Microsoft Security

Protect your data intelligently with trainable classifiers

Table of contents

Introduction

Today, organizations across various industries are generating massive amounts of data, and the volume grows exponentially every year. With the adoption of a cloud and remote work model, data is no longer locked behind your corporate network's perimeters, but instead is spread across many nodes. Recent statistics tell us that 80 percent of business data is dark[1], which means it's unclassified and unprotected, while another study shows that 38 percent of incidents involve misuse of this data[2] (Figure 1). Because of this, detecting and protecting sensitive data is at the core of our focus.

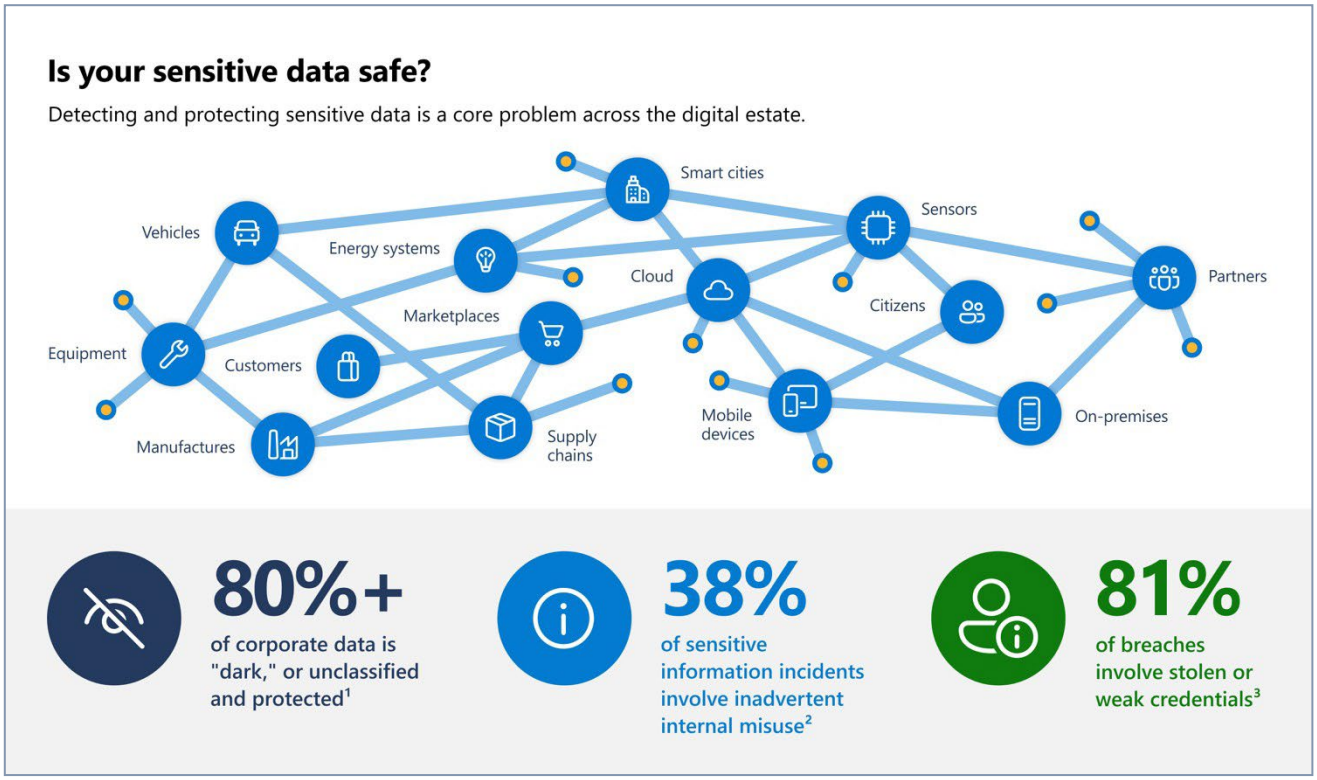

Figure 1: Data across the digital estate

[1] Andrew Trice, "The Future of Cognitive Computing," The IBM Cloud Blog, November 23, 2015.

[2] Jeff Pollard, "Security Budgets 2019: The Year Of Services Arrives," Forrester Research, December 17, 2018.

[2] Clare Ward and Nilesh Pritam, "Cyberespionage and ransomware attacks are on the increase warns the Verizon 2017 Data Breach Investigations Report," Verizon News Center, April 27, 2017.

You need a comprehensive governance plan that can help you decide what business data to protect, retain, or delete, but first you must effectively identify the data. Data classification is the process of organizing data into categories so that it can be protected and handled correctly, acting as the starting point for an information protection discipline. But no matter how large your workforce, manual and rule-based approaches to data classification cannot work effectively on their own. A better approach is to automate data classification with machine learning technology that can train a model to predict the class of new, unseen data, giving you more efficient data protection and minimizing false positives more efficiently as compared to manual approaches.

What is machine learning?

Machine learning (ML) is the process of applying mathematical concepts to data to help a computer learn. ML uses algorithms to identify patterns within data and attempts to learn these patterns to create a model. Once an ML model has been trained to find these patterns, it can be used to make predictions on unseen data. The more data the model is exposed to during the training process, the better the model will be able to perform, just as humans improve with more practice.

