CONCENTRIC Ai

Is Copilot safe? Microsoft Copilot security concerns explained

Note: this article was updated and refreshed as of 11/15/24

If you've used ChatGPT, you know how powerful and helpful it can be. Today, ChatGPT is only one of many tools for generative AI. In late 2023, Microsoft rolled out its own AI chatbot for the enterprise called Copilot. Google also introduced Gemini (formerly Bard), which also has significant benefits and drawbacks.

At the end of 2024, Copilot has already been deployed in many organizations. But it hasn't been an easy ride.

In March, the U.S. House of Representatives banned congressional staff from using Copilot due to concerns about data security and the potential risk of leaking House data to unauthorized cloud services. In June, Microsoft announced it was discontinuing the GPT Builder feature within Copilot Pro, which allowed users to create custom Copilot GPTs. Also in June, in response to backlash over the Recall feature, which captured and stored screenshots for AI analysis, Microsoft decided to make it an opt-in function rather than enabling it by default. The reality is generative AI tools like Copilot can be wonderful for productivity but introduce some notable red flags for the enterprise.

If you don't have time to read the entire blog, here are the key things you need to know:

1. Copilot has overly broad permissions

One of the primary concerns with Microsoft Copilot is its potential for over-permissioning, which can lead to unintended data access across an organization. As a generative AI tool, Copilot aggregates data from Microsoft 365, potentially creating vulnerabilities if permissions aren't carefully restricted. Mismanaged permissions can result in widespread, often inadvertent access to confidential files, including intellectual property, financial documents, and personal information, underscoring the importance of vigilant data governance.

2. The risks of data exposure are significant

With the integration of AI, there's a heightened risk of sensitive data being accessed unintentionally.

With the integration of AI, there's a heightened risk of sensitive data being accessed unintentionally. Copilot's advanced features allow it to handle significant amounts of data, but without strict access controls, this same functionality can increase exposure risks. Data mishandling incidents or unauthorized data sharing can occur, impacting compliance and security.

3. Organizations need proactive access management

To mitigate the risks, organizations must implement proactive access management measures — before, during and after Copilot deployment. This includes regularly reviewing permission settings within Microsoft 365, conducting data audits, and ensuring users only access data essential to their roles. Proactive management also involves using tools to identify and correct over-permissioned accounts to maintain optimal data security and limit exposure.

What is Microsoft Copilot?

Imagine having a personal AI assistant tucked into every Microsoft 365 app you use – from Word and Excel to Teams and Outlook. The purpose of Copilot is to strip away the tedious bits of an employee's workday so they can be more productive and creative than ever.

What differentiates Copilot from other AI tools like ChatGPT is its ability to deep dive into your organization's Microsoft 365 content. Think of Copilot as having a 24/7 virtual assistant who can quickly and efficiently remember every bit of your work and whip up a summary, spreadsheet or document quickly and efficiently.

The potential for Copilot is endless; the expected productivity surge may exceed that of ChatGPT. Simply open a new Word doc, tell Copilot to draft a client proposal using elements from your notes and past presentations, and you've got a complete proposal in seconds.

Copilot can even summarize Teams meetings, keeping track of the key points and to-dos. It can also be your email wingman in Outlook, helping you sort through your inbox. In Excel, it becomes your data analyst.

How does Copilot work?

If you've been using Microsoft products long enough, you'll remember Clippy. Copilot is as easy as Clippy, but much less annoying and with much better results.

If you've been using Microsoft products long enough, you'll remember Clippy. Copilot is as easy as Clippy, but much less annoying and with much better results.

With Copilot, the process is very simple and works like ChatGPT:

- Open the app/sidebar and type in a prompt
- Microsoft looks at your 365 permissions to get the context
- The prompt is sent to the Large Language Model (think GPT-4) to do its AI magic
- Microsoft puts it through a responsible AI check

Sounds great! What's the catch? Copilot risks explained.

Reviewing common Copilot security risks

In theory, Copilot is a dream come true. But there are security risks that must be addressed. While Microsoft does its best to keep security in mind with its product, data security teams need to know this: Copilot essentially has the keys to the kingdom. It can access all the sensitive data you can, which – to be honest – is sometimes more than it should.

Plus, Copilot can do more than fetch data; it can create new sensitive data quickly and in large quantity.

The big issue here is overly permissive data access, which happens in organizations far more often than you think.

Here at Concentric AI, we publish a Data Risk report twice a year based on our comprehensive findings. Using advanced AI capabilities, Concentric AI processed over 500 million unstructured data records and files from companies in the technology, financial, energy and healthcare sectors. This report underscores the risk to unstructured data in the real world by categorizing the data, evaluating business criticality, and accurately assessing risk.

Our most recent report analyzed over 550 million data records and found that 16% of an organization's business-critical data is overshared. That adds up to a lot of data: on average, organizations have 802 thousand files at risk due to oversharing.

Let's explore some other staggering statistics:

- 83% of the at-risk files were overshared with users or groups within the company
- 17% were overshared with external 3rd parties
- 90% of business-critical documents are shared outside the C-suite
- Over 15% of all business-critical files are at risk from oversharing, erroneous access permissions and inappropriate classification and so can be seen by internal or external users who should not have access
- Over 3% of business sensitive data was shared organization wide without concern for whether it should have been shared or not

All about Copilot's access to sensitive data

Let's assume Copilot becomes as embraced by the enterprise as many expect. In that case, companies must approach its use like any other application: walking a tightrope between productivity and restricting access to employees who need it to do their work. Remember the rush to get employees going with access to work remotely during the pandemic? It was a massive challenge to set permissions and security settings promptly.

The good news is Copilot will only work with your M365 tenant data and won't be able to access other companies' data. Plus, your data doesn't train the AI for other companies to leverage.

