Microsoft Purview Data Security Governance Foundations for Copilot for Microsoft 365-20240418_185807-Meeting Recording

April 18, 2024, 5:59PM

58m 21s

Greg Wartes 0:13

Good afternoon, everyone. Thank you guys for taking some time on your day to join us here today. We're gonna probably start at 2 after just to give some people time to funnel in looking at.

The lobby. It doesn't look like anyone's not in. So it looks like everyone's getting in. It needs to be here, so that's good. But two after we will go ahead and get started. But thank you guys again for take some time out of your day.

Alright, as promised it is 2 after it looks like the.

Attendee count is holding pretty steady, so we'll go ahead and kick things off. I believe the recording has already started, so we can check that box.

First and foremost, we just want to say thank you for taking some time out your day today to join us. We're we're here to talk about 2 components of security. One is AI with Co pilot and the other is compliance with purview. We've got 2.

Offer or I should say, one offer with two great speakers. This is a rinse and repeat of a webinar we did about a month and 1/2 ago to a different subset of customers, and when we got done with the presentation.

A lot of customers came back to us and said, you know, I've got a little, a little. Specialty in my environment, I've got something that's not the norm and I was hoping that you could help me answer some questions is how we could leverage AI. So we can keep our data secure using AI. So for this round of the webinar we have included and Microsoft loves our acronym something called the CCC or their Co pilot connect consultation. It is a no fee one hour consultation with lighthouse. If when we get to the end of this that is of interest to your organisation you can either scan the QR code now or you can send Eric at Lighthouse an e-mail.

His e-mail is there on the slide. I will also ask that Eric drops his e-mail into the chat so we all have it, but the feedback from these consultations.

That we've conducted thus far has been it's been immensely helpful to the organisations, so if that's something that's of interest, I would encourage you to

contact Eric at Lighthouse to get that set. As far as the presenters for today, we've got two phenomenal security speakers, one from here at Microsoft, Ameren Gill. He's a senior technical specialist with a focus and data security and compliance. And then Chris Baird from lighthouse.
That's got a very impressive resume that goes back years where they focus on both security and compliance. So with that being said, I will turn it over to Amirin and if there's any questions at the as the slides progress, please feel free to drop questions into the chat. We've got myself some teams from Microsoft and a team from lighthouse monitoring the chat. We'll make sure that all questions get answered. So Amarin, all yours there.

AG **Amren Gill** 5:02
Thanks, Greg. Good, so hey everyone. Thanks for joining us today. I wanna just so start by setting the stage of you know, acknowledging the fact that we are well and truly in the era of Gen AI, right. So this the image above kind of shows that you know depicts to how quickly the adoption of Gen AI has really taken place. So you know two months compared to you know some of these other significant technology trends really sets the tone of how significant this adoption is and that this rapid adoption is moving right, so.
The in the business world specifically, this is absolutely nothing sort of short of a game changer, right? So it's like we're at the starting line of a whole new race in technology and productivity for companies everywhere, right. So this isn't just a case of using fancy new tools, it's about total digital transformation, right. So how businesses operate, how they compete, how they think about growth is being completely turned on its head, right? So.
The Gen AI is like having a technology superpower. Your fingertips will have the ability to crunch data, spit out insights, streamline tasks, you know, spark innovations at a pace that we've never seen before.
I I equate this you know paradigm shift to significant shift that we've seen previously, things like you know moving to cloud from on Prem storages, right? So things like the the move to a virtual workplace when we had COVID lockdown everybody rapid adoption of teams.
This is one of those significant paradigm shifts that will really scope the future of how we work right and a common denominator with those changes that we've seen is that Microsoft is absolutely at the forefront of those shifts from a technology

standpoint, right? So these these technology shifts and these developments are all great, but what it does is obviously present obvious or associated risks that must be considered, right. So some of the innovation that Microsoft has brought to the market.
You know, particularly with Gen AI, has has really been incredible. So you know in this in this real chart I think some from our first announcement of Co pilot, we've expanded the reach from GitHub to dynamics to our power platform. And so many other services that when I was putting these slides together, I was talking over with my colleague and he was saying that you know he said don't try and list them all out because he said that I think we're up to 92 Co pilots either active or being worked on, right. So you know really.
Seems significant development being in being done in the back end.
And we're absolutely not stopping there. Right, so more topically today, our focus is gonna be on copilot for M365.
This, this, this specifically this copilot from GT5. You know, it's the copilot that works with you alongside your apps in your everyday work, right. So this incredibly powerful tool that can reason over your data landscape, right, your business data and significantly boost the efficiency and digestion of creation of that business information. Right.
And.
With copilot, Friend 365's ability to almost instantly surface data from all corners of your data estate. It presents the necessity to build controls and protection around that your most sensitive data, as it suddenly is so accessible, right? So, so immediately, and so quickly accessible. So this obviously prompts the question of how do I protect my data from the evident risk associated with Geni I right, specifically copilot, and this is where.
Microsoft, per view, plays such an integral role to this process right the the first step really is understanding why and how copilot is in enterprise ready right? So this is really down to the fact that the protection capabilities of purview are built into copilot, right? Which ensures that any data security or data governance policies that you build are honoured by copilot. So when they try to access data.
Those security those policies are in place are honoured by copilot, right? So this means that as long as you're adequately safeguarding your data, copilot will not be able to access this data on a user's behalf. That it shouldn't, right?
And there's a very simple concept which is depicted by this. This access of access try

out of M365 data quite nicely and the first requirement is that the data is made available right? So is the content indexed. So once it's indexed it's available for the Microsoft graph and then subsequently the semantic index to be able to reason over and then return results with copilot, right? So then we get into the protection controls that dictate whether copilot can actually interact with the data or not. Right. So.