Manual approaches to classification cannot scale to handle the massive data that organizations have across various business functions. Instead, you can automate your data classification with Microsoft Purview Information Protection trainable classifiers, an artificial intelligence (AI)-based solution that identifies the type of content by analyzing the elements of the content itself. Our advanced classification algorithms powered by state-of- the-art intelligence can quickly adapt to changes in regulatory and dynamic business contexts. (Figure 2)

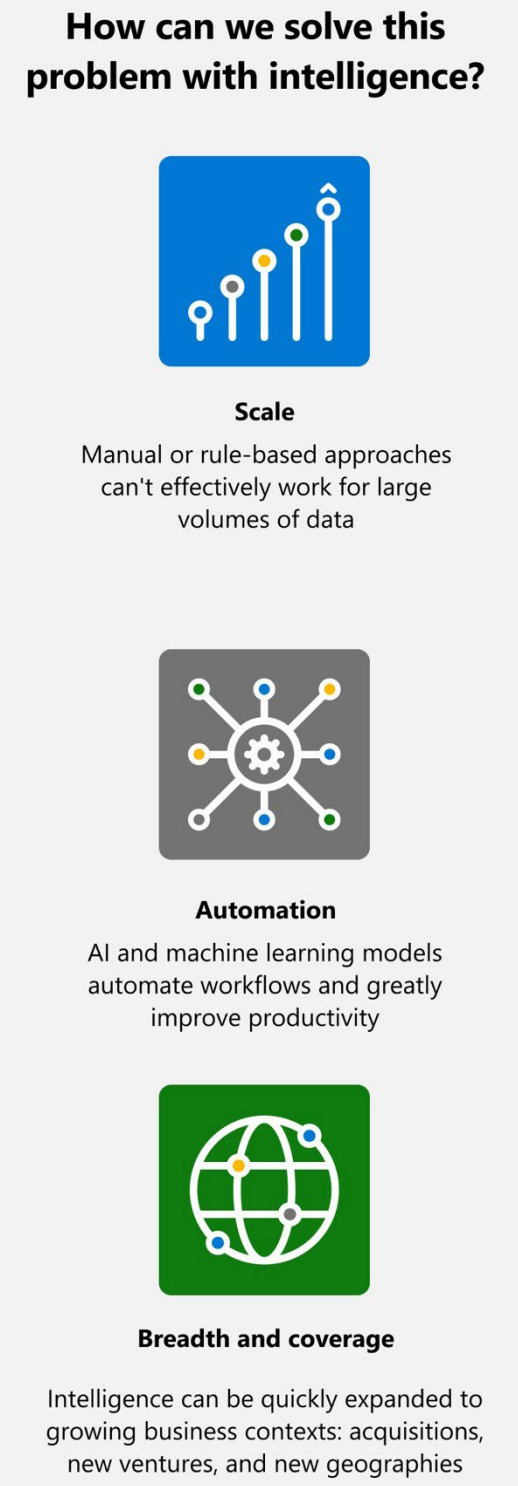

Figure 2: How AI can address business challenges in protecting sensitive data

Protect data with Microsoft Purview Information Protection trainable classifiers

A typical data map of an organization comprises data originating from multiple data sources. AI-based applications are best suited for this data geography because they can adapt to meet dynamic requirements. For example, organizations should use pre-trained, ready-to-use classifiers to discover and protect generic documents and data for common business functions like legal, human resources, sales and marketing, research and development, and finance. For proprietary, organization-specific, or market vertical-specific documents, it's best to use custom classifiers that are trained using organizations' own document examples.

For regulatory functions such as the pharmaceutical, banking, and insurance industries that have standard regulatory templates and policies, fingerprinting is best suited to discovering and protecting standard documents. Document fingerprinting enables system administrators to create a fingerprint sensitive information type (SIT) of a specific document, which can be used later to detect if the same document or part of the same document is found elsewhere in the organization. For more information, visit our Document Fingerprinting webpage. (Figure 3)

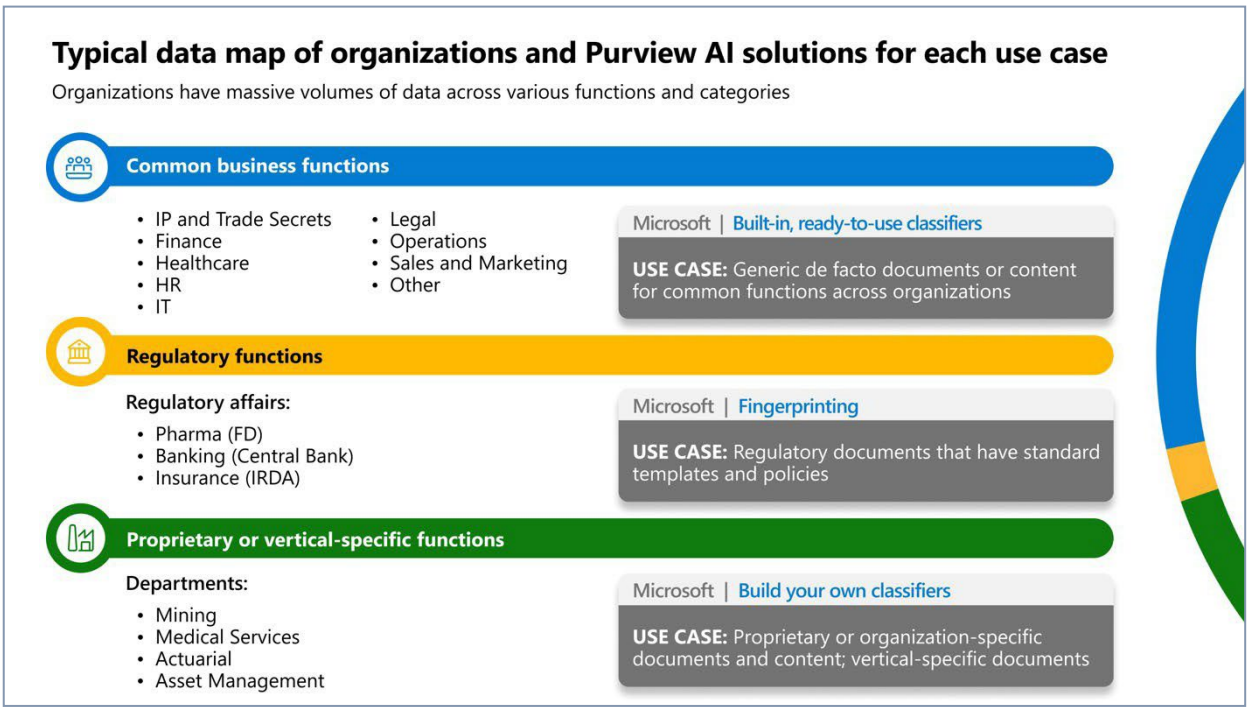

Figure 3: Data map of organizational functions and categories

Microsoft Purview Information Protection simplifies this process with a unified set of capabilities for data classification, labeling, and protection. Our solution addresses information stored in Office apps as well as other popular productivity services where information resides, such as Microsoft Teams, SharePoint Online, Exchange Online, and endpoint devices. Microsoft is focused on delivering built-in, intelligent, unified, and extensible solutions to protect sensitive data across your digital estate—in Microsoft 365 cloud services, on-premises, in third- party software as a service (SaaS) applications, and more.

