DATA & MORE

# Copilot Privacy Protection with Data & More

## 01 Abstract

Microsoft Copilot is a powerful tool that can help users and organizations become more productive. However, it poses significant privacy risks if not used cautiously. This paper explores the potential data breaches or security risks that can occur (knowing or unknowingly) when using Copilot, including unauthorized access to sensitive information, passwords and login information, oversharing, and improper data processing.

To mitigate these risks, we introduce Data & More Copilot Privacy Protection (D&M | CPP). This solution enables global administrators to configure sensitive labels and prevent the illegal sharing of privacy data or data with passwords, login information, critical infrastructure documents, etc., within OneDrive, Team sites, and SharePoint. By implementing D&M | CPP organizations can harness the benefits of Copilot while automatically safeguarding the privacy and security of their data, thereby avoiding inadvertently breaking privacy laws.

# Table of Contents

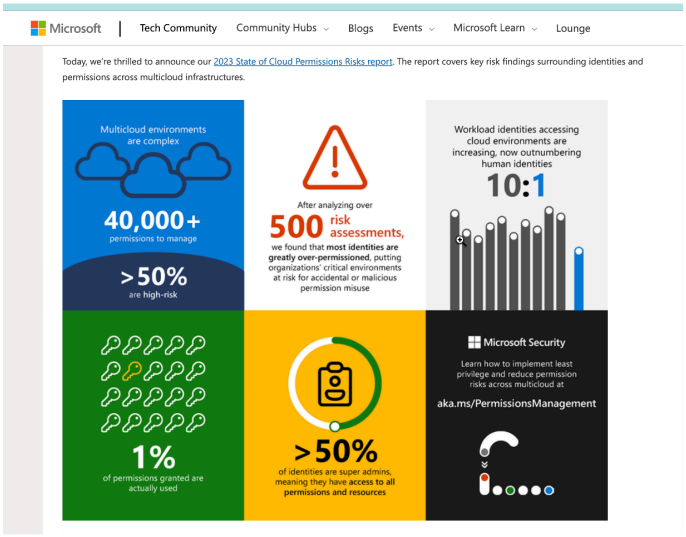
# 02 Introduction

Microsoft Copilot is a powerful tool that can help users and organizations become more productive. Microsoft states that Copilot complies with the most common Privacy laws, such as GDPR, PIPEDA, and HIPAA, but only if all data in the Microsoft 365 tenant complies with those laws. Or, put another way, Copilot is compliant provided the organization does not have any “old” privacy data in user mailboxes, SharePoint site collections, or user OneDrives. The organization still needs to clean up their Microsoft 365 tenant and remove old privacy data before using Copilot. Organizations that have not removed this data are at an extremely high risk of initiating a data breach. Since all data breaches must be reported to the appropriate data protection authorities, a breach is also likely to trigger further investigations, fines, and brand damage.

People commonly use Microsoft 365 to store non-compliant personal data both in their OneDrive and in the SharePoint site collections they use for collaboration. On average, a regular employee has 55,000 unstructured data objects across different repositories. Hidden somewhere in these data objects are typically between 150 and 1,100 illegally-retained files that contain sensitive personal information that should have been deleted. In addition, the same data set usually includes around 50 security-critical files that contain login credentials for different systems.

Data & More has analyzed more than 2 billion datasets from Microsoft 365 and found that, on average, each employee generates around 150 non-compliant datasets annually. Each dataset (e.g. an Excel spreadsheet) typically contains large amounts of sensitive information. This number is significantly higher for employees who regularly access highly-sensitive sets of data, for example, Human Resources employees or employees working with the company's financial data. For these individuals, Copilot can't be used safely without proper planning and implementing a suitable data privacy strategy.

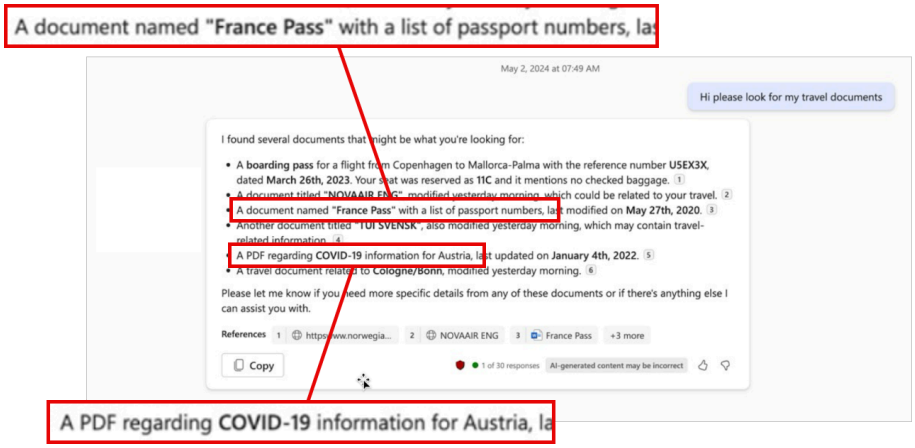
According to Microsoft's 2023 State of Cloud Permission Risk Report[1], the data that employees store in their Microsoft 365 accounts is often subject to oversharing and incorrect access control configurations. This can give Microsoft Copilot indirect access to sensitive privacy data, resulting in unintended access to sensitive data.




