WHITEPAPER

Powering Up Purview with Cyera

A Practitioner's Guide to the Multi-Cloud Data Security Journey in the AI Era

cyera | Microsoft

Executive Summary

Microsoft Copilot is moving into the M365 tenant faster than most data estates are ready for it. Closing that gap starts with a simple premise: Microsoft Purview remains the enterprise standard for compliance, policy configuration, and core data estate management. However, the scale of the data underneath it has changed. Modern data volume, velocity, and multi-cloud sprawl - now amplified by AI-driven retrieval and use - has outpaced what native classification and labeling were built to handle. Purview was designed for a more manual compliance cadence than AI deployment now demands.

That's where Cyera fits in. Cyera operates as a complementary data intelligence layer combining Data Security Posture Management (DSPM) and Data Loss Prevention (DLP) alongside Purview: automating initial discovery, sharpening classifier precision, and extending governance to the clouds and SaaS platforms Purview doesn't cover by design. Together, they replace manual, one-time compliance projects with an automated, continuously synchronized security pipeline - built for the Day-2 reality of live AI deployments, not a single configuration exercise.

That shift matters because AI adoption has turned data hygiene from a background compliance task into a frontline security priority. Before AI, a mislabeled asset or an over-permissioned directory was an audit finding and something to fix on the next review cycle. With Microsoft Copilot and enterprise agents now inside the tenant, that same gap becomes an immediate, natural-language path to data exposure, accessible to anyone who knows how to ask. The numbers make the urgency concrete:

- 97% of sensitive M365 data carries no label.
- 83% of existing sensitivity labels are incorrect or incomplete.
- 40% of file types can't receive a label at all.

Left ungoverned, these gaps sit directly in Copilot's path. Closed, they clear the way for a faster, more confident Copilot rollout, one that Cyera and Purview reach considerably faster working together, with up to 70–80% fewer false-positive DLP alerts along the way.

The Production Environment Reality (The Technical Landscape)

Before any AI rollout can be called safe, look at the production environment as it actually exists: the sprawling, multi-cloud, partly-dark estate most enterprises actually run. The following realities define that landscape.

Uncover the Data Proliferation Baseline

Dark and unstructured data across enterprise storage systems doesn't surface through standard indexing. Finding it requires purpose-built discovery.

Cyera Research, 2026: 435,080 orphaned M365 datastores including abandoned OneDrive accounts, dissolved Teams sites, and archived project folders. These datastores hold 44 billion unmitigated records. This is dark data most security teams don't know exists.

Optimize Custom Policy Management

Keyword and regular-expression classification rules drift out of date fast. They need contextual signal to stay accurate, without constant manual tuning.

Cyera Research, 2026: 335,927 files across enterprise M365 environments are missing confidential sensitivity labels, leaving 73.2 billion records unprotected. Manual classification doesn't scale; leaving labeling to user judgment produces consistent, predictable gaps.

Expand File Tagging Universality

Certain file formats such as raw logs and code blocks can't carry standard metadata labels at all. Covering them requires discovery methods built outside the native labeling path.

Cyera Research, 2026: Across the top 20 M365 policy violations, the average mitigation rate is just 13.6%, which means that 6 in every 7 policy-violating records remain at active risk. The gap isn't awareness; it's the inability to enforce at scale, across file types and formats native tools don't reach.

Extend Boundaries into Non-Microsoft Repositories

Purview's classification definitions can extend to third-party data lakes, SaaS tools, and infrastructure providers outside Microsoft's ecosystem, but only with a bridge built to reach them.

Cyera Research, 2026: 809,898 issues involve European personal data stored outside Europe, creating 22.6 billion unmitigated records of direct GDPR liability. When M365 becomes a global content repository without residency controls, data crosses borders that governance programs haven't mapped.

Accelerate Secure Generative AI Deployments

Preparing a data environment for M365 Copilot means auditing data permissions and exposure before the engine ever ingests content, because retrieval-based AI surfaces sensitive material without the user needing to know it exists.

The AI Amplification Effect: Why the Stakes Are Higher

AI adoption accelerates and exposes data security problems that already exist.
Three structural shifts define what that means operationally.

A. Discovery Risk → Retrieval Risk

Dark and unstructured data across enterprise storage systems doesn't surface through standard indexing. Finding it requires purpose-built discovery.

Cyera Research, 2026: 388,549 violations involve external organizations accessing sensitive files; 178,224 more involve org-wide access to sensitive data. That's 566,000+ instances of overpermissioning. Copilot would treat every one as authorized access..This risk isn't theoretical: the permissions to retrieve that data already exist.

B. Point-in-Time Governance → Continuous Drift

AI rollouts don't end at deployment. New Teams sites, SharePoint sharing links, data owners, and file types appear continuously after launch, and a clean baseline on Day 1 can degrade within weeks. Implication: data security governance has to operate as a continuous Day-2 function, not a one-time Purview configuration project. An architecture that can't detect and respond to drift in near-real time isn't built for an AI-enabled environment.

Cyera Research, 2026: The same 13.6% average mitigation rate tells the continuous-drift story too: point-in-time audits found these violations, but the backlog grew faster than teams could remediate. Governance programs that run on audit cycles fall behind because new violations appear faster than quarterly or monthly reviews can catch them.

