cyberhaven

From Visibility to Control

A Practical Guide *to* Modern DSPM

Welcome to The Future of Data Security

We stand at the beginning of the Intelligence Age, a revolutionary era driven by AI that has democratized knowledge while magnifying the risk to your most valuable asset: proprietary data. Data has broken free from the systems designed to contain it. It now exists in fragments that move across endpoints, cloud infrastructure, SaaS applications, and AI systems at a pace security teams were never asked to manage before.

Just as cloud computing expanded access to computing power, AI has not only expanded access to knowledge, but is rapidly integrating itself into core business workflows. That expansion and subsequent reliance creates extraordinary opportunity, yet it also reshapes the data risk landscape. Data no longer lives quietly in repositories. It flows continuously and is consumed by AI systems both inside and outside organizational boundaries. The assumptions that once defined data security no longer hold in this environment.

In this new era, data has never been *more valuable*, and protecting it has never been *more complex*.

Data security posture management (DSPM) has emerged to help organizations understand where sensitive data resides and how it is configured across endpoint, cloud, and SaaS environments in this volatile era. It provides essential visibility and answers foundational questions such as where sensitive data exists and who can access it.

However, traditional DSPM approaches are limited in both scope and capability. They typically stop at inventory and configuration, offering only point-in-time snapshots of data posture without accounting for how data actually moves across the organization. Many CNAPP solutions now offer DSPM "add-ons" but focus primarily on cloud data stores while excluding endpoints, yet still position themselves as comprehensive. As a result, security teams are left with an overwhelming volume of information that is difficult to interpret and act on, often without visibility into the most critical data flows. This creates a damaging combination of signal noise and blind spots.

In the age of AI, proprietary data is the true differentiator. That data must be continuously identified, deeply understood, and actively protected as it changes state and context.

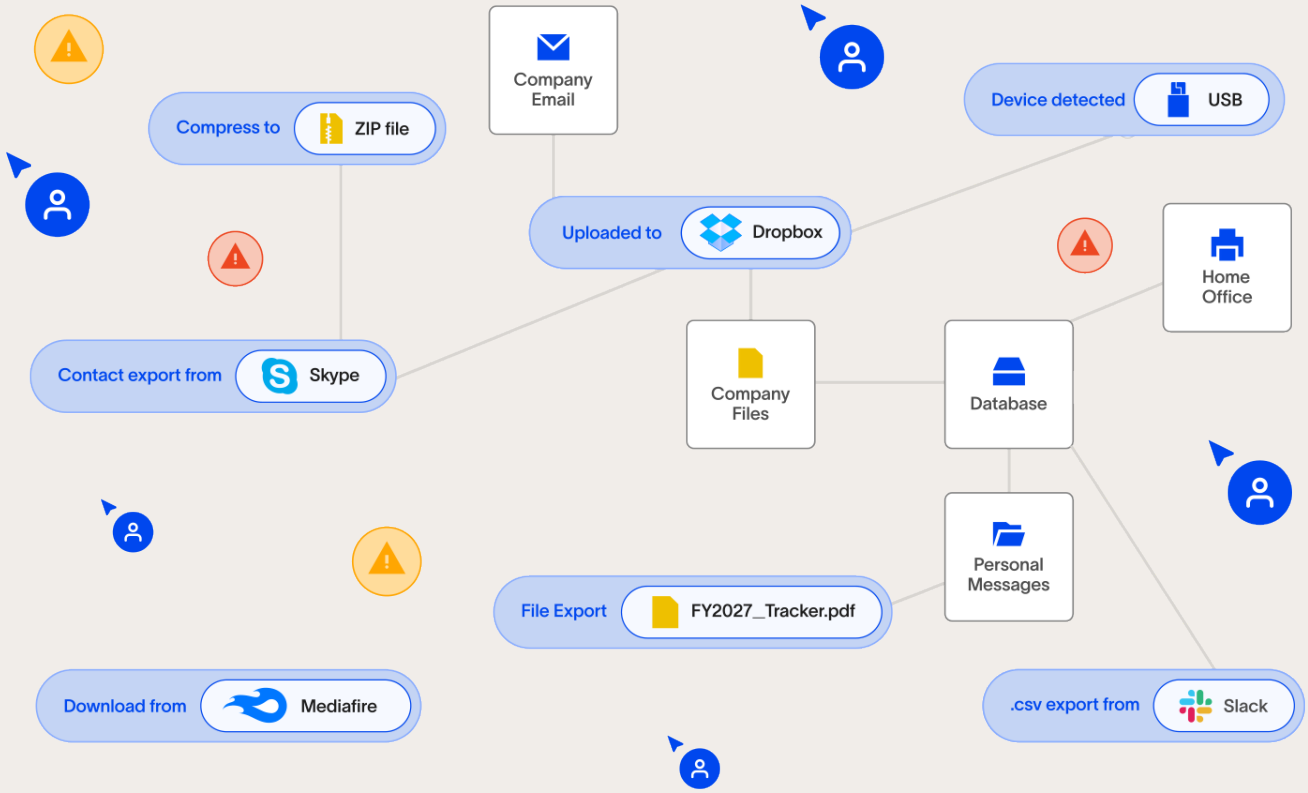
Without modern DSPM, organizations struggle to:

- Discover all of their data across environments and ownership boundaries
- Identify and understand which data carries the highest business and security risk
- Reduce and right-size access as data spreads and permissions drift
- Ring-fence their most critical and valuable data as it is copied, shared, and consumed

That is why DSPM has become a foundational component of modern data security. It creates the data intelligence required to secure information at scale and to support more advanced protections as data flows accelerate.