Information Protection trainable classifiers help your organization manage data security and compliance needs efficiently and more easily.

- **Our out-of-the-box trainable classifiers** for sensitive business document discovery and classification have been trained using a wide sample of data to minimize false positives.
- **You can create customized trainable data classifiers** to meet your unique content labeling and categorization requirements. (Figure 4)

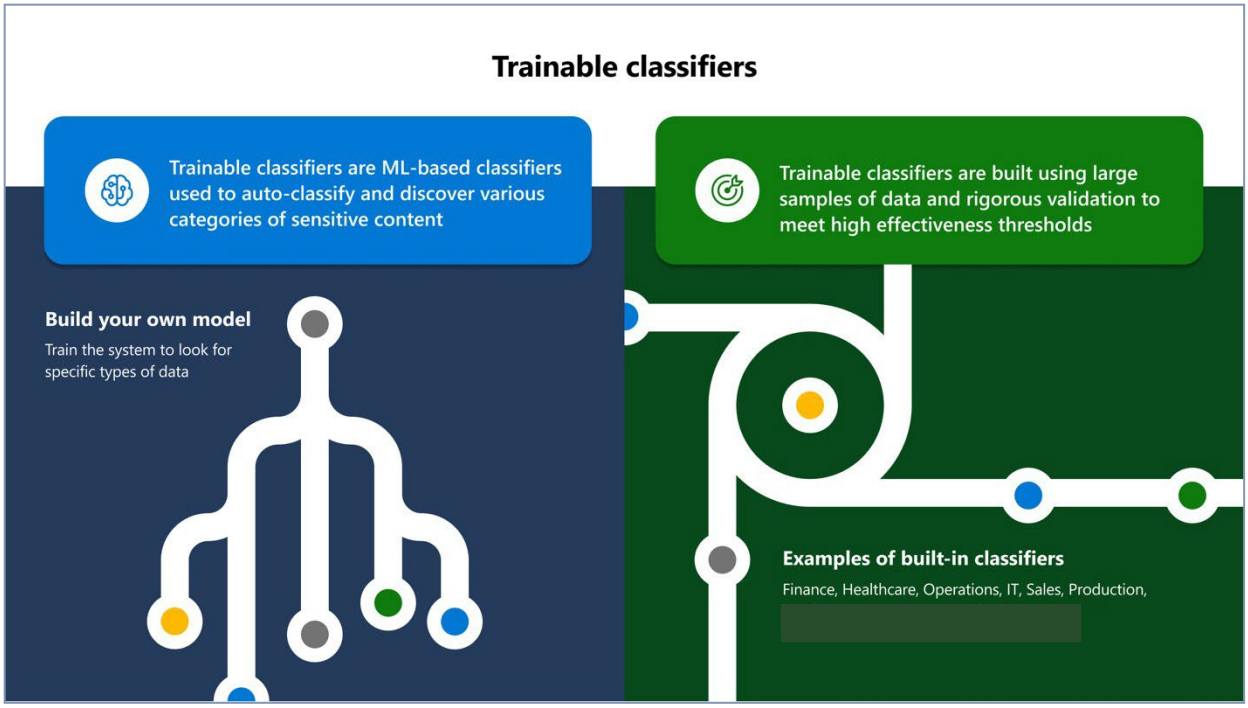

Figure 4: Types of trainable classifiers

Overcome data classification challenges

Compared to the traditional manual approach for data classification, auto-classification can help information workers stay productive by more quickly and comprehensively discovering, labeling, and protecting massive volumes of sensitive data across your organization's digital estate. Auto-classification solutions must address three key challenges faced by customers: scalability, breadth and coverage, and automation. Information Protection trainable classifiers helps you solve each one.

Enable scalability

Through our trainable classifiers, you can use the power of machine learning to identify more data categories with increased performance and quickly classify massive volumes of data. Our classifiers were built and improved with some of Microsoft's latest AI technology and have been pre-trained across a large, diverse number of real-world samples.

Improve breadth and coverage

Significantly improve the speed, performance, and coverage of sensitive data identification at an enterprise scale. We provide coverage for broad common business categories required by global enterprise customers with our ready-to-use, optimized classifiers.

Employ automation

Customers need a solution to work behind the scenes automatically and adapt to ever-changing business needs and regulatory context. Our trainable classifiers can also be used for auto-labeling policies to automatically label and protect sensitive data in key business categories. They're also fully integrated with different Microsoft compliance solutions, such as information protection, data loss prevention, and data lifecycle management, that can help your organization effectively respond to and protect against unauthorized access. (Figure 5)