C. DLP Volume → DLP Signal Quality

AI multiplies the pathways sensitive data can move through: summarization outputs, excerpts inserted into prompts, agent-triggered file reads, copy-paste into external interfaces. More events mean more noise, not better security. Implication: the priority has to shift from alert volume to alert quality, surfacing only the events where data is simultaneously sensitive, exposed, and accessible to the wrong actor. Risk-based prioritization is the metric that matters, not raw detection counts.

D. The Complementary Division of Labor

Microsoft Purview: The system of record for policy configuration, compliance reporting, and endpoint enforcement.

Cyera: The continuous, autonomous discovery engine and AI-readiness monitor, operating persistently to maintain the classification signal AI deployments need.

The Data Security Maturity Roadmap (The Phased Journey)

The journey from a stalled or partial Purview deployment to a continuously governed, AI-ready M365 estate runs through four phases. Each builds on the one before it, moving an organization from a static inventory to an operationalized, self-correcting security pipeline.

Phase 1: Assessing and Discovering the Environment

Step 1

Agentless Cloud Integration

API-driven connectivity scans the environment without disrupting active systems or altering routing configurations.

Step 2

Automated Data Mapping

Zero-configuration indexing catalogs assets and feeds directly into Purview's existing classification taxonomy, locating unencrypted records at rest, evaluating nested group security, and mapping identities and access controls.

Step 3

Verification of Sharing and Access Configurations

Auditing active public and organizational sharing links confirms that data exposures align with corporate governance guidelines.

Phase 2: Aligning to Baseline and Sanitizing for AI

Step 4

Machine-Learning Data Classification

Unsupervised data modeling organizes unstructured text automatically, feeding accurate classifications straight back into Purview's definitions.

Step 5

Pre-Flight Data Cleansing for Enterprise LLMs

Pre-indexing sanitization resolves over-permissioned directories before Microsoft Copilot goes live.

Phase 3: Operationalizing and Integrating the Engine

In this phase, Purview and Cyera operate side by side on the same milestones. Each step below shows what Purview enforces natively, what Cyera adds, and the impact of running both together.

Step 6

Sensitivity Label Orchestration

Purview: Executes policy rules, applies native encryption, and enforces downstream endpoint data restrictions.

Cyera: Autonomously discovers and categorizes assets; writes classification findings directly to MIP via Microsoft Graph API.

Impact: Replaces manual tuning with automated, API-driven asset tagging to accelerate label deployment, and keeps labels accurate as the environment evolves after AI rollout.

Step 7

Signal Noise & DLP Tuning

Purview: Monitors known communication channels and flags unauthorized data movement based on established rule sets.

Cyera: Automates DLP investigation by enriching alerts with identity and behavioral context; surfaces only events where data is sensitive, exposed, and accessible.

Impact: Shifts DLP from volume management to signal quality, reducing alert noise so SecOps can focus on validated, risk-ranked threats. This is critical when AI multiplies event volume.

Step 8

Access Risk & Sharing Remediation

Purview: Maintains centralized compliance registries, identity configurations, and group structure definitions.

Cyera: Tracks anomalies in public links and communicates remediation steps directly to business data owners.

Impact: Moves data triage out of the SOC and into the hands of business owners, cutting time-to-remediation with single-click actions.

Step 9

Data Minimization & Life Cycle Execution

Purview: Provides data retention tags, archiving templates, and defensible disposal scheduling.

Cyera: Scans for and isolates redundant, obsolete, and trivial (ROT) data across multi-cloud infrastructure.

Impact: Applies Purview's lifecycle rules in bulk to continuously shrink the organization's attack surface.

Phase 4: Scaling Across Cloud and Expanding the Ecosystem

Step 10

Governance Extension to Non-Microsoft Environments

The enterprise's core Purview security principles extend to cloud buckets, non-Microsoft databases, and external SaaS platforms.

Step 11

Enterprise Incident Orchestration and Automation

Enriched data event details route directly into central IT service desks and SOAR platforms for unified security operations.

Production Validation: Case Study Analysis

The Operational Opportunity	A Microsoft Copilot rollout at a telecommunications enterprise stalled at the data classification layer. Manual mapping projections put the scoping effort at multiple years.
The Integrated Approach	Cyera was deployed alongside Purview's native framework to classify and organize data across the full M365 environment.
The Metrics Achieved	A 90-day engagement mapped over 200 million corporate documents and reconciled legacy records against regulatory requirements.
The Strategic Conclusion	The engagement shows how automated data discovery scales what a multi-year manual project couldn't, and how much additional value it draws out of an enterprise's existing Microsoft security investment.

Conclusion

The Combined Platform Value

Combining Purview's enforcement capabilities with Cyera's automated data intelligence delivers operational efficiency and stronger security across the M365 estate and beyond.

Acceleration of Copilot Rollout

Automated data discovery extends an enterprise's existing Microsoft security investment without adding headcount, cutting the path to a safe, compliant Copilot rollout from years to weeks.

Find out if your M365 environment is ready for Copilot. Run a free Copilot Risk Report - without installing agents or disrupting production - to surface over-permissioned data, mislabeled assets, and sharing exposures before Copilot goes live.

About Cyera

Cyera is the AI Security Platform built for the age of agents. Enterprises like Paramount, Chipotle, and Valvoline use Cyera to control exactly what data their AI can reach, and govern what happens next. The platform secures data at rest, in motion, and in use, whether touched by humans or AI agents. Valued at $12 billion and backed by over $2.3 billion from Accel, Blackstone, Cyberstarts, Georgian, Lightspeed, and Sequoia. Protect your data. Secure AI.

Learn more at www.cyera.com.