Let this be your guide to understanding why DSPM has moved from an emerging capability to a strategic requirement, and how security leaders can use modern DSPM solutions to regain control of data in the Intelligence Age.

— **Nishant Doshi**, CEO of Cyberhaven



Section 01

When Data Escaped the Perimeter

Protecting data used to be straightforward because the operating assumptions were stable. Users worked inside the enterprise. Files lived in known systems. Access patterns changed slowly. Security teams could scan, classify, label, and enforce policies around relatively static data. Early DLP and first-generation DSPM tools were built for this world, and for a time, they worked.

That world no longer exists.

Over the past decade, cloud platforms, SaaS applications, and collaboration tools reshaped how work gets done. Most security leaders recognize that this shift dissolved the network perimeter. What is less widely acknowledged is that it also dismantled the file perimeter. Data is no longer shared as complete files; instead, it is copied, rewritten, and pasted in fragments across dozens of tools, rarely carrying labels or behaving like traditional files.

Research from Cyberhaven Labs underscores the impact of this shift. More than 80 percent of data exfiltrated from modern organizations consists of fragments, including pieces of strategic plans, acquisition details, and customer information. These pieces move quietly through browsers, chat tools, SaaS workflows, and cloud services, often without ever triggering file-based controls.

80%

of data exfiltrated consists of fragments

As data began moving freely, security architectures became more complex. Each endpoint, SaaS platform, cloud infrastructure, and identity system introduced its own controls and telemetry.

THE RESULT: a fragmented security stack trying to protect fragmented data.

Teams deploy multiple tools, each seeing only part of the problem, generating competing signals, and providing limited insight into what truly matters.

The consequences are visible. Sensitive data slips through gaps between tools and leaves the organization in pieces that appear harmless in isolation. Most of these incidents are not driven by malicious intent. They are the predictable outcome of modern work moving faster than security models designed for a different era.

Security teams today are not suffering from a lack of information. They are struggling with a lack of clarity. Understanding where that data is, how it is used, and when it becomes risky has become both more critical and more difficult.

Enterprise AI adoption further amplifies this challenge. Even before autonomous agents, AI workflows already consumed data snippets through prompts, summaries, and automated workflows. Small pieces of proprietary information can carry outsized strategic value when aggregated across systems, raising the stakes for data security at the exact moment when traditional assumptions are breaking down.

The perimeter did not fail all at once. It eroded as data became dynamic, fragmented, and continuously reused. Any modern data security strategy must begin with an honest understanding of this new reality.