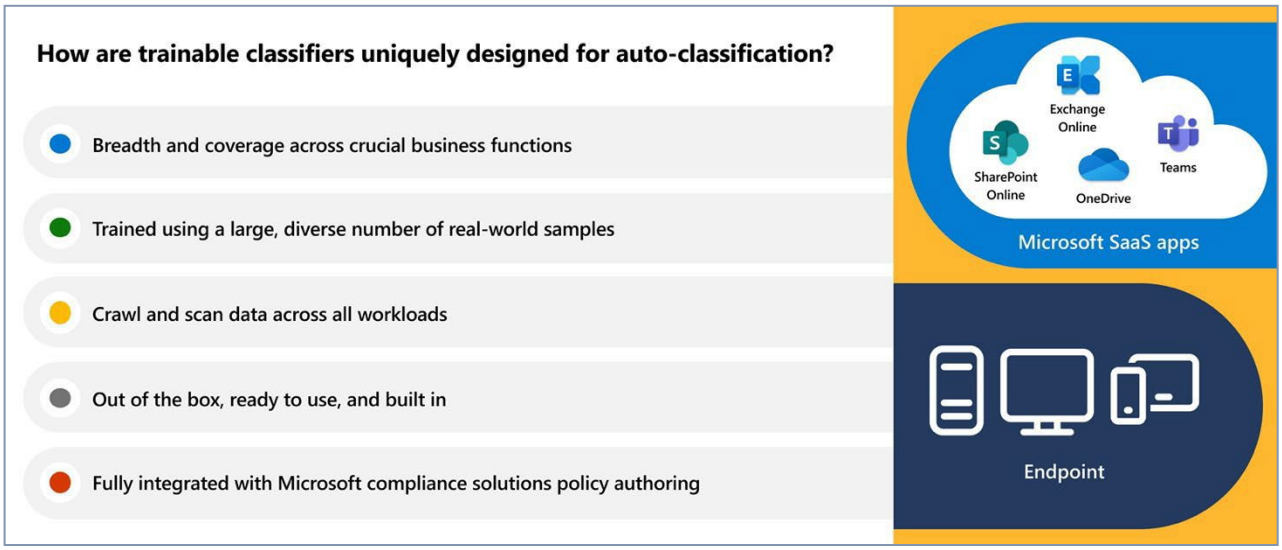

Figure 5: Value propositions that makes trainable classifiers stand out

Merge automation with subject matter expertise

Subject matter experts (SME) define business concepts, such as information protection considerations, regulatory context, and organizational policies, which are infused into Information Protection as human knowledge to create robust ML models for data classification. This improves the effectiveness of the models, thereby reducing false positives and maximizing recall, so that sensitive data are not left unprotected. (Figure 6).

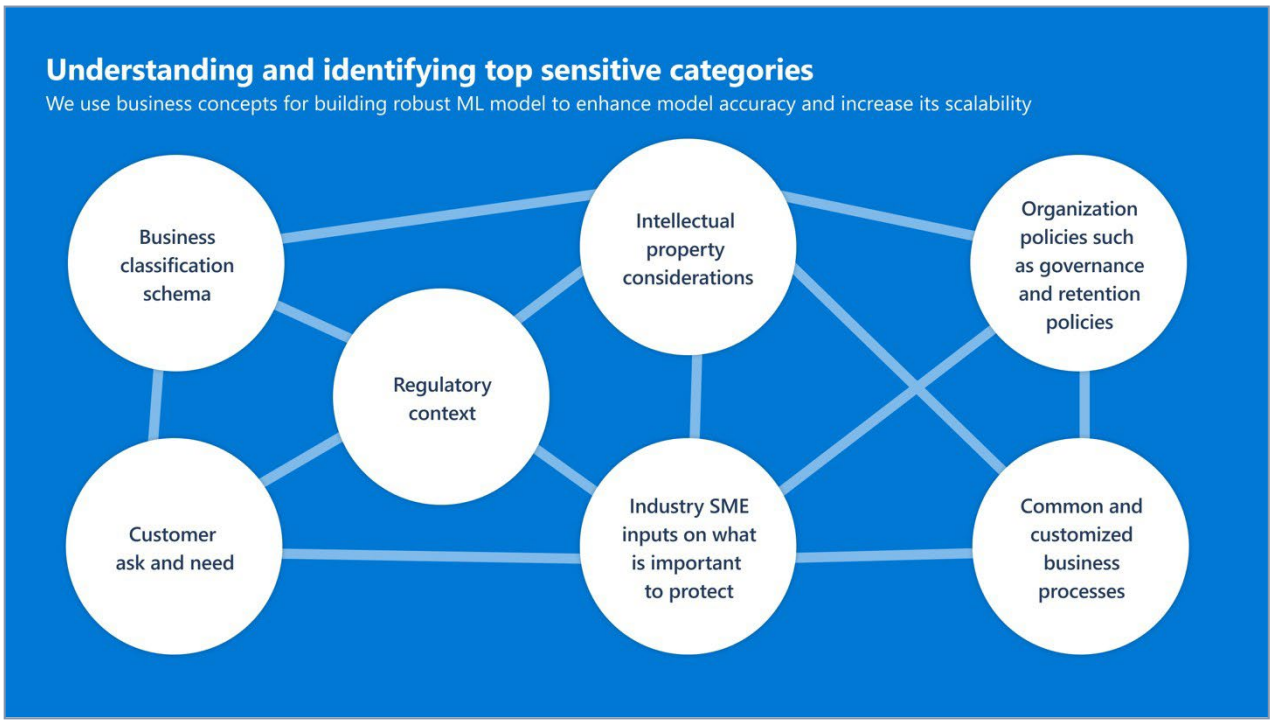

Figure 6: Understanding sensitive business content categories

During the model building phase, SMEs gather a list of business concepts relevant to each of the business categories for training the classifier. In the healthcare industry, for example, the model learns that information like patient name, address, and diagnosis are generally associated with patient health forms, and so it begins looking for this data in documents. The model classifies target content based on the probability of concept abundance and concept diversity. Once the model is trained, our data science team conducts rigorous peer reviews and tests the classifier's quality and performance. We ensure our models possess low latency and support high throughput analysis, enabling them to perform auto-classification quickly for a large volume of data. Finally, we solicit qualitative and quantitative feedback from customers in a private review and further refine the classifiers before making out-of-the-box models generally available to all customers. (Figure 7)