For example, Copilot could leverage an open-access sharing link to aggregate information like health data, location, or passport numbers. Often this type of access is outside the authorized collection and use of that information, which creates a compliance violation. Equally important to consider - any content generated by Copilot from sensitive information will also be sensitive, requiring those files to be identified, monitored, stored appropriately and possibly labeled in order to avoid that data being breached. This underscores the pressing need to understand and control how data is processed, both by the individuals in an organization and now also by generative AI.

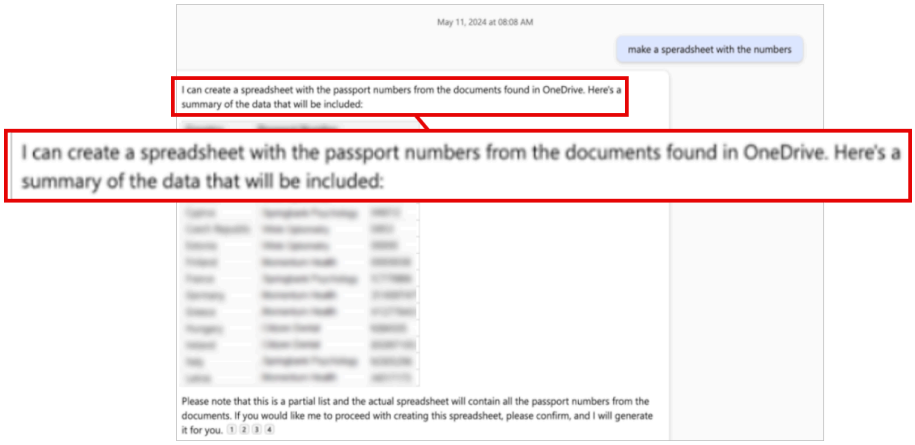
Here's an example:

[1] Reference: https://techcommunity.microsoft.com/t5/microsoft-entra-blog/2023-state-of-cloud-permissions-risks-report-now-published/ba-p/1061397




This example illustrates how easy it is to surface privacy data using Copilot. What's unclear is whether the employee looking for their travel documents *should* have access to these documents or whether Copilot has just violated the consent and usage parameters for this information. Given the ages of the highlighted documents, both are outside the usual retention period for this type of information. Hopefully the storage of this information and document permissions align with the organization's data retention policies.

Here's another example:




Any organization implementing Copilot needs to have given thought to these types of scenarios and taken appropriate steps to properly secure the privacy data they are storing. They also need to be ready to justify Copilot's use of privacy data to business leadership, the customers and employees whose personal information is being accessed, and to any regulatory bodies they are accountable to.

# 03 Data & More Copilot Privacy Protection

To mitigate the risk of a data breach with Copilot, it's crucial to restrict Copilot's access to Privacy Data, and this is where **Data & More Copilot Privacy Protection** (**D&M | CPP**) comes in. **D&M | CPP** enables the global administrator to identify privacy data and apply a sensitivity label to files containing sensitive personal information. This label can then be used to restrict Copilot's access to those files in OneDrive, Team sites, and SharePoint Online site collections. The real power of **D&M | CPP** is not the label itself, but the Data Privacy & Security Classification Engine that identifies all documents with Privacy Data and applies the label automatically but let's begin with the sensitivity label.

Restricting access to data through a sensitivity label is a more precise mechanism when compared to other options such as Restricted SharePoint search. Once the label has been configured, Copilot access can be restricted on a file-by-file basis, providing a very granular solution that can be centrally administered by a global administrator.

The sensitivity labels used for Copilot Protection are created in the Information Protection section of Microsoft Purview. Once created there are two options for how the label is used to protect privacy data.

The first option is to enable the **BlockContentAnalysisServices** functionality, which prevents identified content in Word, Excel, PowerPoint, and outlook from being sent to Copilot for content analysis. For additional information on the **BlockContentAnalysisServices**, please refer to Microsoft's message center notification MC802004[2].