However — there are several issues:

- Copilot will leverage all the data that the employee has access to; from our data risk report, it is clear that most employees' permissions to sensitive data are far greater than what they should be entitled to
- Copilot results do not inherit the security labels from the source files. This represents a serious risk for source files containing sensitive data.
- Now, it's up to the employee to double-check the AI's work and ensure data is classified and assessed for risk properly.

Here's what Microsoft says about access rights in its Copilot data security documentation: "It's important that you're using the permission models available in Microsoft 365 services, such as SharePoint, to help ensure the right users or groups have the right access to the right content within your organization."

When it comes to permissions in the ideal world, zero trust is always best, where, like in the CIA, access to information is on a need-to-know basis. Microsoft suggests using M365's permission models to keep things locked down, but in reality, most setups are far from that ideal.

As for the labels and classification methods that companies rely on to keep data protected, they can get messy, and AI-generated data will only make it messier. With so much data to manage, organizations should not expect their employees to be perfect stewards of data risk. We know it's hard enough for security teams.

What about the potential for Cloud data leaks with Copilot Studio?

Recent research has raised concerns around Microsoft Copilot Studio's potential exposure to cloud data leaks. Security researchers discovered a vulnerability (CVE-2024-38206) in Copilot Studio, which allowed for external HTTP requests that could leak sensitive information about internal services within the cloud environment. The issue stems from a server-side request forgery (SSRF) flaw, which enabled authenticated attackers to bypass SSRF protection and access Microsoft's internal infrastructure.

The vulnerability allowed researchers to make HTTP requests that accessed internal cloud services, including Microsoft's Instance Metadata Service (IMDS) and Cosmos DB. While cross-tenant data wasn't immediately accessible, the shared infrastructure across tenants poses a risk: if one tenant's data is affected, other customers could potentially be impacted as well.

The HTTP request exploit in Copilot Studio can be manipulated to target sensitive internal resources inadvertently accessible due to permissions within the cloud environment. Despite Microsoft's prompt response to mitigate the issue (it was patched quickly), the flaw is an example if how Copilot's features, if exploited, may become a gateway to broader data exposure risks.

What are some real-world examples of Copilot security risks?

Let's explore four potential scenarios that are bound to happen across organizational departments in finance, HR, R&D, and marketing.

Finance: A financial analyst uses Copilot to generate a quarterly financial report. The input data includes a mix of public financial figures and sensitive, unreleased earnings data. Due to an oversight, the sensitive data is not correctly classified at the input stage. Then, Copilot generates a comprehensive report that includes sensitive earnings data but fails to classify it as confidential. What if this report was inadvertently shared with external stakeholder?

HR: An HR manager uses Copilot to compile an internal report on employee performance, including personal employee information. The source data has overly permissive access controls, allowing any department member to view all employee records. What if the Copilot-generated report inherits these permissions, and sensitive employee information becomes accessible to all department members, violating privacy policies? It could potentially lead to internal chaos and legal challenges.

Marketing: A marketing team uses Copilot to analyze customer feedback and generate a report on customer satisfaction trends. The input data contains sensitive customer information, including criticism of unreleased products. Since Copilot outputs are unclassified by default, the generated report does not flag the sensitive customer feedback as confidential. What if the report is uploaded to a shared company server without appropriate access restrictions, making the critical feedback—and details about the unreleased products—accessible to unauthorized employees? Internal leaks and competitive disadvantage become a significant risk.

R&D: A product development team uses Copilot to brainstorm new product ideas based on existing intellectual property (IP) and R&D data. The team's input includes confidential information about upcoming patents. Copilot, lacking context on the company's sensitivity towards this IP, incorporates detailed descriptions of these patents in its output. What if this output is then shared with a broader audience, including external partners, inadvertently exposing future product plans and risking IP theft?

How to keep sensitive data safe with Concentric AI

To best manage any type of data risk, sensitive information — from financial data to PII/PHI/PCI to intellectual property to confidential business information — needs to be identified, classified and remediated if at risk.

Remember, sensitive data can be stored in the cloud, on premises, structured or unstructured data. While most classification methods are better than having none at all, most paths to classification — like end-user, centralized and metadata-driven — can be time-consuming, ineffective and full of unnecessary obstacles.

As difficult as this sounds, organizations need to have a clear understanding of data risk before fully deploying Copilot.

How Concentric AI helps secure Copilot

Concentric AI leverages sophisticated natural language processing capabilities to accurately and autonomously categorize data output from Copilot into categories that include privacy-sensitive data, intellectual property, financial information, legal agreements, human resources files, sales strategies, partnership plans and other business-critical information.

Concentric AI can analyze the output from Copilot to discover sensitive information – from financial data to PII/PCI/PHI — and label the data accordingly to ensure that only authorized personnel have access to it. This also ensures that employees don't have to worry about labeling the output, resulting in better security.

Once that data has been identified and classified, Concentric AI can autonomously identify risk from inappropriate permissioning, risky sharing, unauthorized access, wrong location etc.

Remediation actions, such as changing entitlements, adjusting access controls, or preventing the data from being shared, can also be taken centrally to fix issues and prevent data loss.

Best of all, Concentric AI can help you address Copilot's security risks without having to write a single rule.

To sum up, with Concentric AI, your organization can effectively manage generative AI output data:

- Discover, monitor and protect all data types, including cloud, on-premises, structured, unstructured, and shared via messaging services
- Auto-label sensitive data output from Copilot
- Gain a risk-based view of data and users
- Leverage automated remediation to instantly fix access and activity violations
- Find risk without rules, formal policies, regex, or end-user involvement
- Secure API-based SaaS solution with no agents required

Concentric AI is easy to deploy —
sign up in ten minutes and see value in days.

Book a demo today