Section 02

The AI Era: When Intelligence Became the Attack Surface

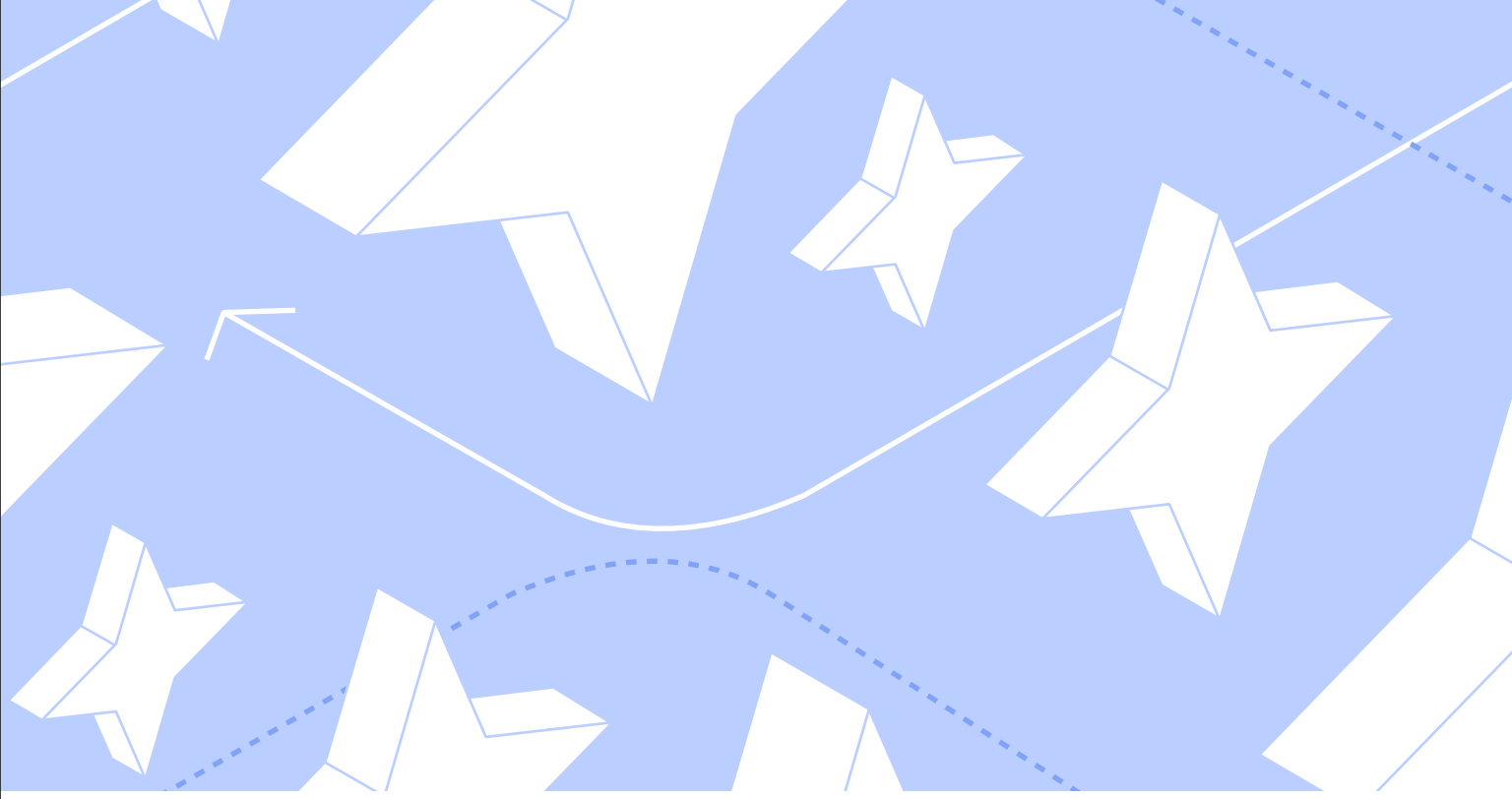
Every major technological era has reshaped who holds power and how value is created. The Industrial Age democratized industry through energy and machines. The Information Age democratized access to information through data and computing power. Today, the Intelligence Age is democratizing intelligence itself, allowing anyone to apply the accumulated knowledge of countless systems at unprecedented speed.

This is the AI era, and the change is coming from every direction. Executives face mounting pressure to adopt AI to remain competitive. Employees are already integrating AI tools into daily workflows, often faster than policies or controls can keep up. Adoption is not a future state. It is already underway. Cyberhaven Labs found that organizations with the highest rates of AI adoption are utilizing over 300 GenAI tools within their enterprise environment.

As AI becomes widely available, a fundamental question emerges. How do organizations gain an advantage when access to intelligence is no longer scarce?

The answer lies in data. Models are increasingly commoditized. Proprietary data, institutional knowledge, and business context are not. Success in the Intelligence Age is defined by the quality, sensitivity, and uniqueness of the data inside an organization, and by how effectively that data is protected as it is used.

Over 300 GenAI tools are being utilized by organizations with the highest rates of AI adoption.



AI accelerates both opportunity and risk. By design, AI systems generate and consume massive volumes of derived data. Prompts, summaries, embeddings, agent memory, and intermediate outputs all create new fragments of sensitive information. This power is amplified by employee use patterns. One-third of employees are utilizing AI from personal accounts, and over a third (39.7%) of all interactions with AI tools involve sensitive data.

These fragments move faster, spread further, and persist longer than traditional files ever did. One individual can now operate with the leverage of an entire team, amplifying both productivity and exposure.

AI agents raise the stakes even further. These systems operate continuously, act autonomously, and access data through APIs and workflows that were never part of traditional security models. They behave like users in many respects, but without human judgment, human pacing, or consistent oversight. They can make mistakes, be misconfigured, or be compromised, all while moving data at machine speed.

For security teams, the implications are profound.

AI adoption is no longer theoretical

By December 2025, nearly half of developers were using desktop-based coding assistants, up from roughly one in five at the start of the year. At the beginning of 2026, nearly a quarter of enterprises had already adopted **agent-building platforms.** AI adoption is accelerating at both the individual and organizational level, with enterprises often running hundreds of generative AI tools in parallel.

AI agents reveal a truth that has been building for years. Risk, data, and behavior converge at the endpoint. This is where data is accessed, transformed, and shared. This is where AI agents operate. In theory, concentrating visibility, tracing, monitoring, and enforcement at the endpoint should bring clarity. In practice, that clarity remains elusive for most organizations.

Legacy security architectures were built to defend network and cloud perimeters and to keep external threats out. The dominant risk model assumed outsiders attempting to gain access. Today, the balance has shifted decisively toward inside-out risk. Data is exposed by legitimate users, sanctioned tools, and autonomous agents operating beyond a perimeter that has already eroded.

