcon Data Protection provides early insights into data movement. Falcon Data Protection offers instant contextual visibility into which type of sensitive data is flowing into GenAI, and this insight can help inform GenAI data protection policies.

Granular data classification: With custom data classification, Falcon Data Protection identifies sensitive data in motion that's trying to egress via web-based GenAI tools in real time. Discover and understand sensitive data that traverses across the organization. See where the data is originating from and egressing to, and understand what sensitive information the data might contain.

Real-time egress protection: After defining custom data classifications, Falcon Data Protection makes it easy to define, test and enforce protection rules for these data classifications to effectively detect and stop data egress via GenAI. Set precedence-based rules to either block by default — i.e., lock down data (with some exceptions) — or allow by default.

Typical Use Cases

Scenario 1 — Accidental Exposure: An overworked analyst in the finance department wants to accelerate the tedious task of drafting an extensive merger and acquisition (M&A) report for a fast-moving acquisition. Fatigued from long hours dedicated to the deal, the analyst decides to streamline the process by utilizing GenAI. The analyst downloads several files like PowerPoint presentations, Excel spreadsheets and PDF files containing confidential strategic information pertaining to the M&A from the organization's managed SharePoint. In an attempt to expedite the report, the analyst copies key sections to a text file and pastes snippets into ChatGPT.

Once data classifications are defined and policies are enforced, Falcon Data Protection:

- Blocks this copy/paste action (clipboard protection) in real time based on the data classification that was identified as sensitive information originating from enterprise-managed SharePoint and file types (.pptx, pdf, .xls, etc.).
- Notifies the user, coaching them that sensitive information sharing is against corporate policy.
- Traces the data back to the original source (in this case, managed SharePoint) with the unique Similarity Detection feature. When the paste action is attempted, Falcon Data Protection blocks the clipboard action because the original file's source is SharePoint. Since snippets of sensitive data were also copied to a text file from several files downloaded from SharePoint, Falcon Data Protection highlights all of the similar files, tracing the sensitive data back to the source and providing end-to-end visibility into GenAI data leakage.

Scenario 2 — Malicious Intent: A disgruntled employee who's about to leave the company decides to feed sensitive information into GenAI. The employee has access to customer PII data — including credit card information and Social Security numbers from Salesforce and managed OneDrive — and downloads them to the endpoint. The employee then decides to copy and paste information from these files into a Word document named "vacation photos" and renames the Word file to "vacation photos.jpg." Afterward, the employee tries to upload the file to GenAI tools such as Canva or Copilot.

Once data classifications are defined and policies are enforced, Falcon Data Protection:

- Blocks this upload action in real time based on the data classification, which detects sensitive information originating from enterprise managed OneDrive and Salesforce along with true file types (.pptx, pdf, .xls, etc.).
- Notifies the user that the sensitive information sharing is against corporate policy.

- Traces the data back to the original sources (in this case, managed OneDrive and Salesforce) with the Similarity Detection capability. When the upload action is attempted, Falcon Data Protection blocks the action because the original file's sources are OneDrive and Salesforce. Since snippets of sensitive data were also copied to a .doc file from several files downloaded from OneDrive and Salesforce, Falcon Data Protection highlights the similar files, tracing the sensitive data back to the source and providing end-to-end visibility into GenAI data leakage — even when the malicious actor tried to evade defenses.

Extend Visibility Across Endpoints and Identities

Falcon Data Protection provides additional contextual visibility by leveraging other Falcon platform modules like CrowdStrike Falcon® Insight XDR endpoint protection and CrowdStrike Falcon® Identity Protection to understand device posture and vulnerabilities, user risk scores, user groups, whether users have privileged access and more. Leverage Falcon Data Protection's rich telemetry to confidently investigate and establish the "**why**" to detect malicious and accidental data exposure on web-based GenAI tools.

"Our old endpoint DLP solution made us navigate different consoles to dig out the data egress incidents and connect the dots manually. CrowdStrike's unified platform approach made it super easy for us to navigate from endpoint incidents to data protection incidents, within the same console, to detect unauthorized data exfiltration."
— Billy Demourelle, Security Analyst at MMR

"Falcon Data Protection has helped provide the visibility we struggled to get with other platforms. Its ability to provide the source and destination of the data has helped us identify high-risk data flows. The CrowdStrike team has also been quick to make product enhancements. Enabling Falcon Data Protection on existing deployed Falcon all-in-one agents minimizes the need for additional resources associated with adopting another DLP agent."

— Data Protection Lead at a Multibillion-Dollar Food Industry Company

Request a demo →

Watch this short video to see Falcon Data Protection in action.

About CrowdStrike

CrowdStrike (Nasdaq: CRWD), a global cybersecurity leader, has redefined modern security with the world's most advanced cloud-native platform for protecting critical areas of enterprise risk — endpoints and cloud workloads, identity and data.

Powered by the CrowdStrike Security Cloud and world-class AI, the CrowdStrike Falcon® platform leverages real-time indicators of attack, threat intelligence, evolving adversary tradecraft and enriched telemetry from across the enterprise to deliver hyper-accurate detections, automated protection and remediation, elite threat hunting and prioritized observability of vulnerabilities.

Purpose-built in the cloud with a single lightweight-agent architecture, the Falcon platform delivers rapid and scalable deployment, superior protection and performance, reduced complexity and immediate time-to-value.

CrowdStrike: We stop breaches.