Do users have access to the sensitivity label? For example, right? So copilot will send a request to look for data on a user's behalf, right? So meaning that it's acting on behalf of the user, right? So if the, IE the user's permission, so if the user and only if the user has authorised access, is this considered a yes, right. And then thirdly, we have while the user may have access to authorization, do they have the necessary permissions to be able to extract?

Content from the file, right? So as we're talking about generative AI, the likelihood is that new content may be generated from this original data, right? So if a user has for example view only permissions to a file, extracting the content from the file shouldn't be possible, right? Copilot shouldn't be able to pull information and create new content from that file.

If and only if, though, all those three conditions are met, that's when copilot for M365 will return results.

So how exactly does purview play a role in all this? Well, purview is a suite of tools that facilitate the security and governance of data, as well as be able to being able to mitigate risk and ensuring compliance right. So the controllers that I mentioned before, so that identification of sensitive content, the you know applying sensitive labels with encryption.

Restricting view and extraction permissions. All of those are implemented with tools in purview. OK, so with purview we can apply those in controls encryption.

Go one step further by building layered security. So once that data has been created, we can actually ensure that that data isn't mishandled or overshed, right? And even use machine learning to develop risk profiles for users that are interacting with copilot and sensitive data together and then take appropriate action based on what that machine learning that AI model has determined as as the risk associated with that user. Right? So and then, you know, going once they're further looking at some of our private preview features which are coming soon, like the AI hub.

We can actually look at activity in copilot and consumer AI apps, so things like ChatGPT.

Bard Bing chat thing over 100 popular generic Gen AI apps, right? So.
To dive into a little bit more about exactly how perv you can do this and help you with this process, I'm going to hand over to Chris from Lighthouse you can walk us through this in a little bit more detail.

Chris Baird 12:30
You should amran. Thank you. Let me just switch the slides over. Hopefully you can see my slides now guys.
All right, great introduction, Aaron. Some really important points and I think there's going to be quite a few things that I kind of pick up on what you said and and repeat throughout my presentation, what we're going to talk about today guys is really kind of diving into what can you do using the purview suite as Aaron says, to kind of mitigate some of those risks, get to that full enablement of of Co pilot for M365.
Little brief intro to me and my organisation, so my name is Chris Bird. I'm a senior security consultant and I lead up the information protection practise for a company called Lighthouse. We are a professional services company. We have a broader business that performs esie and E discovery, but our consultancy practise works on this kind of technology on a daily basis. We've got a really strategic alliance with Microsoft that goes way back as well and there's a lot of benefits to that in terms of it brings US clients, introduces clients to us but of course we get key access to people like Cameron.
Access to the product group as well and really good insights. I mean Amaran, I'm really, really keen to see more about that, that private preview for AI. Hope it's I was reading about that recently sounds brilliant. It's some really good technology coming there. So we're going to go through a couple of things. Firstly, I just want to kind of bring this back to the importance of establishing an information governance programme that can really help you adopt Co pilot for M365 in a compliant manner.
So we'll talk just just kind of briefly you know firstly from Lighthouse, we do take a standards based approach. So people in my team, we do recruit people from.
Security.
Disciplines. So, you know, colleagues come across with experience of common frameworks like NIST, ISO CIS and those kind of things. We've got an army of legal experts in lighthouse and regulatory experts as well. And this is going to be important within your organisation to have these kind of people engaged in your programme as well. We'll see some of that shortly.

So many may be familiar with with the information governance framework from armour. This is something that we use. It helps us kind of help our clients build up models and and look to their implementation. So there's an information governance implementation model. There is an information governance reference model that we are going to look at in a little bit more detail. This is really important to help kind of drive and structure your programme for Co pilot adoption. And then the other thing that we look at is your maturity, right? So one of the things that we appreciate is every organisation has a different level of maturity.
There's a slide later where I talk about what's all been new and perv. You being something that has been around for a long time, some of you may be ahead of others in terms of your implementation.
One of the things that we really need to do with any kind of programme is look at where you are in your maturity. What are the quick wins that you can bring in to enable Co pilot and what can you put on that kind of longer timeline?
So in terms of what does an IT reference model look like, the key things really for the model is bringing in the right people from your business. And there's there's so many reasons to to discuss this one. But the two that we're going to focus on primarily are what are your drivers for business productivity in terms of enhancing your data quality, removing errors and bias from your data. Aaron's talked on some on some of those in the introduction, but the other important thing, the reason we need the business stakeholders and this is with any IG product project, not just with Co filer. Is to make sure we understand data in its business context. We're able to describe that data in terms of its use. It's handling how it should be managed, life cycle and so on. And we really need to bring the business to the table for that. You're also going to bring in people from it. Maybe your modern workplace team, you're going to bring in the people that manage your M365 platform, you're going to have colleagues from security, maybe security, architecture, digital forensics investigations. You see Cert team and you sock.
We're gonna bring in people from privacy. Really important, not just to look at the data that's going into Co pilot and review that from a privacy perspective, but also in terms of the controls that I'm going to show you today. They come with requirements to see data, to see audit information and those kind of things as well. What we don't want is is to upset the privacy team. We want to bring those people in as well and make sure make sure we've got them on board and then we're going to contact risk. We're going to talk to legal records management. You can see we've got