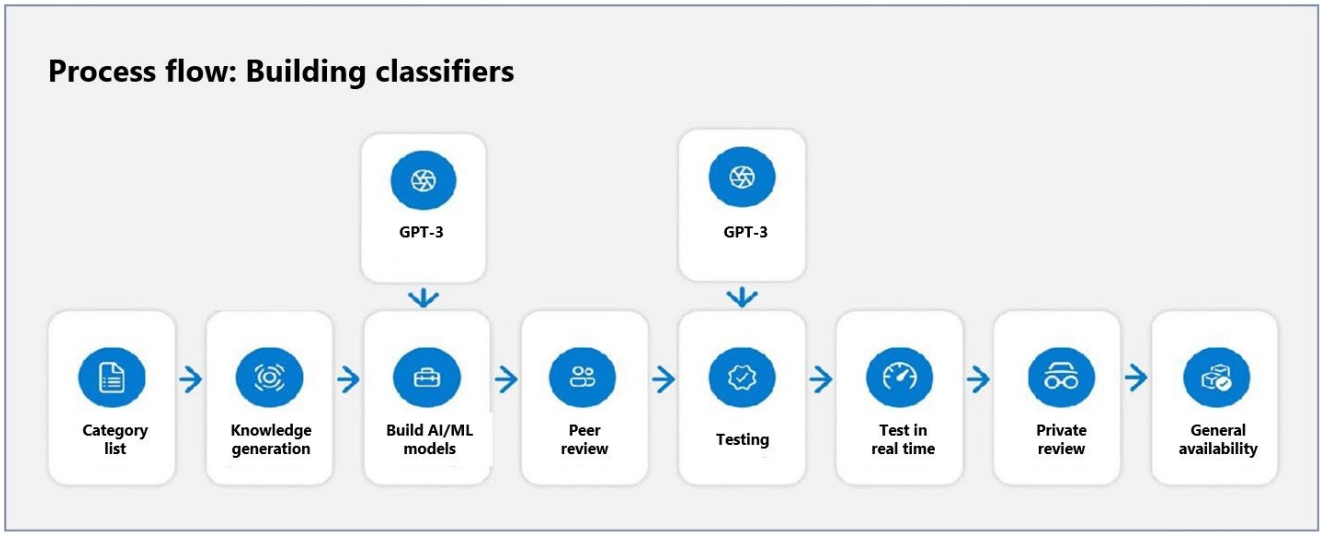

Figure 7: How we create trainable classifiers

To build and test Information Protection classifiers, we employ multiple platforms and frameworks from PyTorch, Hugging Face, GPT-3, scikit-learn, OpenAI, DeepSpeed, and vNext Technologies, as well as models from Microsoft's Project Turing. We also use many types of large language models as part of offline data generation and validation of model performance.

Mitigate risk with models tested for optimized performance

We provide classifiers with reduced false positives while maximizing recall to create the most performant models. We fine tune the hyperparameters to optimize performance in multiple ways.

- We develop models to detect diverse content by training on diverse samples and integrate enough positive and negative indicators to represent real-world documents that appear similar but have small differences.
- To detect bigger documents with large volumes of mixed content, we perform normalization of documents to ensure business concepts learn the appropriate weights during model training.
- To remove noise, we perform extensive error analysis and human reviews.

Once done, we run our models first on public data, next on synthetic data generated from GPT-3 and other large language models, and then on Microsoft's proprietary data. Finally, we validate them on data from our design partners and release them in stages for private preview with enrolled customers, then for public preview, and last for general availability. Before making them generally available, we address all performance issues at each stage to ensure we ship the best quality models to our customers.

Deploy ready-to-use trainable classifiers

Microsoft has created and pre-trained multiple classifiers that can help increase the coverage and performance of data classification while reducing false positives. Over 46 pre-trained, ready-to-use trainable classifiers can identify more than 100 categories of sensitive content. They are currently available in English, with common and critical categories soon to be released in other global languages, allowing you to scan generic de facto documents with content from everyday organizational functions—including finance, IT, intellectual property and trade secrets, legal, healthcare, human resources, and operations. These classifiers are pre-trained using a diverse number of real-world samples. This ensures they provide broad coverage of various types of business functions, and are: used to discover and automatically label files and documents, used to apply retention labels for records management, used as conditions in data loss prevention policies, and used to monitor inappropriate content for communication compliance These classifiers also can be used to discover and classify standard types of sensitive data found across the following business functions shown in Figure 8.

Figure 8: Categories of built-in, ready-to-use trainable classifiers

Create custom trainable classifiers

If you have sensitive proprietary, organization-specific, or vertical-specific content that requires identification and categorization beyond the pre-trained classifiers, visit Microsoft's Get started with trainable classifiers page to learn the prerequisites for creating your own trainable classifiers. The process is as simple as giving the classifier human-picked samples that positively match the category for which you're training, and then testing the classifier's prediction ability by using a mix of positive and negative samples.

Explore Microsoft Purview Information Protection

By efficiently categorizing and labeling content, Information Protection enables organizations to protect sensitive data across multiple fronts.

Know your data

The content explorer capability makes discovering your sensitive information easier with a current snapshot of items that have sensitivity or retention labels or have been classified as a sensitive information type in your organization. Content explorer helps you better know your data by:

- Giving visibility into the amount and types of sensitive data and allowing users to filter by label or sensitivity type for a detailed view of locations where the sensitive data is stored.
- Providing administrators with the ability to index sensitive documents stored within supported Microsoft 365 workloads and identify sensitive information being stored.
- Identifying documents classified with sensitivity and retention labels.
- Discovering and displaying categories of sensitive content matching trainable classifiers and specific files containing sensitive data in Teams, SharePoint, OneDrive, and Exchange Online. (Figure 9)

