Microsoft 365, AI & Unstructured Data: Safeguarding Your Business Data

eBook

WeActis by Mondata

Securing Unstructured Data in an AI-Driven Workplace

The way we work has changed. Organizations increasingly depend on strong collaboration, cloud tools, and AI-powered automation to remain competitive. Microsoft 365 has emerged as the cornerstone of business communication, generating and housing extensive volumes of **unstructured data** – such as emails, documents, chats, and beyond. Although these advancements enhance efficiency, they pose considerable security, compliance, and data governance challenges.

AI accelerates opportunities and risks, **exacerbating existing data access**, sharing, and control vulnerabilities. Without a clear preparation strategy, organizations risk exposing sensitive information, enlarging their attack surface, and becoming targets for threats like ransomware and data leaks.

This ebook offers a **roadmap** for securing unstructured data in an AI-driven workplace, covering:

- **The intersection of AI and unstructured data presents both opportunities and threats.**
- **Key risks in Microsoft 365 collaboration and the role of AI in exacerbating them.**
- **Proven strategies to enhance access control, data hygiene, and compliance.**
- **Actionable steps to bolster cyber resilience without disrupting business operations.**

Are you ready to take control of your unstructured data? Let's dive in.

Unstructured Data, AI, and the Security Challenges Ahead

Artificial Intelligence (AI) has quickly transformed how businesses operate, with organizations keen to leverage its potential for enhancing efficiency and maintaining competitiveness. However, AI's extensive integration into collaboration tools like Microsoft 365 introduces new security and compliance challenges—especially concerning unstructured data.

Enterprises generate **massive amounts of unstructured data daily** – documents, emails, chat messages, spreadsheets, etc. While collaboration platforms enhance productivity,they expose organizations to risks such as data breaches, compliance violations, and cyber threats like ransomware. AI further amplifies these risks by automating data processing, sometimes without clear visibility or control over sensitive information.

As AI reshapes the digital workplace, organizations must proactively adapt their security, risk, and compliance strategies.

