GOVERNANCE / AI DATA SECURITY

Copilot oversharing: what actually controls the files your AI can read

> *Permissions decide what a person can reach. Labels and policy decide what the AI will hand over. Those are two different sentences, and most owners learn that the expensive way.*

Pedro Bandeira, 3DH Consulting · August 2026 · 10 min read

Get this guide as a PDF

The full guide plus the pre-Copilot checklist, offline-ready.

Download PDF

A lawyer on a client's team stopped me last week with a question I did not have a clean answer to. She asked what stops Copilot, once we switch it on, from reading the documents her firm marks privileged. The ones a court would protect. The ones most of the company is not supposed to see.

My first answer was the obvious one, and it was wrong. Those documents sit in folders only the right people can open, I said, so permissions handle it. She was right to keep pushing. Microsoft 365 Copilot does not respect your folder structure the way you assume it does. It surfaces whatever a person is technically allowed to reach, and in most companies people can reach far more than anyone ever intended. That gap has a name now. Microsoft calls it oversharing, and it is about to become the thing every owner running Copilot has to answer for.

THE ONE LINE TO REMEMBER

> Copilot inherits every access mistake you have ever made. If a document is reachable by someone who asks the right question, Copilot can surface it, summarise it, and quote it back, even if nobody was ever supposed to open that folder. Permissions were the control surface for humans. They are not the control surface for AI.

What oversharing actually is

Start with the scale of the mess. IDC's figure, cited across the industry this year, is that around 90% of enterprise data is unstructured, the emails, documents, spreadsheets and notes that no system was ever designed to read. A Harvard Business Review Analytic Services survey put it more bluntly: only about 15% of organisations have their data in enough order to safely use the agentic AI everyone is now selling them. For years that pile just sat there, too big and too messy for anyone to trawl. Its safety came from the fact that nobody could be bothered to look.

Copilot is a search engine over exactly that pile, pointed at whatever the person asking is allowed to see. Ask it to pull together everything on a deal, a salary review, a redundancy, a legal dispute, and it will search across SharePoint, OneDrive, email and Teams and return what it finds. The retrieval is only as tight as your permissions are. And here is the uncomfortable truth about permissions in a small company: they grow by accident. Somebody was added to a site for one project and never removed. A folder was shared with "anyone with the link" to get a file to a supplier in a hurry. An assistant inherited a manager's access and kept it. None of that mattered while the data was too vast to search. Copilot makes it searchable in one sentence.

That is oversharing. Not a leak, not a hack. It is the quiet accumulation of access nobody cleaned up, suddenly made legible by a tool that is very good at finding things.

Why folder permissions are the wrong mental model

The instinct of every owner I talk to is the same as mine was. The sensitive things live in a locked room, so we are fine. The problem is that the room has more keys cut than you remember, and Copilot tries every key the asker holds. Two things break the folder model specifically.

First, reach is not the same as intent. A person may be technically permitted to open a document they would never think to go looking for. They will not stumble across it in a thousand clicks. Copilot will surface it the moment a question touches it, because it does not browse the way people do, it retrieves. The obscurity that protected that file, the fact that no human would ever find it, evaporates.

Second, permissions in Microsoft 365 are layered and inherited in ways almost nobody has mapped. Access flows down from a site, across from a shared link, sideways through a group somebody was added to years ago. To know what any one person can actually reach, you would have to reconstruct all of it. In a company without a dedicated IT team, that map does not exist. So "it is in a protected folder" is a statement of hope, not a control.

> THE DISTINCTION THAT MATTERS
>
> Permissions govern what a person can open. That is a human control, and it leaks over time. What governs what Copilot will process is different: it is the label on the document and the policy you set around that label. You can leave the messy permissions exactly as they are and still stop Copilot from touching the things that matter, if you control it at the right layer.

What actually controls it: four layers, not one

The tempting takeaway is "just label your sensitive files and you are done." That is closer to the truth than the folder model, but it is not the whole truth, and selling it as a one-step fix is how firms get caught. The real control is a short stack of layers. You do not need all four on day one, but you need to know they exist.

1. Sensitivity labels plus a Copilot DLP policy

A sensitivity label is a tag you put on a document: confidential, privileged, internal only, whatever your scheme is. On its own, a label is just metadata. The control comes when you pair it with a rule. Microsoft lets you build a data loss prevention policy scoped to Copilot itself, using the "Microsoft 365 Copilot" location and a condition of "content contains a sensitivity label." When that is on, Copilot will not use a labelled document to build an answer, even for a person who is

allowed to open it. This is the mechanism that answers the lawyer's question: label the privileged documents, set the policy, and Copilot leaves them alone.

2. Restricted Content Discovery for whole areas

Labelling works file by file. Sometimes you want to fence off an entire area, the HR site, the board folder, the legal workspace, without labelling every item inside it. Restricted Content Discovery, part of SharePoint Advanced Management, removes a site from Copilot's retrieval pipeline entirely. People who have access can still open the files by hand. Copilot simply cannot see them. For a small firm, pointing this at two or three genuinely sensitive sites is often faster than a labelling project and closes the biggest holes first.

3. The permissions themselves, eventually

Labels and discovery rules keep Copilot out. They do not fix the underlying fact that too many people can reach too much. That still matters, for the day someone leaves, for an audit, for the next tool you turn on. So the permissions cleanup does not disappear, it just stops being the thing that blocks you from switching Copilot on safely this month. Do the fast controls now, schedule the deeper cleanup after.