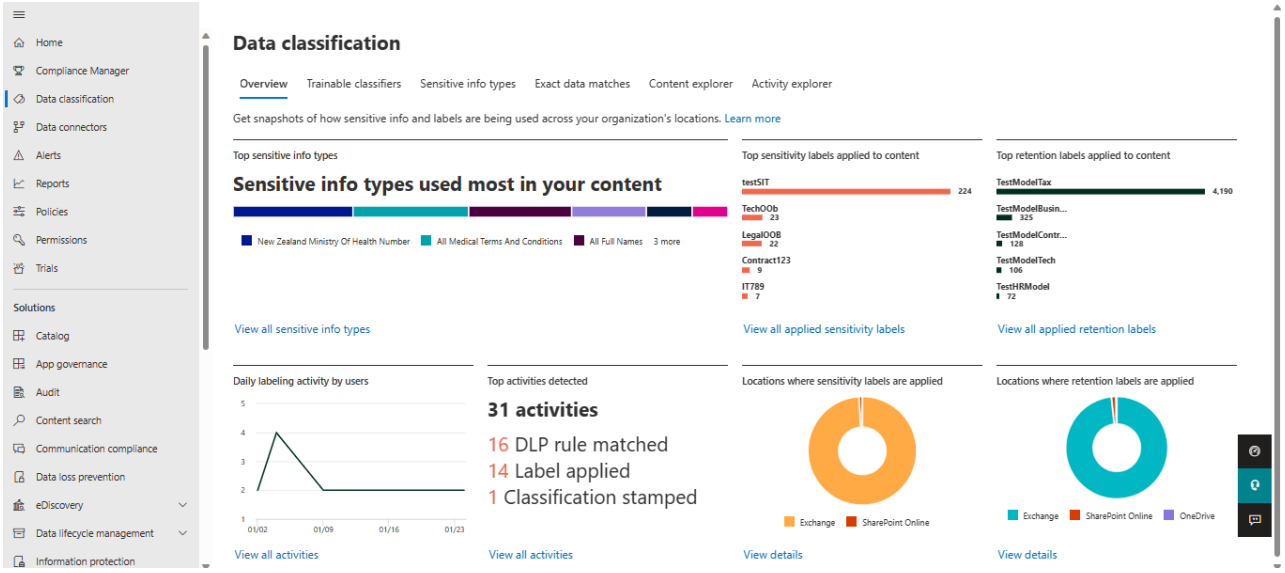

Figure 9: The Data classification Overview page

View trainable classifiers

Categorizing and labeling content so it can be protected and handled properly is the starting place for the information protection discipline. In the Trainable classifiers tab, you can see pre-built trainable classifiers or create your own classifier (Figure 10).

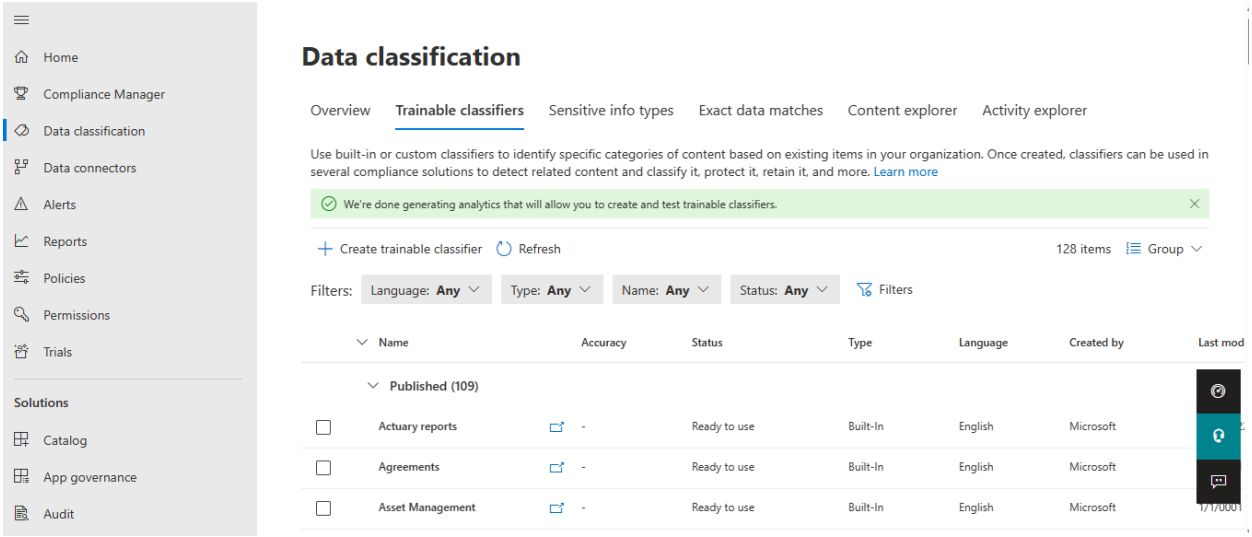

Figure 10: The Data classification Trainable classifiers page

View tagged content

In the Content explorer tab, you can view source content labeled by the trainable classifiers by expanding Trainable Classifiers in the filters panel. The filter will automatically display the number of incidents found in SharePoint, Teams, and OneDrive, without requiring any labeling (Figure 11).