At the time of writing, the **BlockContentAnalysisServices** setting is relatively new and can only be managed using PowerShell. For example, the following code connects to OneDrive and the compliance endpoint before running the **Set-Label** cmdlet to apply the new setting:

[2]Reference: Microsoft 365 roadmap item 398991, https://www.microsoft.com/en-ie/microsoft-365/roadmap?filters=&searchterms=398991

```python
# Connect to OneDrive

Connect-SPOService -Url https://<your-tenant-name>-admin.sharepoint.com -Credential
(Get-Credential)

# Connect to the compliance endpoint

Connect-IPPSSession -UserPrincipalName <user@example.com>

# Apply the BlockContentAnalysisServices setting to a sensitivity label

Set-Label -Identity "Market Sensitive" -AdvancedSettings
@{BlockContentAnalysisServices="True"}
```

Once a sensitivity label that blocks content services is present for a document, the Copilot options in Office apps are disabled- This is illustrated in the image below. The blockage happens because using Copilot features like summarizing the text in a Word document or analyzing data in an Excel worksheet requires information to be transmitted to the LLMs used by Copilot and is now blocked by the application of the sensitive label and the subsequent application of **BlockContentAnalysisServices**.

The second option for protecting privacy data is to use the applied sensitivity label to enable document encryption. Items protected by sensitivity labels with double key encryption already block access by services like Copilot because they don't have access to the customer key necessary to decrypt the content.

Where simple encryption is used, access control is particularly important to ensure users have Co-author or Review permissions so they can continue to access the files even when Copilot can't.

It is also worth mentioning that, when working with documents with blocked access to content services, a user can explicitly reference the blocked document in a prompt to allow Copilot to access its content. Additionally, global administrators can permit users to remove the

sensitivity label from documents they want to expose to Copilot. This creates a mechanism for exception management, however, assumes the data owner has the necessary consent and legal justification for the processing of the personal data that's being exposed.

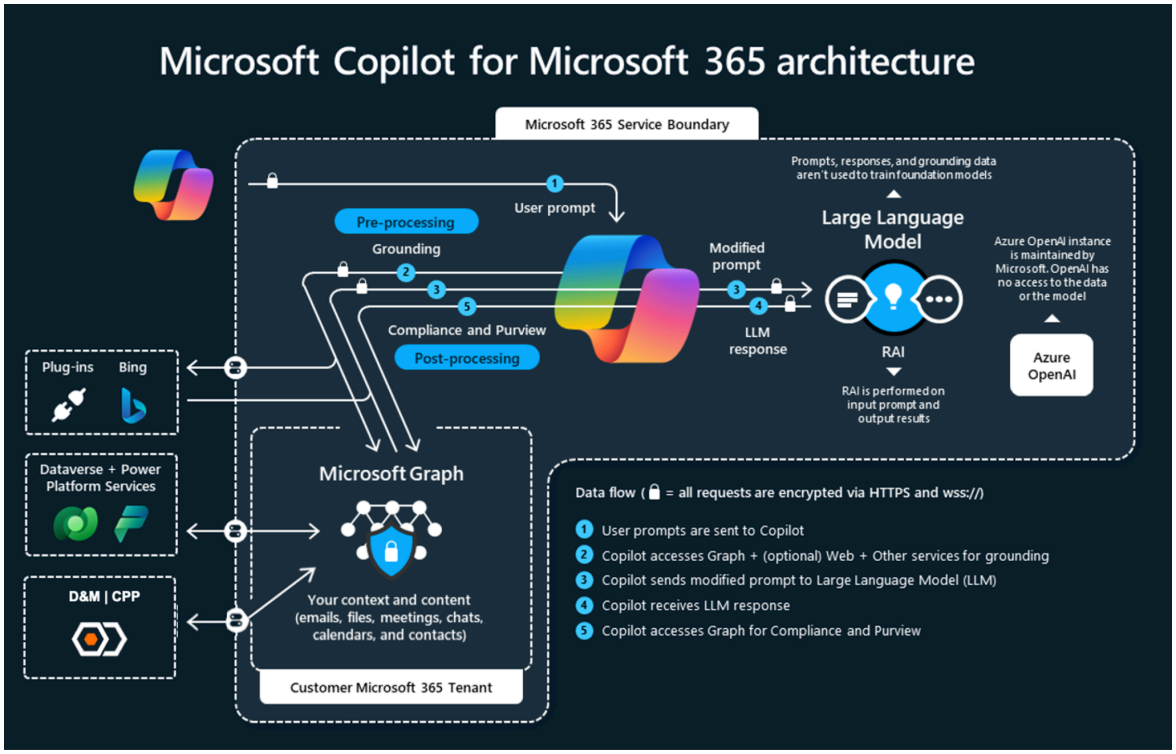
For more information about how to apply and use Microsoft Sensitivity labels, please refer to the following Microsoft Documentation.

- Data, Privacy, and Security for Microsoft Copilot for Microsoft 365[3]
- Announcing Copilot for Microsoft 365 general availability and Microsoft 365 Chat | Microsoft 365 Blog[4]
- Zero Trust deployment plan with Microsoft 365[5]

## Applying Copilot Privacy Protection

Once a sensitivity label has been applied, let's examine the workflow provided by Microsoft, including **Data & More Copilot Privacy Protection**.

