Office of the Victorian Information Commissioner

PUBLIC STATEMENT

Use of Microsoft 365 Copilot in the Victorian public sector

This public statement is made pursuant to section 8C(1)(f) of the *Privacy and Data Protection Act 2014* (Vic) (**PDP Act**).[1] The statement relates to the potential adoption and use of Microsoft 365 Copilot (**Copilot**) by Victorian public sector (**VPS**) organisations.

The Acting Information Commissioner has issued a public statement specifically about Copilot on the basis that Copilot will soon be available to organisations that have an existing Microsoft E3 or E5 licence.

The statement outlines the Office of the Victorian Information Commissioner's (**OVIC**) minimum expectations regarding the adoption and use of Copilot, and it outlines privacy and information security considerations for VPS organisations.

OVIC will continue to monitor developments in relation to Gen-AI and will provide further guidance on other tools where appropriate.[2]

What is Copilot?

Copilot is a tool that, when integrated, will appear across all Microsoft 365 applications such as Word, Outlook, Teams, Excel and PowerPoint. It is a form of generative artificial intelligence (**Gen-AI**) that uses the GPT-4 Large Language Model (**LLM**) and access to an organisation's information holdings and systems to create new content based on prompts from a user.[3]

[1] While the public statement is made in accordance with the Information Commissioner's privacy functions, the information contained within this statement also relates to organisations' data security responsibilities under Part 4 of the PDP Act.

[2] Other GenAI tools such as Stable Diffusion, Adobe tools, or those technologies that create content including text, images, code or audio will be considered and guidance provided in the future, where appropriate. If your organisation is considering the adoption of Copilot or other Gen-AI tools and have privacy, security or freedom of information questions, please contact OVIC via enquiries@ovic.vic.gov.au.

[3] For more information, see Microsoft, *Introducing Microsoft 365 Copilot – your copilot for work* (accessed on 19 September 2023), https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/.

The Gen-AI landscape is rapidly evolving and, while many organisations may have watched this from afar up to now, the general release of Copilot – which is due to be made available from 1 November 2023 – could mark a point of departure where the use of Gen-AI becomes an everyday occurrence.

Copilot appears to offer opportunities to optimise routine tasks across the VPS. However, the capabilities and productivity impact of Copilot will depend on the specific context, the circumstances in which it is used, and the level of training and knowledge of the user.

The adoption and use of Copilot may also magnify existing information security and privacy risks for VPS organisations and create new ones (some examples are included in this statement below). Organisations must therefore take the steps outlined below to improve their information security maturity and privacy practices before adopting Copilot.

Organisations are responsible for all actions and content generated by Copilot. Where an information security incident or interference with privacy occurs through the use of Copilot, organisations will not be able to simply say that the incident or interference was "caused by the Gen-AI".

Minimum expectations before adopting Copilot

Before using the features available in Copilot, at a minimum, organisations should:

- assess the maturity of their existing information security program prior to signing up for the integration of Copilot. This includes:
    - identifying existing information holdings and systems that may be impacted by the introduction of Copilot, including due consideration of the security value of these;
    - considering how any newly generated information by Copilot will be assessed, valued and securely managed, including applying appropriate protective marking for newly generated information assets;
    - conducting an updated information security risk assessment considering the integration of Copilot, including due consideration of adjusted risk profiles and development of treatment plans to address relevant Copilot features;
    - implementing any updated treatment plans by rolling out any new or changed controls. Specific consideration should be given to a review of the organisation's existing logical access controls, informed by an audit of all current access controls to protect, detect, and handle against unauthorised access to information holdings and systems;

  - implementing a formal process for the ongoing monitoring and review of risks and controls, especially critical given the dynamic development and release of enhancements and new features of Copilot.

- *explicitly disable* data sharing with Microsoft in platforms such as Power Platform and Dynamics 365;

- undertake a privacy impact assessment to understand the ways in which Copilot will be utilised, the risks it presents to the privacy of individuals, and how to mitigate those risks;

- develop and implement clear guidance and training on the use of Copilot, that:

  - ensures personnel and systems *do not* have the option to send feedback or report content or bugs to Microsoft, as doing so may result in the unauthorised disclosure of public sector information; [4]

  - covers prompt engineering, risk assessment and management, human review of generated language, and active monitoring and reviews to identify potential misuse; and

  - ensures all users of Copilot are aware of how the tools may generate, collect, use or disclose personal information and public sector data, and understand how to ensure the quality and security of the information;

- ensure all users undertake training with a Microsoft Accredited partner prior to accessing and using Microsoft Copilot features; and

- develop an incident response plan for dealing with inadvertent disclosures or misuse of information through the use of Copilot.

Privacy considerations

The 10 Information Privacy Principles (**IPPs**), as set out in Schedule 1 of the PDP Act, will apply to the collection and handling of personal information that may be generated or used by Copilot.