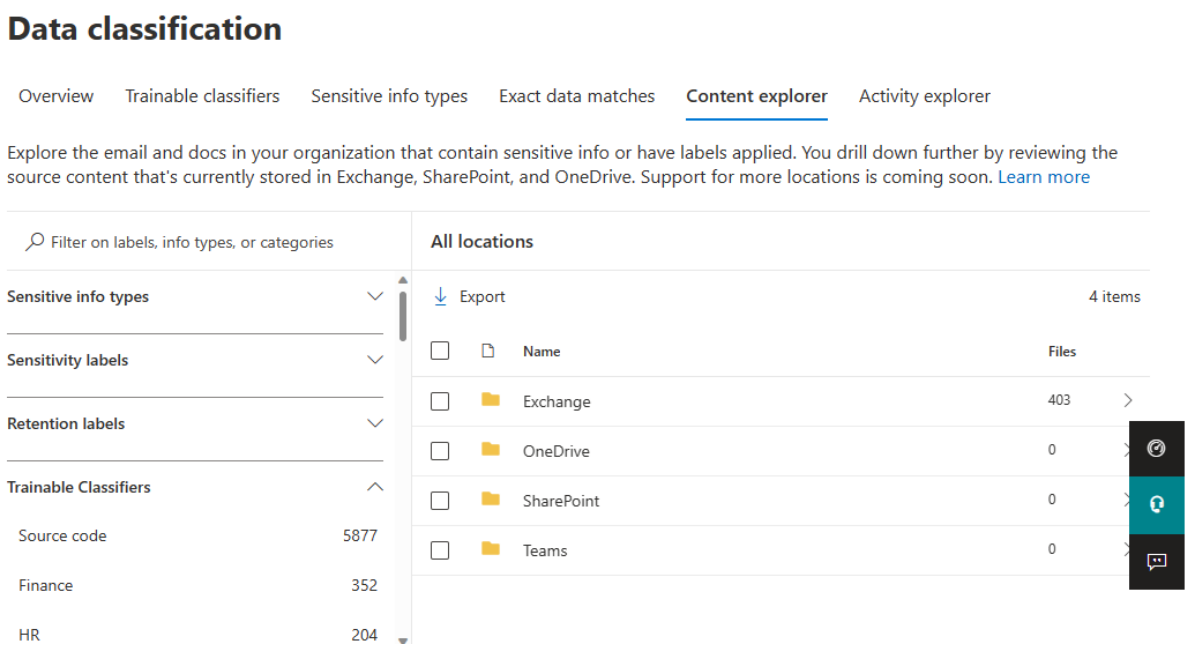

Figure 11: The Data classification Content explorer page

Employ sensitivity auto-labeling

Information Protection can use trainable classifiers in server-side auto-labeling policies for Microsoft SharePoint, OneDrive, and Exchange. You can now take advantage of this capability to more quickly and comprehensively discover, label, and protect massive volumes of sensitive data across your digital estate with pre-trained models optimized for performance and scalability. The screenshot in Figure 12 shows how to add trainable classifiers in auto-labeling for files and emails.

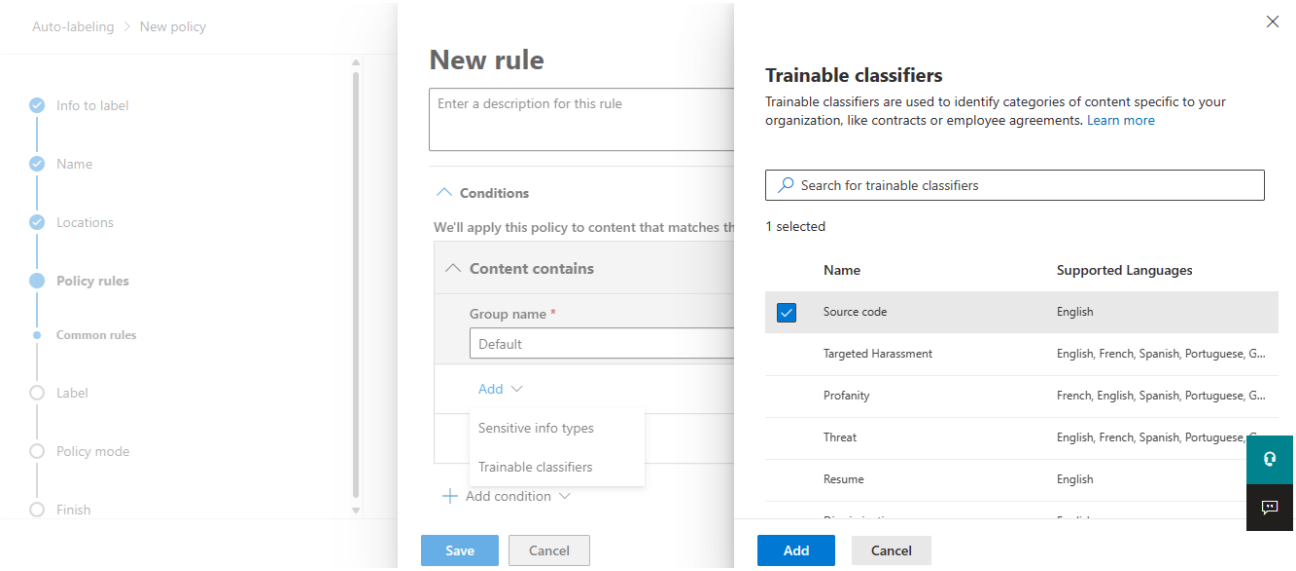

Figure 12: Creating labels with trainable classifiers

Data loss prevention

To help protect sensitive data and reduce the risk of data loss, organizations can use Microsoft Purview Data Loss Prevention (DLP) to prevent unauthorized data sharing and data exfiltration. Our DLP solution now supports all advanced classifiers, including trainable classifiers, on various DLP workloads such as SharePoint, OneDrive, Teams, Exchange, and endpoint devices. Visit our Learn about data loss prevention webpage to learn more about which applications and services, workloads, platforms and endpoints, on-premises file shares, and non-Microsoft applications are supported by our DLP solution.

With our DLP tools, you can:

- Efficiently monitor the activities users take on sensitive items at rest, sensitive items in transit, or sensitive items in use and take protective actions.
- Enable system administrators to create DLP rules specific to a category, such as Health, and designate specific actions when a DLP rule matches specific content or files, like sending incident reports and alerts to system administrators. (Figure 13)