Endpoint detection and response (EDR), legacy DLP, and traditional DSPM were not designed for this environment. They were built for a world where humans were the primary actors, applications were known and managed, and data flowed through networks security teams could observe. Autonomous agents operating continuously on endpoints break these assumptions. Each tool may capture a narrow slice of activity, but none provide a coherent view of how sensitive data moves, changes, and accumulates risk across systems.

The mismatch is architectural.

Pre-AI security tools assume that data moves through controlled networks, that actors behave at human speed, that applications are relatively static, and that risky events have recognizable signatures. Agentic AI violates every one of these assumptions. Data moves locally and via APIs outside traditional inspection points. Activity occurs continuously and at machine speed. New tools appear faster than governance can track. Risk emerges gradually through accumulation rather than through a single detectable event.

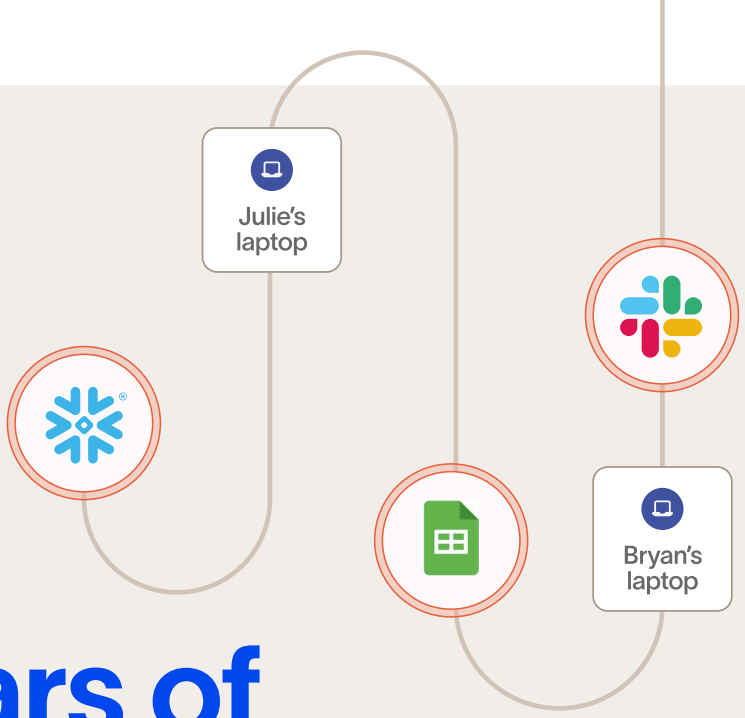
How to Secure AI

Many AI tools operate beyond the visibility of legacy security controls, and autonomous agents introduce variability that cannot be governed through behavior-based rules.

The durable control point is the data. Effective AI security requires continuous understanding of sensitive data and how it is used, at machine speed. This means shifting to data-centric insight that tracks how data relationships change and which human or autonomous actors are involved. Security teams must observe data flows and enforce context-aware policies that distinguish low-risk use, such as testing with synthetic data, from high-risk exposure of production data.

As legacy vendors race to add AI-related features to existing platforms, risk continues to compound. Shadow AI spreads across endpoints. Sensitive data is absorbed into agent memory (persistent data stored by autonomous AI agents) that sit outside established governance frameworks. Security teams are left trying to manage AI adoption. Their tools cannot see how data is actually used at the point of execution.

This is the challenge of data security in the Intelligence Age. It demands a fundamentally different approach, one grounded in how data behaves at the endpoint and how risk emerges when intelligence becomes ubiquitous.



Section 03

The Pillars of Data Security 3.0

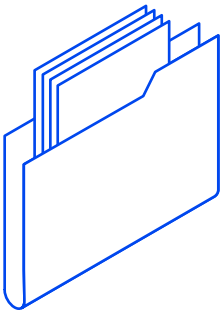
Data security has evolved alongside how data is created, stored, and used. Each generation of tools reflects the assumptions of its time. Many organizations today are still relying on approaches rooted in earlier eras, even as data behavior has fundamentally changed.

Legacy DSPM solutions have failed to meet this moment.

Security teams face obscured visibility, constant alert fatigue, and growing operational burden. Sensitive data now moves continuously across endpoints, SaaS platforms, cloud infrastructure, and AI systems, often outside the reach of any single control. Tools designed to monitor static environments struggle to track data both at rest and in motion, leaving teams with fragmented insights and limited confidence in their security posture.

This gap is not theoretical. The cost and frequency of data breaches continue to rise, while insider-driven and misuse-based incidents remain among the hardest to detect. The challenge is no longer simply finding sensitive data. It is understanding how that data is used, when risk emerges, and how to intervene effectively.

Today, many security teams can discover sensitive data, but they cannot operationalize that insight. Findings arrive without insights, prioritization, or a clear path to action. As a result, DSPM outputs are reviewed periodically rather than used continuously, and data risk remains understood in theory rather than managed in practice.