a significant number of stakeholders here. So one of the things that we really kind of suggest you guys look to put in.
Before you.
To adopt Co pilot and and these technologies is really establishing IG Steering Committee. You don't need to change your operation. You don't need to build an op model, you know, bring in these stakeholders, meet on a regular basis so that you can take accountability as a group, as a forum for things like process standards, policy documentation. And we can really drive that down through your policy into what we're going to show you today into implementing that, that process through the controls in purview.
So how does a standard based approach work in purview? Well, at the centre of that model we can see the IG RM talks about the creation and receipt of data. With Microsoft 365, we have outlook, exchange and teams. We've got data coming in and coming out of the environment. We also have connected apps, right? So we need to factor that in as, as Amin said in his intro, anything that the user has access to through the M365 universe is potentially accessible to Co pilot. So we need to look at that data that's coming in. We need to protect that data.
We need to potentially retain that data.
For for business reasons, but also to ensure that we don't over retain data or maybe perform disposition. And then we've got data transfer management and and ensuring that we've got alerts of of data transfers across business units and entities within your organisation.
This is just a slightly different view now that kind of shows kind of from a standards perspective that information access layer that I talked about, which could be connections to on Prem repositories where you guys have data, it could be connection to cloud applications where you've got data and also your Microsoft Microsoft 365 data assets in SharePoint, OneDrive and teams. We've got this cloud based access through M365. You then builds on that. It gives us Microsoft information protection that allows us to do really creative things.
With sensitivity labels, classification of documents and so on.
We can manage information life cycle through SharePoint through enterprise content management to capture the collaboration. The version control of documents.
We have purview records management which allows us to set retention policies and define the storage that we use for records, so we can declare information as a record.

We can store it immutably.
Many of you will be familiar with ediscovery. We've got ediscovery standard and ediscovery premium within Microsoft per view as well. And then finally, we've got the disposition and archiving implemented through data lifecycle management and disposition review. Now all of these per view controls are available through the E3 and E5 product offerings and various different Sku's, but we're going to talk about where these can fit in and where you can get the best out of these to ensure that that really awesome Co pilot adoption.
So I'll start just briefly talk about the copilot architecture, and I think Cameron really touched on the key point of this slide for me earlier, the concept of grounding. So I think everybody who's here understands what Co pilot is, you know, Ameren talked about however many copilots there are. We're focusing on M365. So it is a set of prompts within your Microsoft 365 canvases. So think about being in a Word document, being able to bring up the AI type and input into a prompt and get a response in a standard language that you could understand. And you can use within your document.
When you enter that prompt into copilot, what's going to happen? As Amarin talked about, is at the bottom left of the diagram here, you've got your tenant, all of your data, all of your Microsoft 365 data. What you enter into copilot is going to be sent to your tenant. It's going to go through the Microsoft graph through the semantic layer, and it's going to ground. It's going to you're going to hear this word ground in quite a bit. It's going to ensure that anything that the requestor asks for is only information that's available to that requester and not information that's not available to them. So if they don't have permission.
If they have a sensitivity label, it's blocking their access. That person is not able to access that information. That's a really important thing to understand with Co pilot today.
Likewise, Co pilot isn't going to go out to Bing chats, not going to go out to the open AI data and and language model. It's only going to look at your Office 365 data.
So that response is going to come back from the graph. It's going to say yes, copilot. I've got this data for this user. I've got these recommendations for documents. I have all this access for Chris. It's going to pass that back across to the the large language model which is going to give you that GNAI response that's going to really improve your quality, right. You're going to take that response and maybe use that to write an

exec summary in your document. But the last thing that copilot does that we're going to really focus on today is it's going to go back and check what are the post processing actions I need to perform before I give that user that response.
Do I need to ensure a document is labelled correctly before I pass that data to that user?
Are there any permissions or restrictions on that data?
So benefits of Co pilot again, I think we all understand the benefits of Co pilot. That's that's why we're here. We want to drive business productivity. We want to want to really improve quality. So you're going to have an experience where you're writing an e-mail copilot's going to suggest content to you. It's going to suggest documents that may be relevant to the emails that you type in. It's going to improve your tone, it's going to improve your language in document creation. It's going to enhance our quality and this is this is really important as well. We're talking about the information in terms of its, its security and risk. But we also need to think about information in terms of its accuracy, its consistency and quality. The last thing we want to do is take bad information.
And feed that into Co pilot because bad information in equals poor information out.
And then finally, it's going to give things like code assistance as well and what we need to think about. As I've talked about, there are one or two of these we need to think about the impact on data exposure from a sensitive data perspective and proprietary information disclosure, data spillage, those kind of things, biases and errors, unauthorised data access and the compliance challenges that come with that.
And some of the capabilities that we'll talk about today in terms of Microsoft purview that that help us with this are going to be really focusing on data classification, you're going to hear me talk about how that's the cornerstone of your information governance programme in terms of Co pilot adoption. We'll always come back to the need to do good data classification. We'll talk a little bit about that shortly. We'll then move on to, once we understand data in its context, we'll start to talk about how you can label that data and how those labels are used by Co pilot to ensure protection.
And the right permissions and the right information is fed to the right requester.
Remember that that grounding that I talked about, which is really important.
We'll talk about auditing. There's two aspects to this. We'll talk about how it's important to be able to audit the prompt and responses that Co pilot gives and have access to those through audit and other tool sets, but also general auditing of the information through its lifecycle. What we don't want to do is lose management