In deciding whether to integrate Copilot, organisations should undertake a privacy impact assessment (**PIA**) to identify:

- what types of personal information will be collected, generated and handled as part of the use of Copilot;

- the functions or activities of the organisation that will involve the use of Copilot;

[4] Ibid.

- any risks to the information privacy of individuals (and potential non-compliance with the IPPs) through the use of Copilot; and
- whether information privacy risks can be properly mitigated, and the steps the organisation will take to do so.

Further to this, organisations should be aware of how Copilot operates across all Microsoft 365 applications and develop and implement adequate policies and procedures to ensure users are aware of how personal information will be collected, generated, used and disclosed when using Copilot.

While all 10 IPPs are relevant to the handling of personal information by an organisation, an organisation should be particularly mindful of its obligations under IPPs 1, 2, 3 and 4 when considering whether to integrate Copilot. Examples have been included below to demonstrate certain features in Copilot and highlight some of the potential privacy risks for organisations to be wary of.

Collection of personal information

IPP 1 sets out that an organisation must only collect personal information if it is necessary to fulfill one or more of its functions. The collection must be lawful, fair and not unreasonably intrusive. Organisations must also provide notice of why the information is being collected, how it will be used and how individuals can access the information.

Organisations should understand that when Copilot generates new content in response to a user prompt (such as "draft me a summary of the main points in the document called 'John Smith's CV'"), this generated content constitutes a new 'collection' of personal information. Organisations must comply with the requirements of IPP 1.[5]

Example – Collection of personal information via Microsoft Teams

An organisation's HR team invites a Senior Operations Manager to a meeting via Teams to provide an update on an internal conduct investigation involving one of the Manager's staff.

The Manager realises at the last minute that they cannot attend so asks Copilot to 'follow' the meeting to receive a recap of what was discussed.

The HR team joins the meeting and see a message from the Manager that they can't attend. They decide that without the Manager, there is no point discussing the matter but, since they have all set aside 30 minutes for the meeting, they decide to use this time instead to discuss another matter – relating to a different member of staff's request for family violence leave.

[5] For guidance on how to interpret and apply the IPP 1, please see OVIC's IPP Guidelines – IPP 1, Collection.

The next day, the Manager sends a prompt to Copilot on Teams asking "what were the main points discussed at the meeting?" Copilot responds with a summary of the details about the other staff member and her request for family violence leave.

The generation of the summary content by Copilot is a collection of personal information which would likely contravene at least IPP 1.1 (as the Manager had no need to know it).

Use and disclosure of personal information

IPP 2 provides that personal information may only be used or disclosed for the primary purpose for which it is was collected, or where an exception applies under IPP 2.1(a) to (h).

Human error is the leading cause of data breaches reported to OVIC involving unauthorised use and disclosure of personal information. Without proper controls in place, the implementation of Copilot could exacerbate this, given the increased speed at which personal information can be generated, used, and disclosed.

Microsoft emphasises that users are in control of what to do with content generated by Copilot. However, there is a risk that humans using a tool described as intelligent may not always exercise an appropriate level of scrutiny over its output.

Further, as Copilot uses documents from an organisation's information holdings to generate content in response to a user prompt, it could provide the user personal information with which they were previously unfamiliar. This creates a risk that VPS employees could be unaware of the original purpose for which the personal information was collected, and cannot properly consider whether a proposed use or disclosure would comply with IPP 2.

Example – Unauthorised disclosure of personal information generated by Copilot

An organisation's Assessment Officer wants to write a letter rejecting Person A's application for a service. They prompt Copilot to draft up this letter and base it off a letter they wrote 2 months ago rejecting Person B's application.

Impressed by how quickly Copilot drafts the 6-page letter, the Assessment Officer quickly reads through and digitally signs it. They then send the letter to Person A by email.

Person A later calls the Complaints Officer and informs them that while the letter was addressed to Person A, it also contained the name of Person B in the body of the letter as well as other personal information relating to Person B's application.

As well as pointing out an interference with Person B's privacy, Person A also suspects that the Assessment Officer may have based her decision on the wrong personal information, and asks that his application is considered afresh.

Accuracy of personal information

IPP 3 requires organisations to take reasonable steps to ensure the personal information it collects, uses or discloses is accurate, complete and up-to-date.

The use of Copilot presents an increased risk of organisations generating, using or disclosing inaccurate personal information. This is because Copilot will not always give you the correct information.[6] In choosing whether to implement Copilot, organisations should assess whether they will be able to mitigate risks associated with Copilot generating inaccurate, incomplete or outdated personal information.

Example – Summary of complaint documents present incomplete picture

Bianca makes a complaint to a Regulator about the actions of a Department. In a 20 page document, she sets out comprehensive details of her allegations, the serious harm she has experienced, and what she wants the Department to do to resolve her concerns, including payment of $20,000 compensation.

The complaint is assigned to a Complaint Officer at the Regulator. After reading the complaint once, the Officer writes a prompt to Copilot asking it to summarise the content into a new document and directs that it should be two pages long. Once generated, the Officer emails the summary to the Department asking for its response to Bianca's complaint.