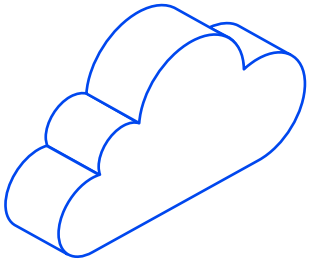
Data Security 1.0

Era: Early 2000s

Security model:
File-centric and perimeter-bound, utilized rule-based engines

Core assumption:
Data lives in known locations and moves slowly

The first generation of data security focused on files and perimeters. Built for on-prem environments, these systems relied on static rules and slow classification engines. As data fragmented and collaboration accelerated, accuracy declined and coverage narrowed. Many teams were forced to concentrate protection on a small fraction of their data, leaving the rest unmanaged.



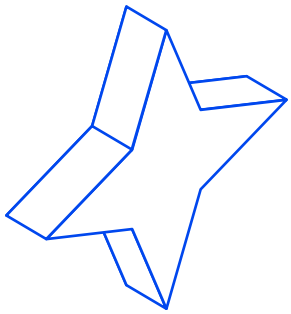
Data Security 2.0

Era: Cloud adoption wave

Security model:
API-based discovery and classification that offered limited coverage, context, or control

Core assumption:
Visibility into cloud data stores is sufficient to manage risk

The second generation expanded visibility through cloud APIs. Discovery improved, classification became more sophisticated, and coverage of cloud data stores increased. This represented meaningful progress, but the approach remained largely observational. Data could be seen, but rarely controlled. Context was limited, endpoint activity was often out of scope, and integration with enforcement systems such as DLP was inconsistent.



Data Security 3.0

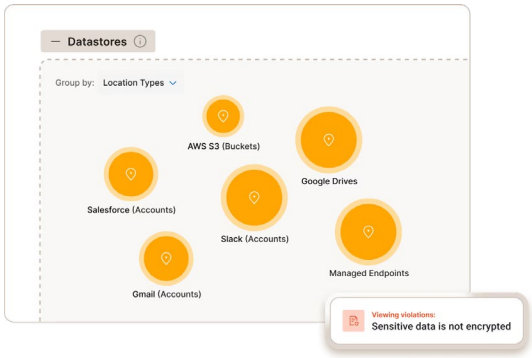
Era: AI-driven, endpoint-centric environments

Security model:
Continuous data understanding with action, end-to-end visibility, deep-data context, and ability to act on risk across the IT environment

Core assumption:
Data risk emerges through movement, context, and use

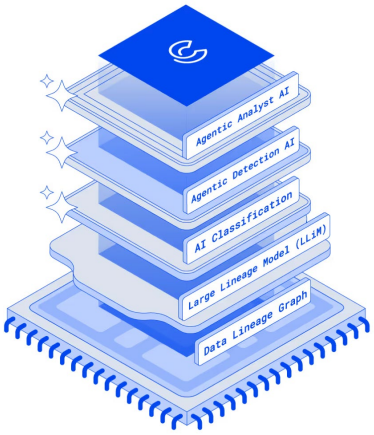
Data Security 3.0 builds on these lessons. It recognizes that data risk is dynamic and contextual. Visibility alone is insufficient, teams also need insight and the ability to act on risk. Modern DSPM must operate across environments, keep pace with data in motion, and support real-time decision-making.

This requires three foundational pillars.



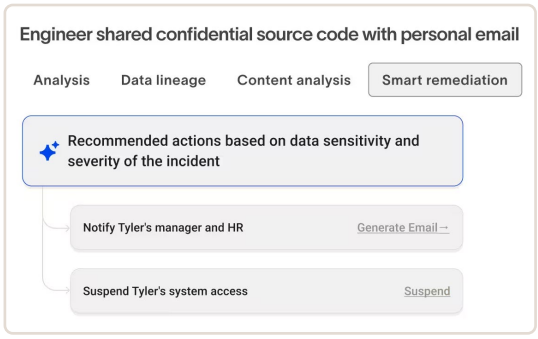
Holistic Visibility and Control

Security teams need a unified view of data across its entire lifecycle. This includes where data lives, how it moves, and how it is used across cloud applications, endpoints, and on-prem repositories. Equally important is the ability to act on that insight, tracing exposure and enforcing protection before risk becomes loss.



Deep Understanding of Data

Modern DSPM must go beyond surface-level classification. It must interpret data in context by understanding lineage, ownership, access patterns, and behavior. Knowing who interacted with data, how it was used, and where it moved next enables accurate risk prioritization and informed response.