control of data which goes from one SharePoint repository in one business unit to another. We need to have auditability of that, to ensure that we can apply the right controls to all of our SharePoint landscape, not just one business unit.
So the key target is really to get to a point where we understand risk. We understand our data and we can start to implement policies that provide protection and purview gives us really, really great way to implement those policies.
So now this this model, we won't go into any great depth on this, but this is useful for you to come back to in the recording and view in future. It's almost a square onion if you like, but what it's showing here is your data is it is at the bottom and then we've kind of got these layers around your data that provide the protections that we need for not just for copilot for M365, but general protection for your business as well from data exfiltration and other threats and purview gives us all of those tools to do that. So we can see we've got a classification engine at the bottom there sits across our Microsoft 365 data.
It's going to run 24 hours a day, seven days a week. Classify your data in a common index. It doesn't mark your data, it doesn't impact your data in any way, but it stores a record of that information containing sensitive information. So if you've got Social Security numbers in lists within documents on SharePoint, that classification engine is going to find them. And this is something folks that if you've got M365 regardless of whether you're an E3 or an E5 customer, this is running in your tenant right now. So you can go and look at this. You can look at these data classifications and really get some.
Great insights out of your tenant today.
As we move up through the layers, we then start to look at now we understand the data in terms of its its classification. We can apply sensitivity labels to it. We can life cycle, manage that data. We can apply auto labelling policies to protect and we're going to show you a bit of that today as well and then have sort of preventative and detective controls through data loss prevention, communication compliance and insider risk policy.
One of the things we like to to kind of highlight is, you know, we'll say what is all this new, what we mean by that is.
The purview is, is called purview now. It used to be called Compliance Centre. Many of you, if you've been on the M365 journey as long as I have, you will have used this portal and these products in one format or another. Maybe if that was just for your incident response team to look at message trace to use content search, you will have

used the pervy portal, so something you might have a head start, right? In terms of you've got access control to allow your administrators in to define further roles to allow a role based model for your organisation to manage. Some of these tools. But really what we're saying is come back, revisit pervy, take a look at it. It's really moved on and we can really use it to to enhance access control and permissions. Data protection. We can eliminate stale data. Make sure we've got the right kind of retention and we're not over sharing documents within within our applications.
So the apparent team has never really been greater. There's a lot of pressures, not just with Co pilot, we've got, you know, data privacy is a bigger risk now than it's ever been. Information security drivers, the the threat landscape is, is is really hot right now and sort of ediscovery and legal matters. You know we've, we've got a lot of things in the public domain that organisations want to make sure they've now got these processes to protect themselves. And then when we layer on the dawn of jet AI, it makes absolute sense to come and review.
What you can do with purview.
I'll go back to Ameren's Point as well around some of these points that are on screen. M365 Co pilot respects all of your existing controls around access control and permissions, data protection and the sharing of documents, right. So that's really important. I'll come back to that grounding again that you know, I want you to go away from this session and realise that maybe work to do, but understand there are some basic things that copilot has out-of-the-box that already give you these kind of protections straight from day one.
So I always like to do information protection first, but I'm not going to do that today and the reason is I want to show you discovery and communication compliance on the basis that if anybody in this session is exploring E5, if you do have E5 or a compliance sq, then these are two of the quick wins. These are really low barrier to entry things that you can do right now. So you can deliver a message back up in your organisation that we have eyes on Co pilot. There are things that we can do to get eyes on it to to get alerts and those kind of things.
And these are really, really great, great tools. So the first one we'll talk about is communication compliance. It's a digital surveillance tool. So if you are looking at digital surveillance, you know, take a look at communication compliance. But I'm going to talk about today is only one policy. I'm not looking at using this for four digital surveillance across your organisation. Communication compliance gives me the ability to monitor and surveil information and communications such as teams,

chats and outlook messages.
Great thing that Microsoft has done recently with this product is they've introduced copilot prompt monitoring and communication compliance as well.
So with a single policy, you guys can right now implement communication compliance to look and track at those messages that go into those prompts and those responses that come can come back and hopefully you can see in the deck if you've got on a big enough screen. Here on the screenshot, we can see we've got a request here made in Co pilot for Microsoft Word. Something was fed in to request some maybe some sensitive information. Copilot gave a response to say couldn't get access to that information.
And we as an investigator, we can receive an alert to tell us somebody is exploring sensitive information in Co pilot, we can get eyes on that quite quickly. So that's a really good message that you can give up to your to your senior managers that you can start to do things right now and get eyes on on copilot interactions.
The other one that I just wanted to touch on was ediscovery premium, so if anybody is using ediscovery as part of their broader ESI and legal processes, they need discovery premium with a small tweak to your standard operating procedures and maybe some role based access control changes, you can use ediscovery to do investigations to go and do searches, and collections on sensitive information that may have been overshared through Co pilot, so that's another tool that you can you can leverage right now to go and do that, particularly if you've already got an ediscovery process in place.
What you're going to be able to do is pick a, pick a custodian like you would normally do, but you'll be able to choose copilot as a location, just like you can for teams. For for exchange, and you'll be able to, as you can see on the screenshots here, you'll see the copilot information, the prompts and responses. The other great thing that Microsoft's got as well is just search and collect. If you if you have process processes connected to graph, you can actually delete data. So if you know you've got data spillage, if you know you've had some information that's been over shared. And you need to delete that data in your organisation. You can do that as well and then from that you can take those alerts. You can take those reference points and fix the problems at source. Maybe fix the permissions that led to that issue in the 1st place. So I just wanted to show those as quick wins before I get into information protection because I think that the messaging that you can give if you've got those available to you right now is is very positive indeed.