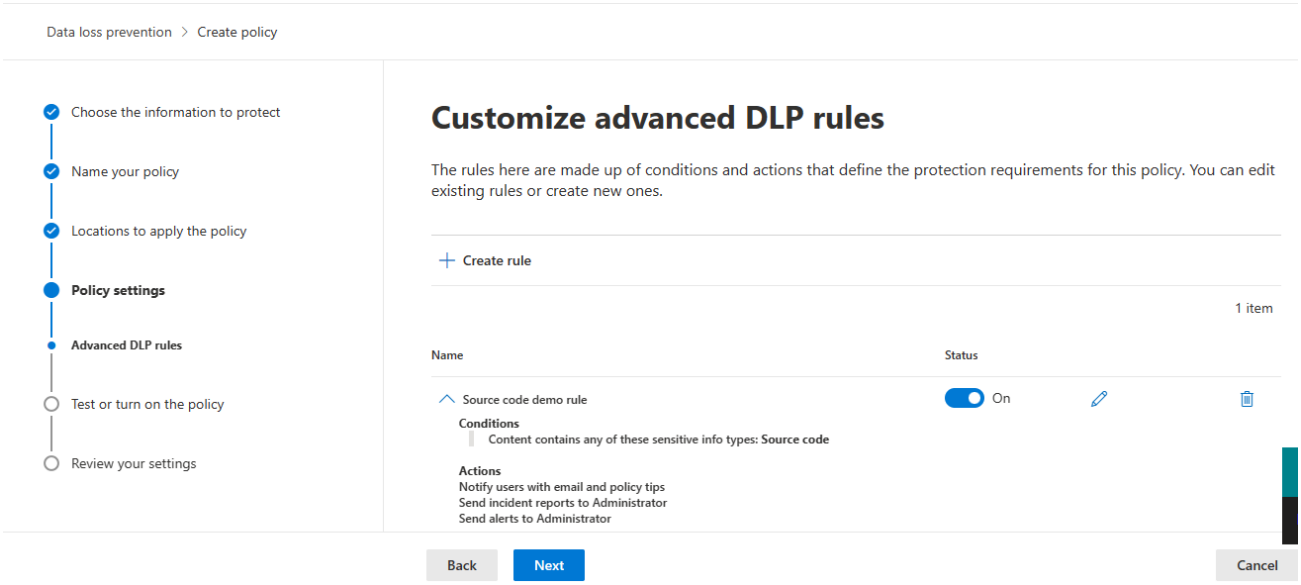

Figure 13: Customize advanced DLP rules

Data lifecycle management

Microsoft Purview Data Lifecycle Management helps you govern your data and meet your legal, business, privacy, and regulatory content obligations. Retaining and deleting content is often needed to meet compliance and regulatory requirements. Retaining high-risk or high-value content protects it from malicious deletion or ransomware, while enforcing deletion of data without business value reduces the risk during a breach. Data Lifecycle Management provides tools and capabilities to help you retain the content you need and delete the content you don't. A key differentiator of our capabilities is that all of them happen in-place, reducing the need for multiple copies and information silos. Our DLM solution supports creating retention policies, as shown in Figure 14, and supports SharePoint, OneDrive, Teams, and Exchange workloads. Visit our [Microsoft Purview Data Lifecycle Management] webpage to get more information on our DLM solution.

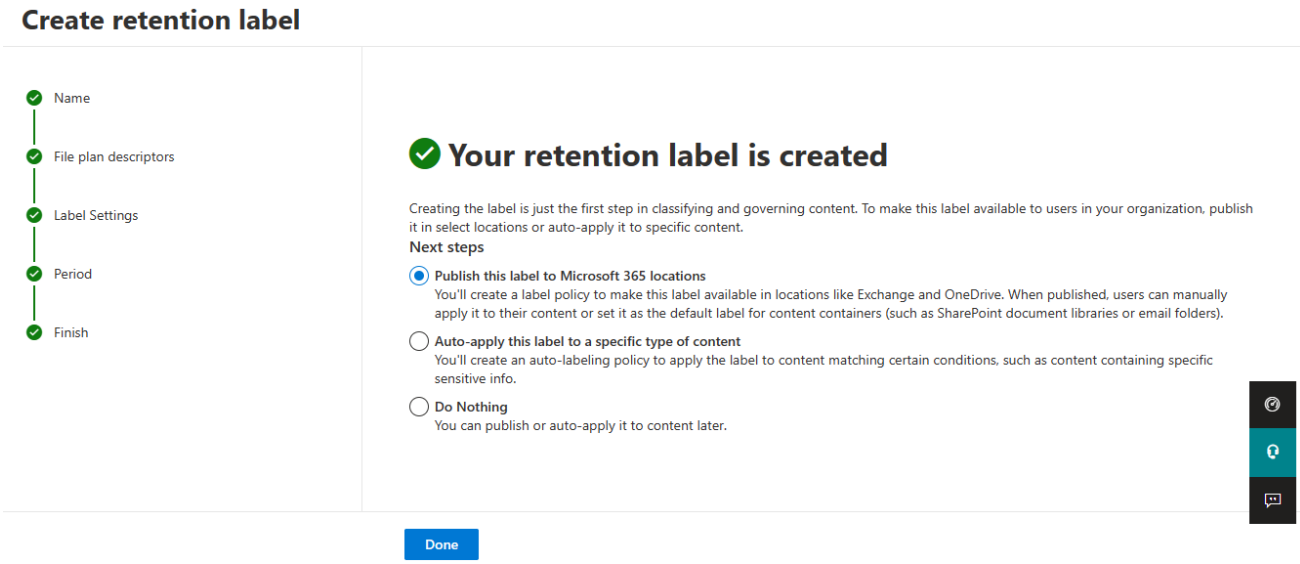

Figure 14: A successfully created retention policy

Once you've published the retention labels in OneDrive or a SharePoint library, you can label not only Microsoft 365 documents, but also non-Office files such as PDFs. In Content explorer, you can get a quick view of files that have retention labels.

Summary

Traditional classification techniques such as regular expressions, manual, or rule-based approaches can't easily handle massive volumes of data. These types of approaches are only appropriate for specific use cases in which a small number of highly sensitive documents are labeled for access by specific groups or individuals.

Machine learning-powered trainable classifiers enable organizations to quickly and comprehensively discover, label, and protect massive volumes of sensitive data across their digital estate with pre-trained models optimized for performance and scalability. Information Protection delivers a unified set of capabilities for data classification, labeling, and protection not only in Office apps, but also in other popular productivity services where information resides like SharePoint Online, Exchange Online, and Microsoft Teams, and endpoint devices.

We invite you to learn more about this game-changing new technology and how your organization can benefit from it.

Learn more

To learn more about trainable classifiers with Information Protection, visit the following reference links:

Microsoft Purview Information Protection product website

Trainable classifiers technical documentation