Ease of Operation at Scale

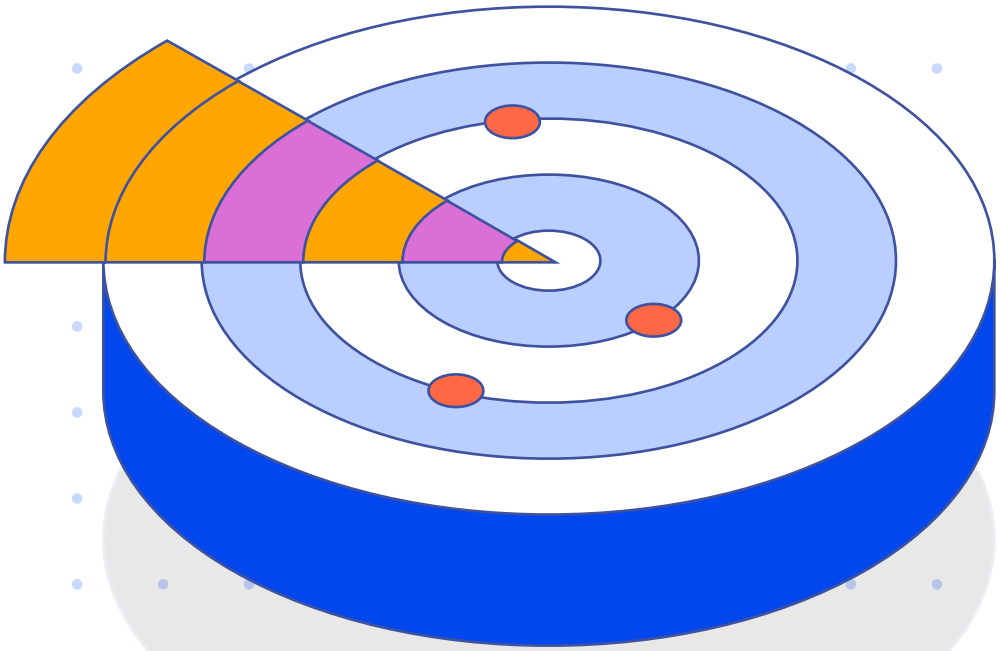
Data security must be practical. Solutions need to deploy quickly, tune easily, and automate intelligently without disrupting users. Tools that generate excessive noise, require duplicated policies, or demand constant manual intervention fail to scale and ultimately fail to deliver protection.

Together, holistic visibility and deep data understanding make security easier to operate. When teams can see clearly and understand context, enforcement becomes more precise, automation becomes effective, and operational burden declines.

Data security 3.0 reflects how data actually behaves today. It provides the foundation security teams need to protect proprietary information in an environment defined by AI, continuous movement, and accelerating risk. DSPM sits at the center of this new reality, connecting visibility, understanding, and enforcement across all data and environments.

Section 04

Turning Visibility into Control: What DSPM Must Do



Modern data security depends on more than discovery or classification alone. Next-generation DSPM transforms visibility into actionable insight, enabling teams to understand, manage, and protect sensitive data across increasingly complex environments. Understanding these capabilities is critical for evaluating DSPM solutions and ensuring alignment with your organization's risk posture.

01 Data Discovery: Finding Sensitive Data Everywhere

The foundation of DSPM is knowing where your data resides. Modern platforms connect to every data source across the organization, including:

- Cloud infrastructure (AWS, Azure, GCP)
- SaaS applications (collaboration tools, CRM, ticketing systems)
- On-prem databases and file shares
- Employee endpoints
- Generative AI tools used by employees

Discovery must be continuous. Scheduled scans cannot keep pace with the rapid creation, replication, and transformation of data. Continuous discovery ensures that newly created data, changes to existing data, and shadow copies are detected as soon as they appear, reducing exposure before it becomes a risk.

02 Data Classification: Understanding What the Data Is

Knowing where data lives is only the first step. Modern DSPM platforms classify data with precision, moving beyond pattern-based or rule-heavy approaches. AI-driven classification enables:

- Higher accuracy with fewer false positives
- Semantic understanding beyond simple regex rules
- Sensitivity assessment in context of business use, not just content patterns

This capability is especially critical for unstructured data and AI-generated content, where sensitivity is determined by use, not form.

The Value of AI-Driven Data Understanding

AI-Driven Data Understanding moves DSPM beyond static labels and fragmented discovery. It combines deep semantic analysis with continuous context and lineage so you can:

- **See what data truly is, not just where it sits.** AI analyzes content and context to improve accuracy and reduce false positives.
- **Distinguish meaningful risk from benign noise.** Provenance and access history help separate internal corporate data from public or low-risk information.
- **Understand how data changes and moves.** Contextual lineage reveals how sensitive information evolves as it flows across endpoints, SaaS, cloud, and AI workflows.
- **Inform smarter protection and prioritization.** Rich context enables policies that align with business risk, not just pattern matches.

03 Contextual Data Understanding: Beyond Labels

Classification alone cannot reveal risk. Modern DSPM enriches data with context, helping teams understand:

- **Provenance:** Was data internally created or sourced externally?
- **Exposure:** Who can access it—internal users, external collaborators, or the public?
- **Location:** Endpoint, SaaS, cloud storage, or on-prem systems
- **Structure:** Document, spreadsheet, database record, or raw text
- **Management status:** Whether the system holding the data is managed or unmanaged

Context lets DSPM differentiate risk between seemingly identical data. For example, an internal document on a managed laptop poses far less risk than the same document publicly shared from a SaaS platform

04 Data Lineage: Tracking How Data Moves and Transforms

Lineage tracks data throughout its lifecycle, showing how it moves and changes across systems. For example, sensitive data might be:

- Created on an employee endpoint
- Uploaded to a collaboration platform
- Exported into cloud storage
- Copied into documents or spreadsheets
- Fed into AI tools that generate derivatives

Without lineage, these actions appear as disconnected events. With lineage, DSPM reveals hidden risk paths, shadow copies, and downstream exposure that static tools cannot detect.