So with information protection, I talked a little bit about data classification earlier be in the cornerstone of everything that you you build your programme on and it's so important for Co pilot. And when we talk about classification, some people on on on the session might think, oh I, I get that it's confidential, it's highly confidential. We'll come to that. That's that's the labelling aspect when we talk about classification, we're really trying to get to what is that data and what does it mean in a business context. So is it health information, is it finance information? Is it legal information? And if we understand that, we understand how we should handle that data, how we should share that data and life cycling and of course, whether or not that information can be shared externally or shared with other business entities or units within your organisation, that's going to improve decision making. It's going to enhance security, it's going to improve your regulatory compliance position as well.

So how do we do that with Microsoft 365 with for for E3 customers on this call right now, go and take a look at your sensitive information types. If you haven't done so already, there's about 320 at the last count, I think maybe it's more of built in sensitive information types that are continuously classifying your data right now and these are going to look at common patterns within data. They're going to look at things like words, expressions within certain distances of one another as well. So if you think you can write a regular expression.

That's going to fit in here if you think you've got a keyword dictionary that's going to fit in here, maybe you're using these today to look for credit cards, Social Security numbers and those kind of things.

The other tool that some of the E5 or compliance customers can use which whilst we're on the subject of AI and ML, Microsoft has a classified type called a trainable classifier which leverages a machine learning model and this is really more focused on that business context where the sit looks at the patterns, the expressions but less around the business context. The training will classify as looking for documents by type. So I'm trying to find a finance document. What does that look like? How is it structured? How is it formed? What is the language within that document and what kind of things does it contain?

And you can build these models through a wizard very, very easily by feeding in seeding documents from known data sets in your organisation. Examples might be proprietary source code. You can feed it information that you've got within your source code repository so it understands how you write your source code, what your comments look like, what your kind of languages that you use are, maybe your

finance documents. You can build your own finance classifiers as well. And when we when we combine these things, when we combine the trainable classifiers and the sits.
We can get a really high level of confidence about what the information is, so think of an example. We've got a sit that detects an invoice number, but we've got a trainable classifier that detects finance document. We could have a rule in place that says if we find something that matches finance and contains an invoice number and contains a name, let's put a label on that because we know that's really sensitive and I'm going to get to labelling on the next slide. And this is really going to come back to Co pilot. You'll see that in the next few slides. But this is the foundation. This is the first place that you need to start is getting these classifiers working for your business. You know, reduction of false positives, making sure that they're matching your content and and you really understand the business context of your data.
So let's assume we have that we have that, that great ability to to understand what our data looks like. Now we're going to go into the next level of being able to apply a label to that. So like I said with that example with an invoice or a finance document, you might want to apply a confidential label to that because we know it's sensitive. There are clever things that we'll get into on the next few slides that you can do with those labels to apply at different levels, different levels of protection that are really going to impact Co pilot as well. So a sensitivity label is a marking.
In a document and it is carried on that document wherever that document is shared and it will display to users. For example, if they open up word documents or PDFs, they will see that the information has been labelled. You have the ability to also apply a watermark on the content for printing as well.
But apply protective and restrictive controls that will explore shortly. You also have the ability to then auto label, so one thing is creation of these labels so your users can use them. The next thing is being able to auto label your content when you find it matches a particular classifier.
It's really going to help with Co pilot as well, because the important point about Co pilot it does it we're we're not highlighting the security issue. We're highlighting the potential for poor access control and governance, right. So if you have libraries out there today where you know you've got poor access control and governance, SharePoint libraries, which you just don't have, the A/C LS that you really wish you had in place on them, you've got SharePoint admins opening sites and making public sites with autolabelling. We're changing the paradigm a little bit to come away from

the access control list and protect the content itself.
So now we're going to apply a label to a document that might be sat in a public SharePoint site, but it's finance document. It's confidential. We're going to apply a confidential label to it, and that's going to have some restrictive controls. So even if I go to that public site and I take that information out, I still don't have access to it. And you can see how that links to copilot, right? So when we talked about copilot only being able to see the things the user can see, if we have a label that blocks me from seeing that document, then copilot can't see it either.
Want to see all the points? Just one kind of quick point to note from for those that kind of look at all of this and at the end of the session are still in a bit of a panic about use of Co pilot. There is a product called Double key encryption which we often talk about avoiding unless you've got a real need for it. But actually if you want to protect your most secret data, double key encryption is going to hide that data from Microsoft that any data protected with DKE is not going to be ingestible into Microsoft 365 Co pilot. So if you know you've got data today.
Not a way. You could create a secret label. I'm gonna show you that shortly. That secret label can apply double key encryption and you know any of that data is just simply not accessible to Microsoft, to the M365 value add like ediscovery DLP. And certainly not to copilot for M365.
So with a label we can apply permissions. The important thing here is to understand that the administrators can apply permissions to labels. So we can say who within the organisation should be able to interact or or view this content. But also we can apply external permissions as well. Less important from a copilot perspective. More important from a data exfiltration data leakage perspective there, but we're going to focus today on kind of permissions within your organisation.
So as an example, just for the next few slides, we'll show a a simple label taxonomy. We'll talk about a public label, which may be a marketing your communications team, have that label available for things that are published into the public domain. Maybe you've got an internal or a general label, which is kind of a default general business information, probably some good practises around that would be put adlp policy in to stop that leaving the organisation, but no need to really encrypt it. We're not concerned that Copilot's got access to that data.
But then where it gets fun is with these confidential and highly confidential labels so confidential we could have that content encrypted. We could have a rule that says all users in the organisation or part of the organisation have access to that. Maybe third