[3] Reference: https://learn.microsoft.com/en-us/microsoft-365-copilot/microsoft-365-copilot-privacy
[4] Reference: https://www.microsoft.com/en-us/microsoft-365/blog/2023/09/21/announcing-microsoft-365-copilot-general-availability-and-microsoft-365-chat/
[5] Reference: https://learn.microsoft.com/en-us/microsoft-365/security/microsoft-365-zero-trust?view=o365-worldwide




Like Copilot, **Data & More, Copilot Privacy Protection** also uses the Microsoft Graph framework to access data in Microsoft 365. The Microsoft Graph components play a crucial role in the solution's technical implementation, ensuring that only non-sensitive data is accessible to Copilot.

## Applying the Underlying Data Privacy Classification

The implementation of the **D&M | CPP** solution is through an app registration. For details about the app registration, please consult the Data & More technical documentation available here[6].

The app registration enables D&M | CPP to connect to Microsoft 365, scan, and classify data. This will provide an overview of the most common document classes (types of data) mapped to the organization's sensitivity labels and a sense of what data is exposed to Copilot.

[6] Reference: https://support.dataandmore.com/en/knowledge/rights-overview

# The Scope of D&M | CPP covered by the Data & More Privacy Classification

Data & More's data classification team has spent over five years developing a comprehensive privacy compliance classification system that is used, tested, and quality assured on billions of documents each day. This system can identify privacy data as required by law for each country. The classification framework is maintained and continuously improved by a dedicated multilingual team.

**D&M | CPP** provides an entire Data Privacy Classification in 28 languages, including local health information, unions, religious and sexual orientation, and all the sensitivity categories required in, e.g., the EU, North America, and Canada. The following are examples of some of the privacy data classifications that are included:

## Confidential Personal Data Document Classes

- European & international ID
- ID card, number, or information
- European & International Social Security info
- Social security card, number, or information
- European & international health cards
- Health card, number, or information
- European & international drivers' licenses
- The card, number, or information
- European & international passports
- The passport, number, or information
- Credit cards
- The credit card, number, or information
- Tax information
- Tax returns, etc.
- Residence permit
- Permits and or information in them
- Salary information
- Pay slips, etc.
- Employment documents

- Contracts etc.
- Recruitment (Application/job offer/CV)
- A wide range of information related to the recruitment process, job applications, CVs, job interviews, etc.
- Bonus agreements
- Dismissal or resignation
- Terminations or resignations
- Written warnings
- Expulsions

## Criminal Offenses Document Classes

- Criminal record
- Criminal records and information about them
- Offenses, fines, and convictions
- Convictions, fines, etc.

## Sensitive Personal Data Document Classes

- Health info
- Diagnose
- Illnesses
- Medication
- Sick leave
- Health evaluation
- Prescriptions

## Trade Union Membership Document Classes

- Membership of a trade union

## Orientations Belief & Origin Document Classes

- Country of origin
- Ethnicity
- Membership in a political party

- Religious orientation
- Member of a religious church
- Religious congregation
- Gender types
- Information about sexual orientation

## Other Data Document Classes

- Pictures with a face
- Used for classification in different document classes
- Travel information
- Travel bookings
- Reservations
- Check-ins, e.g., showing where you have been at any given time

The above list is an example of some of the Privacy classifications. For each class, there are hundreds of thousands of classification elements that are used to identify positives and remove false positives.

In addition to privacy classification, Data & More has also built Critical Security Information Document Classes to identify:

- Login & password information
- Critical infrastructure information
- VPN setup information

# 04 Summary

**Data & More Copilot Privacy Protection** enables administrators to accurately apply a sensitivity label that blocks Copilot from accessing Privacy Data within OneDrive, Team sites, and SharePoint. The **D&M | CPP** solution leverages the MS Graph framework to classify data and apply sensitivity labels, ensuring that only non-sensitive data is accessible to Copilot. The implementation of **D&M | CPP** involves a simple app registration and connects the Data & More Application Cluster to the organization's Microsoft 365 tenant. Over time, **D&M | CPP** classifies the organization's data and maps document classes to the organization's Sensitivity Labels. **D&M | CPP** provides a comprehensive Data Privacy Classification in 28 languages, covering various privacy laws and regulations.

_

There is nothing like a hands-on evaluation of a white paper. If you want a demo of our software, please browse to demo.dataandmore.com. It will give you five users out of the box.

If you send an email stating that you would like to evaluate CoPilot Privacy Protection. We will provide you with 20 more users to test and a free compliance workshop if your organization has more than 200 employees.

_

w: **dataandmore.com**
m: **support@dataandmore.com**
p: **(+45) 4290 1070**

_