Effortless Experience: DSPM Should Fuel Innovation, Not Hinder It

Modern DSPM provides visibility while enabling teams to protect data comfortably. By unifying discovery, classification, and enforcement into a single workflow, teams can:

- **See and act on risk instantly.** One platform, one view, no fragmented dashboards.
- **Reduce noise and false positives.** Context-rich AI insight highlights what truly matters.
- **Move faster with confidence.** Automated policies and continuous monitoring let teams focus on decisions, not busywork.

Simplify operations at scale. Unified controls mean fewer tools, less manual effort, and consistent enforcement across endpoints, SaaS, cloud, and AI workflows.

Outcome: Security teams gain clarity, confidence, and control without the complexity and friction of legacy systems.

05 Data Risk Assessment and Prioritization

Effective DSPM continuously evaluates risk across multiple dimensions:

- Public exposure or misconfigured access
- Overly permissive entitlements
- Cross-border data transfers
- Dormant or orphaned sensitive data
- Risky movement patterns

Instead of generating thousands of alerts, modern DSPM correlates sensitivity, access, exposure, and usage to prioritize the issues that matter most. This helps teams focus on actionable risk instead of drowning in dashboards.

06 DSPM and Generative AI: Protecting Data in AI Workflows

AI fundamentally changes how data risk manifests. Employees interact with AI to summarize, analyze, and generate content, producing new derivatives and sharing sensitive data outside controlled systems. Next-generation DSPM provides:

- Detection of sensitive data being fed into AI tools
- Tracking of AI-generated outputs across environments
- Visibility into AI workflows that create unacceptable risk
- Enforcement based on data sensitivity and context

Without DSPM visibility into AI-driven data flows, organizations are blind to one of the fastest-growing sources of exposure.

07 Identity and Access: Understanding Who Can Touch the Data

Modern DSPM integrates identity, access, and data into a unified model, linking datasets, datastores, human identities, service accounts, and AI-driven entities. This transforms posture from a static view of "where sensitive data lives" into a dynamic understanding of "who can interact with it and under what conditions."

This enables organizations to answer critical risk questions out of the box, including:

- Which sensitive datasets are accessible to large groups of users or the entire organization?
- What customer or regulated data can contractors or external collaborators access?
- Which sensitive data is exposed to non-human identities, such as service accounts or AI agents?
- If a specific identity is compromised, what data could it reach?

By mapping access relationships directly to sensitive data, DSPM surfaces identities with excessive or risky permissions and allows teams to drill into access patterns, usage behavior, and downstream exposure. This eliminates the need to rely on fragmented IAM tools or manual analysis to understand blast radius.

Core Capabilities of DSPM

Capability	What It Does	Why It Matters
1. **Continuous Data Discovery**	Finds sensitive data wherever it lives: cloud, SaaS, endpoints, on-prem, and AI tools	Reduces blind spots and shadow copies before they become risk
2. **AI-Driven Data Classification**	Identifies sensitive and regulated data using semantic and contextual analysis	Improves accuracy, reduces false positives, and captures fragmented/ unstructured data
3. **Contextual Data Understanding**	Enriches data with provenance, exposure, location, structure, and system status	Differentiates risk between similar data and informs prioritization
4. **Data Lineage**	Tracks how data moves and transforms across systems, workflows, and AI outputs	Reveals hidden exposure paths and downstream risk invisible to traditional tools
5. **Risk Assessment & Prioritization**	Continuously evaluates exposure, access, entitlements, and movement patterns	Focuses teams on actionable issues instead of overwhelming dashboards
6. **AI-Aware Data Protection**	Monitors sensitive data usage in AI workflows and tracks AI-generated derivatives	Prevents high-risk AI interactions and maintains control over emerging data flows
7. **Identity & Access Context**	Maps who can access sensitive data across human identities, service accounts, and AI agents.	Shows who can touch data and how broadly, enabling accurate prioritization and faster breach impact analysis.
8. **Stopping Data Loss**	Take action, automatically, through natively integrated DLP capabilities.	Turns visibility into action by stopping data exfiltration at the source.

Data Lineage: See Every Move Your Data Makes

Cyberhaven pioneered data lineage to map data at its origin, not just its location. Every move, copy, edit, or share is captured, creating a complete map of how sensitive information flows across your enterprise.