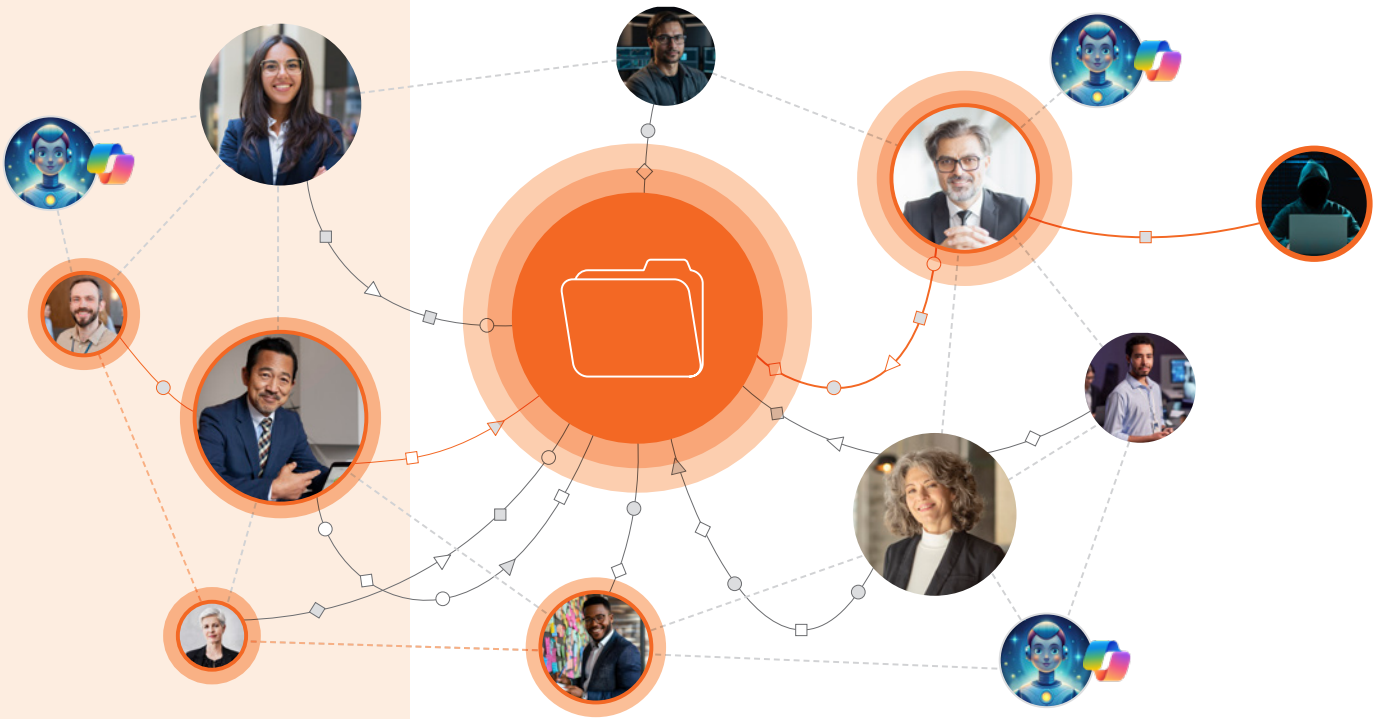
Chapter 1

Rise of the unstructured data

What treasures live in those collaborative environments?

The simple answer is unstructured data.

We have made significant strides in **securing structured data in the past.** This type of data typically resides in applications and systems, among others. It is often planned and organized, making it somewhat more predictable in content, though surprises can still arise. While this remains a constant challenge, it is almost taking a backseat compared to the **rising challenges** posed by its counterpart: unstructured data.



Unstructured data comes in different file formats.

It can include pictures, Word documents, Excel files, presentations, audio, and more. Therefore, it is always different and, you could say, unpredictable. The adoption of collaborative environments and several other factors have driven the amount of unstructured data to a staggering annual growth rate of 55-65%. Some may refer to an "**explosion**" that is currently happening.

Clearly, data has become the new currency, and numerous innovations are designed to generate more data, even though there is often considerable duplication. Many governance initiatives, particularly within the public sector, have established some structure for categorizing this data to ensure adequate protection. This effort aims to guide organizations and effectively address the risks associated with this data.

Unstructured data grows at the annual rate of 55% to 65% and **accounts for more than 80% of all enterprise data.**

Congruity 360 - 2023

Ransomware – a Billion-dollar Industry

The growth of unstructured data has also been causing headaches in cybersecurity and other industries. Criminals have recognized and seized the opportunity presented by unstructured data and organizations' dependence on it to disrupt operations and generate revenue, whether by making it inaccessible to the organization (through encryption) or exposing it publicly (through leaks), all to compel the victim organization to pay.

We have observed significant growth in the market for systems and applications that assist organizations in understanding unstructured data, following the cybersecurity principle that one must know what needs to be secured. These systems can examine an organization's data stores and report on the types of data identified, and in some cases, even label files and enforce certain controls to a limited extent.

Companies require strong strategies to handle the growing amounts of potentially sensitive data; cybersecurity solutions designed for this situation should be rapidly implemented to ensure optimal use and prevent costly cybersecurity and privacy incidents.

Unstructured data and generative AI

An artificial intelligence model is a mathematical structure that enables predictions or generates content based on input data. There are numerous direct links between unstructured data and generative AI.

1 The first point is quite fundamental: **most generative AI models exist because they are trained on unstructured data.** These models are based on large language models (LLM).

For instance, an LLM (large language model) such as ChatGPT consists of a network of neurons designed to comprehend and generate text coherently. It is trained on vast amounts of text to learn the relationships between words and sentences.

Parameters are internal variables of the model that are adjusted during training (or learning) to optimize predictions. LLMs possess billions of parameters that influence how the model interprets a text input and generates an output. These parameters are adjusted during model training to reduce prediction errors related to the training data.

2 Second, **generative AI's biggest strength is to help generate unstructured data.**

As mentioned, LLMs are designed to predict the following output, making them very efficient at generating content from unstructured data when prompted. Consider your chat agent, chatbots, or other applications that assist in writing content. This significantly contributes to the growth mentioned earlier, as we now have systems generating content, not just users. Moreover, these systems don't sleep or eat, allowing them to operate continuously on a 24/7 basis.

3 The third one is **generative AI, which skillfully navigates a maze of unstructured data, whether for legitimate or potentially malicious purposes.**

When you "prompt" an AI assistant within a collaborative environment, it first undergoes a « grounding « process. This process aligns a model's responses with accurate, real-world data to ensure relevance, reliability, and consistency. It gathers information from various sources specific to the organization. Depending on its configuration, the sources and access permissions may vary; however, in an MS 365 environment, this could include "any" data that the end user can access, such as shares from other users, previous accesses, and more.

When a user with unauthorized access prompts the MS 365 copilot with a question, the response could easily reveal sensitive information that this same user **did not** know existed and **had access to.**

Chapter 2

Collaboration at a Cost

The Growing Risks of Unstructured Data

When users ignore Microsoft 365 data hygiene, IT faces a growing mess without the necessary business context to fix it safely.