party need to validate who they are with a one time pass code but highly confidential documents. You can you can imagine you could have multiple labels for these. You could have project labels for highly sensitive projects and you're only going to apply permissions to the people who should see that content and again back to my point around grounding that means.
Not going to see this data.
And then finally, the secret label with Double key encryption.
So what does this look like in a couple of examples? We've got a patterns team here that use a SharePoint site. The administrators just made the site public wasn't quite thinking because when he set it up, he was just putting in some general process how to create a patent? What is the application template? But now we've got people actually putting patent applications in that SharePoint library. The access control is dead on it, which means everybody can go to that library and pull the patent application. But now we've got auto labelling. We've got the ability to detect that that information contains a pattern.
So we can apply the highly confidential label. That means we can apply the permissions that come with that label, which means that Chris Bird when he's typing into his Co pilot prompt show me a list of patterns within the organisation. If I don't have access to information defined by that label, I don't see those patterns.
Similar, another example will be a letter of intent for a merger or acquisition. This may be even more secret than your patents or, or maybe less. Whichever way around you look at this, but this example uses that double key encryption to really kind of make that data opaque so that Co pilot can't see it whatsoever.
So how this looks through Co pilot, what we can see here is I've crafted a prompt, a request to Co pilot using the patterns, process document file and how to file a patent document and other documents from the patents library. Go and generate the assure FAQ about patents. What we can see at the top of screenshot number one on the left hand side is I don't have any label on this document. This is a brand new document that I'm writing. I've got no label on this. I'm going to click generate against copilot, remember through that architecture it's going to go through the graph. See what I've got access to past that to the LLM.
And at that final point, when I said it's going to do those cheques with purview, what you can see on the second screenshot is it's detected, hey, there's patent information here that's really highly confidential information that's been detected in some of these. Some of this information. So Co pilot's now going to automatically label the

content based upon the information that it read from. So again, we've got, you know, we'll go back to those two kind of controls, control number one, if the label prevents the user from making the request in the 1st place, copilot's going to come back to me only with an FAQ. It's not going to come back with any patent information.
If I ask it for punt information, it's going to say I don't have access to that, Chris, but the other one here is if I do have access to that, at least the document I've created now, the new data that I've generated carries that confidential label and that's going to happen automatically.
So now we'll we'll talk about data loss prevention. Next, I want you to kind of open your mind a little bit to this from some sort of the traditional ways of thinking of DLP from an exfiltration point of view. Microsoft gives us tools that can really focus on the concept of data loss within your organisation. So between business units, R&D data, for example, leaking into the hands of sales and marketing that then expose that data to the wrong people inadvertently by again through by patents, example asking for information that they're then feed into a marketing pamphlet or something like that. So we can use the DLP.
Tools that purview gives us straight away to really implement some controls here as well. There's a lot you can do with E3, by the way. You know there's there's E5 here as well, but if you're an E3 customer, there's a lot you can do with purview DLP.
So the control points are really around e-mail, preventing sensitive information from being shared. So if you think about the the SharePoint access control discipline issue, people picking stuff up on SharePoint and emailing it across your organisation that you've lost management control that goes into another SharePoint, suddenly people are going to ask questions to Co pilot. It's going to go to SharePoint repository B&C. That data we didn't intend for that to happen. We lost the management control so we can implement DLP policies through teams and exchange that really recognise data that should not be transferred should not be moved to other parts of the organisation and we can stop that or even just alert to it, provide auditing as well.
And the way DLP works, the engine is. It's really great. It's an if this then that type condition and action cause and effect type engine. But I'll give you one example here of what I just talked through there as that data left the SharePoint site and it went on to the e-mail, we could have adlp policy that says if we spot this type of information being emailed internally, put a label on it, even if we don't have an auto labelling policy. If you guys don't have that maturity yet, but you've got DLP, you can apply the label so that those permissions apply to the recipient.

Which means they're not gonna be able to access that document, and either is copilot.
The other thing we can do with DLP just before I sort of close and move to data lifecycle management is there is a product called Defender for cloud apps. Any of my kind of security counterparts on this call will know that's Microsoft's CASB. But what that has is the ability to create file policies that can help you look at where documents have been shared. So it has a link been shared externally, has it been shared internally outside of your SharePoint access control. So you know somebody shares a link, maybe anonymous anonymously or direct with somebody and you can go into defender for cloud apps and see.
Reports of the information that's been shared.
The other thing I suggest you take a look at SharePoint Advanced Management. There are some additional licencing for some aspects of SharePoint Advanced Management but there is one aspect that I'm pretty sure you get with E5 which is the information governance reporting and that's going to give you a view of oversharing and and over permissive access.
So data lifecycle and records management, what we want to talk about here is legacy and stale data. And just to give you an example, right, all data in equals all data around poor data out. Think about if we ask a child to draw his pictures of animals, but we only ever show that child dinosaurs from the prehistoric age, we're not going to get pictures of dogs and cats and elephants and those kind of things at the modern era because we trained the child on the wrong things and Co pilot is going to be exactly the same and you guys are going to want to use this to.
Enhance the accuracy and.
Consistency of your data and use really, really accurate information, so it's important that yes, we retain data for business purposes, but if we have the opportunity to delete data, we should do that to ensure that Co pilot isn't pulling in that legacy data and giving you poor things out.
How can you do that in purview? There's a few options. The first thing that we'd suggest is a baseline retention policy that many of you probably have. You can apply these different policies within purview to the various products the various storage like e-mail, like SharePoint, and some examples here. Let's take e-mail. We may choose to retain data in a mailbox for 30 days. After that 30 days, the user is allowed to delete that data themselves, but after a year, that data is forcefully deleted from the mailbox.