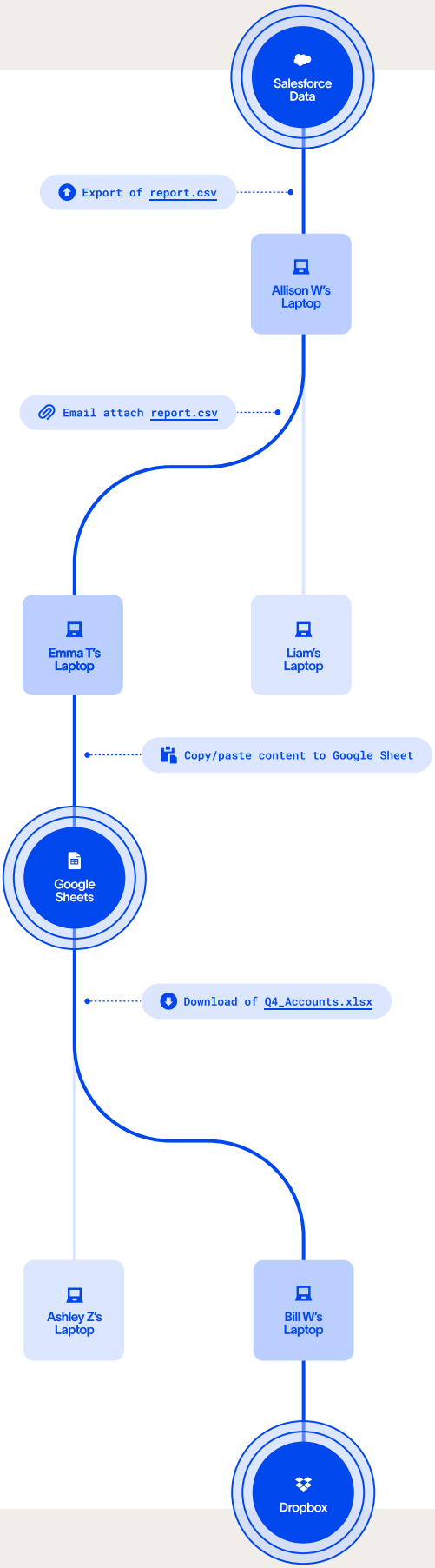
Data lineage powers three critical outcomes:

- **Accurate classification:** Context informs what is truly sensitive and reduces false positives.
- **Actionable insight:** Teams can enforce policies across endpoints, SaaS applications, and cloud environments.
- **Risk visibility:** Shadow copies, AI derivatives, and cross-system flows are all tracked in real time.

Unlike DSPM solutions built on static scans, Cyberhaven's lineage-first approach provides continuous, real-time insight. The result is a DSPM that sees more, understands better, and acts faster. This enables organizations to secure sensitive data confidently, even in AI-driven environments.

Why Lineage Matters:

- Detects hidden data exposure paths
- Feeds AI-driven classification with real context
- Enables risk-based prioritization and enforcement at machine speed



Section 05

The Cyberhaven Difference

The future of data security demands more than visibility. It requires control, context, and the ability to act at the speed of modern work. Cyberhaven approaches this challenge by unifying discovery, classification, enforcement, and AI-aware protection into a single platform. By addressing both the complexity of data and the pace of AI-driven workflows, Cyberhaven empowers security teams to reduce risk without adding operational friction.

How We Solve Modern Data Security Challenges

Data Visibility

Challenges: Security teams often lack a clear understanding of where data resides, with only surface-level context and outdated snapshots.

Cyberhaven Approach: Continuous discovery, AI-driven classification, a comprehensive data catalog, and rich contextual insights (including data lineage) give teams a real-time, complete view of their information ecosystem.

Data Risk Management

Challenges: Sensitive data is scattered, duplicated, or unmanaged, increasing exposure and insider risk.

Cyberhaven Approach: Cyberhaven combines classification, sensitivity labeling, and a centralized data catalog to identify, prioritize, and reduce high-risk data. AI-powered insights help analysts act quickly and confidently.

Data Privacy & Compliance

Challenges: Organizations struggle to locate regulated data, enforce retention policies, and ensure compliance across complex environments.

Cyberhaven Approach: Classification, regulatory labeling, and the data catalog support compliance requirements, while "data minimization" techniques reduce unnecessary exposure and enforce the principle of least privilege.

A Unified Platform for the AI Era

Cyberhaven combines DSPM, DLP, Insider Risk Management, and AI Security into a single, unified platform. By integrating these capabilities, the platform eliminates the fragmentation that slows teams and creates blind spots, giving security leaders a complete view of sensitive data wherever it lives and moves.

Key Capabilities of the Cyberhaven AI & Data Security Platform:

Data Security Posture Management (DSPM): Continuously discover and classify sensitive data across endpoints, cloud, SaaS, and AI workflows. Detect risk in motion and automatically enforce protection.

Data Loss Prevention (DLP): Stop leaks before they occur. Coach users, block risky actions, and protect data across email, web, cloud, and endpoints.

Insider Risk Management (IRM): Combine data and behavior signals to identify insider threats, clarify intent, and detect slow-burning risks before they escalate.

AI Security: Secure AI adoption with visibility into shadow AI, assessment of AI risk posture, and enforcement that prevents leaks without slowing innovation.

Cyberhaven unifies visibility, understanding, and enforcement, enabling organizations to manage sensitive data continuously, reduce operational complexity, and protect proprietary information at the pace of AI-driven workflows.

Request a demo