4. A posture assessment so you are not guessing

Microsoft's Data Security Posture Management for AI, inside Purview, runs an automated weekly assessment of your busiest SharePoint sites, flags files that carry sensitive content, and highlights anything shared through "anyone" links. It then suggests the exact policies to fix what it found, and it lets you run them in simulation first, so you can see what would have been blocked before anything actually is. This is how you replace "I think we are fine" with a list. It is the same move I make on every engagement: turn a vague worry into something you can read in ten seconds.

> **THE HONEST CAVEAT**
>
> Labels are not magic, and anyone who tells you they are is selling. Two things to know. The metadata of a protected document is still indexed, so Copilot can sometimes reference that a document exists and cite its title without reading the contents. And there have been documented cases where Copilot mishandled labelled files. This is exactly why the answer

> is layered. Labels plus a DLP policy plus restricted discovery on your most sensitive areas is defence in depth. Any single one on its own is a single point of failure.

The order to do this in

You do not need a six-month programme before you are allowed to use Copilot. You need to close the loud holes first and schedule the rest. Here is the sequence I run with a small company, roughly a week of real work spread over a month, not a quarter of consulting.

1. **Name the untouchables.** Sit with whoever knows the business and list the categories Copilot must never surface: privileged legal, HR and payroll, board papers, live deals, anything under a confidentiality agreement. This is a conversation, not a project. It takes an afternoon.

2. **Fence the worst areas with Restricted Content Discovery.** The two or three sites that hold those untouchables come out of Copilot's reach immediately. This is the fastest, highest-value control and it needs no file-by-file work.

3. **Label and set the Copilot DLP policy.** For documents that live outside those sites, apply your sensitivity labels and switch on the DLP policy that excludes labelled content from Copilot. Start in simulation mode and read what it would have blocked.

4. **Run the posture assessment and read the list.** Let DSPM for AI scan your busiest sites. Fix the "anyone" links it finds. This surfaces the accidents you had forgotten.

5. **Only now, widen Copilot access.** With the untouchables fenced and the loud holes closed, roll Copilot out to the people who will actually use it. Adoption is a separate problem, and a real one, but at least you are not adopting on top of an open vault.

6. **Schedule the permissions cleanup.** Book the deeper work to tighten who can reach what. It is no longer blocking you, but it is still owed.

Notice what is not in that list: buy a platform. The controls above are already inside the Microsoft licences a firm running Copilot is very likely paying for. The work is configuration and a few decisions, not procurement.

The clock: this stops being optional in October

None of this is theoretical, and the timing is not mine. Microsoft is building Copilot oversharing alerts and DLP controls directly into the Microsoft 365 admin center, targeted for October 2026. Read that for what it is. Microsoft is putting the oversharing problem in front of every administrator by default because enough companies switched the AI on before they did the groundwork, and the results were bad enough to warrant a permanent warning light on the dashboard. The tool arrived faster than the housekeeping, and now the housekeeping is being made unavoidable.

The firms that will be comfortable in October are the ones that did the sequence above quietly beforehand. The ones that will be scrambling are the ones who treated "we turned Copilot on" as the finish line rather than the start. This is the same pattern as every governance deadline: the work is small if you do it early and painful if you do it under a light that is already blinking red. It also sits directly alongside the wider EU AI Act obligations that landed in August 2026, where the same question, who can see what and can you prove it, is the one a regulator asks.

> **WHAT TO DO THIS WEEK**
>
> You do not need a consultant to start. Open Purview, find Data Security Posture Management for AI, and let it run its weekly assessment on your top sites. Read the list it gives you. That single act turns oversharing from an abstract fear into a specific, fixable set of items, and it costs you nothing but the time to look.

The point most owners miss

The lawyer's question was the right one, and my first answer was the wrong one, because I reached for the human control, permissions, when the AI honours a different one. Permissions decide what a person can reach. Labels and policy decide what Copilot will hand over. Those are

two different sentences. Most small companies do not have a Copilot security problem so much as a visibility problem: the sensitive material is reachable in ways nobody has ever mapped, and the AI is simply the first tool patient enough to find all of it.

So before you roll AI across your business, the dull question is the whole game. Not who can open the folder. What is written on the document, and what rule sits around it. Get that right and Copilot is what it is supposed to be, an assistant that knows your business. Get it wrong and it is the most thorough intern you ever hired, reading everything, forgetting nothing, and answering honestly to whoever asks. The good news is that the controls are already in the licence. The work is deciding what matters and spending a week saying so out loud. Getting your data house in order this way is the same foundation that makes every other obligation easier, from AI governance to the structured-invoicing rules now landing across the EU, which is exactly why we treat data readiness as one job, not five.

Turning Copilot on without opening the vault

3DH configures the label, DLP and discovery controls that stop Copilot oversharing, using the Microsoft licences you already pay for. No platform to buy. A week of real work, not a quarter of slides.

See how we do it

SOURCES

- Microsoft Learn: Use Microsoft Purview to manage data security and compliance for Microsoft 365 Copilot
- Microsoft Learn: Microsoft Purview DLP for Microsoft 365 Copilot and Copilot Chat
- Microsoft Learn: Purview data security posture management for AI

- Windows News: Microsoft to put Copilot oversharing alerts and DLP controls in the M365 admin center, October 2026
- Reltio, citing IDC and HBR Analytic Services: 90% of enterprise data is dark, 15% of firms are AI-ready
- VentureBeat: cases of Copilot mishandling sensitivity labels

---

Published by 3DH Consulting. Practical AI adoption, Microsoft 365 governance and EU compliance for small and mid-sized firms. This is a business translation, not legal advice.