The way we know that no data older than a year is going to be accessed or ingested into copilot.
So that's kind of systematic deletion and systematic retention, but the one thing that purview also gives us is then how we can layer over that with selective retention. So there is something called retention labels and similar to those sensitivity labels, these can be rules based. So if we see finance data based upon your data classification, right, we we find a trainable classifier with finance. It's got some key financial information in that document. We can have a retention label policy configured to apply a seven-year retention label.
To that document automatically. So now that e-mail reaches its one year, goes into disposition, but there's a flag on the file to say hey, I'm sticking around for another six years. That file is not going to be deleted, it's going to be retained longer. So you can do this. You can implement these controls to think about making this data inaccessible to copilot, but still doing the things that your business requires to keep that key data.
Some of the other aspects available for for for Co pilot in terms of risk perspective, one is PREDA, go and take a look at this, speak to Amaran your Microsoft team about this one as well. It's not part of E3 or E5, it is an additional module. You may qualify for a trial on this one. I won't talk about all the data privacy things in terms of Dsa's, but some really exciting stuff happening in Preva right now. Follow the road map, you know keep keep an eye on those things with Microsoft. But the things that I want to highlight here are in privacy risk management. That's a particular module within preva.
So when we talked about DLP having those kind of binary controls, you know cause and effect, something happens. We want to we want to work on that, stop that information being shared and so on what previous going to give you is more dynamic insights and alerts based upon overexposure data transfers that may occur between business entities and it's a, it's a really, really great product. So if we look at data minimization, you can get started in preview with some policies that look at, we've got sensitive PII here. It's appeared in the SharePoint repository and it seems to be getting over retained. We seem to be keeping it beyond what maybe GDPR or the regulations or other organisations in our industry keep that data.
So it's going to focus on content age and it's going to give you contextual alerts to say this data is now at risk.
In a similar way, data over exposure, maybe data has been shared with the wrong

people. The permissions you know we've we've got poor access, control and governance and we've got data shipped sat in SharePoint repositories accessible to the wrong people. Previous got to give us those insights again quite in a quite dynamic way. And then finally data transfers. So this isn't, excuse me, this isn't just data transfer to external. Think about this maybe business unit or business entity within your organisation and preva's got to give you those insights into where? Data is maybe gone from your your research team, your intellectual property managers, those guys that work in your R&D function.

And it's been picked up and moved to a different business unit or a different entity all together. And the risk with that is if you share a common M365 tenant or even if you don't like that, data is going to be accessible to Co pilot if it loses the permissions and controls that should be applied to it. So remember, if you apply sensitivity labels and permissions even with a data transfer, you de risk in that in the 1st place. But if you don't have that level of maturity, then privacy just gives you that additional capability.

One of the final controls that I want to talk about today is insider risk management. We don't often think about this through the Co pilot lens, but it is really important to ensure that you've got a posture across your entire data asset landscape and across your entire organisation. The Inside risk management capability if you guys do use a behaviour, analytics or insider threat within a seam or anything like that. Today, Microsoft offers a really, really great product called Insider Risk Management that does that same thing. And if you're going to maybe consider investment for auto labelling.

Or for some of the more creative DRP controls. Then just check your licencing type one because this may be something that that you can use as well and it is going to help us from a copilot perspective. So what inside of risk does is it looks at indicators and triggers that occur across data within your organisation that may suggest that data is at risk now. That could be something as simple as a colleague resigns or maybe they missed out on a job on a job or something like that. Right. But it could also be somebody is using.

Threatening language within a communication.

Now where that comes in through the Co pilot lens is we might see somebody persistently asking for sensitive information from Co pilot. Maybe they're getting frustrated and the prompts are getting frustrated and they turn into threats. Communication compliance, remember, can detect that language. It can see the

prompts and responses. For copilot, it can detect negative sentiment, threats, corporate sabotage. And it can tell inside a risk. Hey, we've got this user that's really trying to access sensitive data that they don't have access to. You might want to check this guy out.

Irm is going to give us a 90 day view of all that activity and those activities could be things like removing a sensitivity label. So you've gone to all this great effort to put these sensitivity labels in place to auto label your content, but now we've got somebody that's taking the labels off, moving that content somewhere else and suddenly we've lost that management control again and that data becomes accessible to copilot. So whilst insider risk doesn't really manifest directly in the Co pilot interface, it's not a control that we think of as being directly associated with copilot. It is going to improve your posture.

We're going to derisk things. It's going to make sure that those kind of actions of removing labels, removing sensitive data from SharePoint repositories, taking it outside the permission boundary, changing that data, you're going to get those insights and be able to do those forensic investigations on that data.