Access reviews

Because there is significantly more data than before, and since it is largely "unstructured," maintaining a regular, up-to-date, and minimal access review is more crucial than ever. This process must occur regularly for each library where data is stored and, ideally, be **conducted by users who are closely aligned with those libraries' business objectives.**

Shares

Shares are "**discretionary**" accesses created for users outside the regular audience listed in the standard access reviews discussed previously. They are great for productivity but should be **temporary and limited**. We create shares constantly; we should review and revoke them as often as possible, ensuring that only the required ones remain.

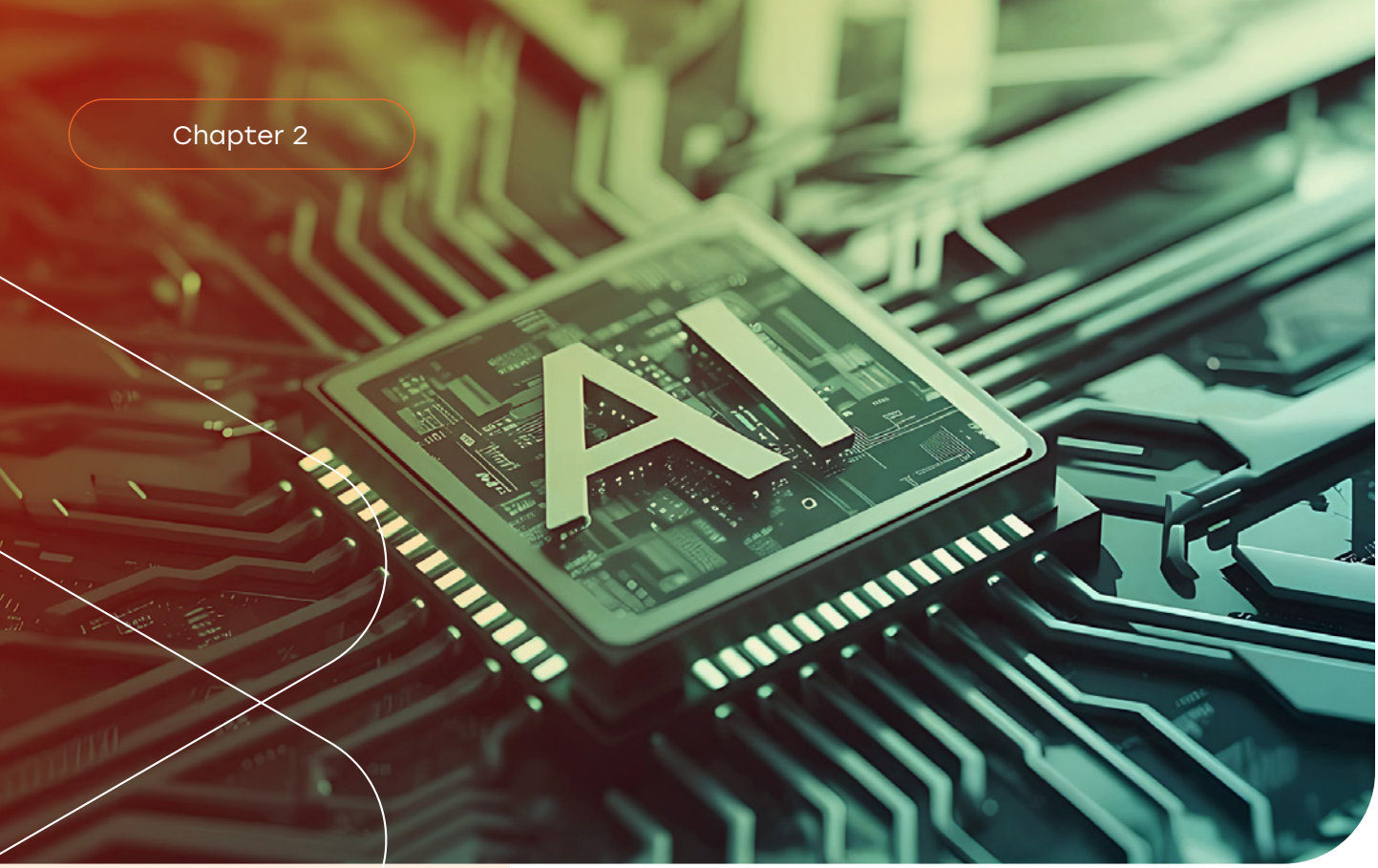
Oversharing

If every file within a library is **shared with everyone**, you have to wonder what the purpose of access reviews is. Oversharing often reveals gaps in a library's structure. Keeping oversharing in check and addressing any gaps in processes as they arise is crucial for maintaining an acceptable level of risk.