And then finally, one of the things you can do with IRM is integrate it with third party connectors as well. So you have the ability, let's say SAP has an event stream that goes into your scene today. We can take those events into inside a risk and correlate those events. So now we can see the user is persistently trying to access and move sensitive information with SAP. Think about this. The copilot can't access data that's in SAP, but now we've got somebody that go into SAP persistently downloading data to the laptop and moving it to SharePoint. Suddenly that data's come out of its management domain in SAP.

And it's ended up in your Microsoft 365 landscaping. It's giving you a headache. That data now needs to be governed. It needs access control. It needs to be labelled. So being able to follow these event flows is going to give you those insights into where new data appears in your organisation that may be accessible to copilot that you didn't expect.

So in terms of review and next steps, just kind of covering the key important things really I I talked about the the absolute importance, the criticality of understanding data in its business context. That is the first thing to go away and do today start to bring in your business engaging business, insure you've got a rich set of classifiers across your landscape. You don't have to be an E5 customer customer to be able to do that. If you didn't know that existed, you'll be surprised to go into M365, look at

data classification and you'll see all sorts of insights.
Into what kind of data you have across your landscape. The next step we talked about is making sure that once we understand that data in its business context, we apply labels and protection to that data. So that Co pilot has the right level of access. Remember, we talked about grounding, making sure that the request that only has access to what they're supposed to see and if we've got a label on a document that denies somebody from seeing that content, Co pilot cannot see that content through that user prompt either.
So the next the next step from that would really be to look across your environment at how you can then implement policies we talked about looking at data loss prevention and Microsoft preba as well offsetting the benefits of those two tools to look at maybe data over exposure data over sharing subtle differences there between the two and look at maybe data transfers. And then we looked at other controls like communication compliance, ediscovery as quick wins and insider risk management as well in terms of improving your security posture.
The compliance posture through quick detection and triage of serious, serious kind of breaches of information that's in your sensitive repositories.
So. So with that, folks, I I guess we can kind of look to any questions or answers. I haven't been able to see whether we got anything in the Q&A. But also I'll pass back to Amrin as well to kind of close us out with any final comments as well amirin. So I'll pass back to you.

AG **Amren Gill** 54:09
Yeah. Thanks, Chris. I think there was there was just on there was one and I'm I'm gonna just quickly put this up on screen just so everybody has this as we kind of close this out. I wanted to just address a couple of the questions.
One of the the last questions that I got from Christopher Ward, it's a great question, which was, are there any white papers on how insider risk management and preva comply or conflict with the many emerging privacy laws in different countries? And it's a great question. It's a great question I think.
With, you know, with some of the technology that Microsoft has bringing out, it's only been it's it's rapidly moving so fast, it's very, very well documented on the Microsoft documentation page. But sometimes, particularly with those emerging privacy laws, it can be difficult for the, you know, white papers or documentation to keep up with that, right.

That is one of the biggest benefits of working with a partner like Lighthouse because they bring together, they bring to the table this really unique blend of privacy experts and technologists.
And those, you know, those those professionals are really technologists and privacy or legal professionals, right? So they have a, you know, a great background in both of the two. So while there may not be documentation and you can't find that information out there, one of the, you know, advances of working with a company like a partner like Lighthouse is they have the expertise in order to be able to relate the technology to those privacy and legal implications, right. Of of adopting those technologies. But it's a great question. So Long story short, I don't think there's a tonne of white paper stuff out there. I've I've. I'd love to be corrected. And I'd love someone to pull it up. But that's again.
One of the benefits of working with a partner like Lighthouse.
And I'm happy to open the floor up if there are any other questions.
That weren't answered, or if there's anything else.
From anybody.

Greg Wartes 56:00
Cameron and Chris, sorry to interrupt. This is Greg. I I did receive one question via e-mail.
That was asking me to ask during the webinar and it was what is copilot data retention look like as far as the prompts and the responses are concerned from from the users and the chat session for lack of better term.

Chris Baird 56:23
So it's really gonna follow. So great question, Greg. It's really gonna follow what we talked about in terms of retention policies, retention labels like like it respects labels in terms of their sensitivity. It will respect labels in terms of their retention as well. Remember that data is stored at source when when you use copilot to generate new data, you're going to store that as new data. You're going to store that maybe as a document file, as an Excel file, as a new piece of data. And just like you saw that data is going to be classified, it's going to get sensitivity label applied to it dynamically. If that data matches a retention rule, it's going to get that retention label applied to that new data as well. So copilot again is going to respect all of those things. But yeah, great. Great question. And as I said, sort of the language model as well. Guys,

any anything you feed in to copilot doesn't go outside of your tenant. It doesn't go to to Bing, it doesn't go to open AI. It stays within your tenant copilot, is really focused on giving you that information to create that new data rather than storing that in a model of its own.

Amren Gill 57:31
Perfect. OK. Any other questions?

Chris Baird 57:32
All.

Amren Gill 57:37
K Well, like I said, the I've left up on the screen the you know the. Please take advantage of the the Nook V1 hour consultation with lighthouse. You know it it really, really will be beneficial. I think some of the questions that we answered him would would benefit from that session as well and you know hopefully some we can help you with your copilot journey and make things easy and give you Peace of Mind from a data security perspective.

Chris Baird 57:58
Absolutely. We look forward to it.

Amren Gill 58:00
Yeah. Thank you so much, Chris. The lighthouse and Greg as well, I appreciate and everybody for spending the time today and I hope you have a great rest of your day.

Chris Baird 58:08
Thanks Cameron. Thanks team and thanks everyone for joining. Appreciate it.

Erick Bronson stopped transcription