Data Hygiene

Maintaining end-to-end control over unstructured data is challenging. However, we can certainly begin with the files and libraries where these files are stored. By implementing appropriate controls in those libraries, **aligned with their business objectives**, and ensuring a comprehensive lifecycle, we can mitigate the risk of unnecessary or unused data posing threats to the organization by being exposed to or utilized by Gen AI. As we continue to generate data daily, we must **cultivate new behaviors** that minimize these risks.

Chapter 2



What are the risks?

By now, you might be starting to draw some initial conclusions. What risks are associated with leveraging powerful tools such as AI within collaboration environments?

Like many other cybersecurity risks, they revolve around the data that lies within. As with any new technology or opportunity, AI introduces new risks. However, as we've seen, **generative AI amplifies or accelerates risks that have existed in organizations long before these AI discussions began.** Although it offers numerous efficiency opportunities, it can also present risks that may disrupt organizations that are unprepared or hesitant to tackle them.

What Strategies Can I Adopt?

1

Start small and build on that

Recognize that this is not a **“project”** with a fixed end date.

In many ways, addressing all those risks can be overwhelming, especially since they have likely existed at different levels for a long time. However, any improvement is a step forward. Recognize that this is not a “project” with a fixed end date. As your organization grows and as the unstructured data increases, **this is part of a new reality.** Whatever measures you implement will endure.

This can be achieved by focusing on shares, reducing excessive amounts, or applying proper controls to the libraries where your most important data should reside.

2

Distribute the responsibility

Who is generating data and utilizing those systems? → **Everyone**

Who has the business context for the data they use and produce? → **The end users and owners**

Who should we engage to see progress and maintain a risk-acceptable environment moving forward? → **As many employees as possible**

Provide your end users with context and simple tools to help them not only leverage those tools but also address risks as they arise. You need everyone’s help.

3 Keep it simple

Align your compliance and controls with the targeted business objectives or data types. To engage your end users and foster a strong cybersecurity culture, you must keep things very simple for everyone. This doesn't mean the backend isn't complex; however, what you present to users and the requests you make of them should be straightforward. You should **establish it as a "habit"** for them.

You must keep things very simple **for everyone.**

4 This is the new reality

You need new indicators to track it. Whether regarding end-user engagement, daily actions taken to reduce risk, trends in share volume, access reviews, or other mechanisms to mitigate risk, you need new indicators and methods to track them. These will either complement or replace some of the existing indicators you may already have, such as phishing rates, click rates, and training. They should be **shared with executives and promoted positively.**

Key Takeaways

AI-powered collaboration is here to stay, bringing a surge in unstructured data that necessitates new security strategies. Organizations must act now to balance innovation with risk management. Here's how:

1

Start Small, Scale Strategically

Focus on quick wins, such as reducing excessive data sharing and applying access controls to high-risk libraries. Security is an ongoing process, not a one-time project.

2

Empower End Users

Security is not solely the responsibility of IT. Equip employees with the context and tools necessary to identify and address data risks in their everyday workflows. Promoting good data hygiene guarantees that sensitive information is managed and protected appropriately.

3

Simplify Security Practices

Align security measures with business goals. When protocols are straightforward, intuitive, and seamlessly integrated, users are more likely to adhere to them.

4

Measure and Adapt

Monitor new security indicators such as access review participation, data sharing trends, and AI-driven data exposure. Communicate these insights with executives to drive awareness and continuous improvement.

Whether your organization fully embraces AI or not, securing unstructured data in Microsoft 365 is non-negotiable. The time to act is now—every step you take strengthens your cyber resilience, improves **data hygiene**, and safeguards your business from emerging threats.

About WeActis

WeActis revolutionizes cybersecurity by turning employees into active defenders and making security a shared responsibility rather than an IT burden.

Integrated into Microsoft Teams, it seamlessly embeds security into daily workflows, guiding employees to mitigate risks in under five minutes per week.

By improving data hygiene in Microsoft 365, strengthening governance, and streamlining risk reduction, **WeActis** helps organizations build a Cybersecurity Resilience Culture—enhancing compliance, reducing data exposure, and driving measurable, lasting security improvements.

info@weactis.com
1-833-666-3282
WeActis.com

WeActis
by Mondata