The Department responds, agreeing that the conduct occurred as alleged but insists that it was a trivial matter that had little impact on Bianca. It offers her $1,000 to resolve the complaint. She rejects this and the Regulator closes the complaint on the basis that it cannot be resolved.

Bianca is disappointed by the Department's approach and asks the Complaint Officer how a government body could read her complaint and still place so little value on what she went through. The Officer explains that it was actually an AI-generated summary of the complaint that was sent to the Department, and provides Bianca with a copy.

[6] Microsoft, *Launch video*, https://news.microsoft.com/reinventing-productivity/ Launch video, https://news.microsoft.com/reinventing-productivity/.

Bianca is aghast when she reads the summary. She believes it has missed out crucial elements of her allegations, and has minimised her description of the harm she suffered. She believes it has created an inaccurate and incomplete description of her complaint, and has led to the Department trivialising her experience and dismissing her complaint.

Security of personal information

IPP 4 requires organisations to take reasonable steps to protect the personal information it holds from misuse and loss and from unauthorised access, modification and disclosure.

The adoption of Copilot could magnify existing risks to the security of personal information for VPS organisations. One particular risk relates to unauthorised access due to Copilot's integration with SharePoint. This is because, in OVIC's experience, access permissions in SharePoint are frequently poorly managed. Misconfiguration of access controls is one of the most common causes of data breaches reported to OVIC.[7]

Organisations should therefore review and update access controls for any tools integrated with Copilot, but especially SharePoint to address the risk of data breaches. Further information on steps organisations should take in relation to security controls is covered in the Information Security considerations below.

Example – Unauthorised access due to misconfigured SharePoint access controls

Rachida is drafting an internal document for a new member of her team explaining the basics of performance development procedures. Looking for inspiration, she prompts Copilot asking about any previous documents about performance development that people have created in her organisation.

Due to an access control misconfiguration that Rachida was unaware of, Copilot sends her a range of documents which includes final copies of the Performance Development Plans of some of her colleagues.

[7] OVIC Incidents Insights Report: 1 July 2022 – 31 December 2022, https://ovic.vic.gov.au/information-security/security-insights/incident-insights-report-1-july-2022-31-december-2022/

Information Security considerations

Before organisations implement any Copilot features, they must ensure their information security programs are of an appropriate maturity.

Irrespective of whether an organisation is subject to Part 4 or 5 of the PDP Act, all organisations should consider the security risks to the confidentiality, integrity and availability of public sector information and systems. This is critical, given the dynamic risks (benefits and drawbacks) Copilot presents VPS organisations.

While all 12 Standards should be considered by organisations, the following Standards are particularly important when considering whether to integrate Copilot into operational environments and should be addressed at a minimum.

Standard 2 – Information Security Value

Standard 2 requires an organisation to identify and assess the security value of public sector information, ensuring that it uses consistent identification and assessment criteria across the information lifecycle. This standard also incorporates the application of protective markings.

Organisations should have procedures that document how the organisation assesses newly generated information by Copilot given the current content and context of the material. This updated assessment must consider any protective markings on the material that was used to generate the new information asset.

Standard 3 – Information Security Risk Management

Standard 3 requires an organisation to utilise its risk management framework to undertake a Security Risk Profile Assessment to manage information security risks. The framework ensures that information security risks are managed through information business decisions while applying controls to protect public sector information.

Integrating Copilot into an organisation's operational environment presents new information security risks, and may further exacerbate known or unknown vulnerabilities in an organisation's controls. It is therefore critical for an organisation to conduct a security risk assessment for the integration of Copilot, to determine the appropriate treatment plans, and ensure the risks identified are monitored and reviewed.

Standard 4 – Information Access

Standard 4 requires an organisation to establish, implement and maintain an access management process for controlling access to public sector information.

A vast number of incidents reported to OVIC under the Information Security Incident Notification Scheme are due to inappropriately applied access controls. There is an increasing number of incidents related to employees having access to information they have no legitimate business need for via SharePoint due to misconfigured access privileges, or failures with access administration.

Using Copilot may exacerbate these issues – given that in response to user prompts, Copilot may use any information holdings to which the user has access. Without Copilot, a user may never realise they have inappropriate access to certain information or act upon it. However, using Copilot, the user may inadvertently use inappropriate content, and may potentially disclose unauthorised material via their Copilot-generated document.

Organisations should therefore take the utmost thoroughness in reviewing and updating their access management controls *prior* to integrating Copilot into their operational environment.

An organisation preparing to integrate Copilot should:

- review its existing access management processes;
- revise how it manages access and requests for access; and
- audit current access privileges, and limit the use of information to the appropriate level.

Standard 5 – Information Security Obligations

Standard 5 seeks to ensure all persons within an organisation understand their responsibilities to protect public sector information.

Before integrating Copilot:

- organisations should develop and deliver information security training specific to Copilot, to promote awareness of user obligations and risks before users are permitted access; and
- initial users, and all subsequent users, should undertake training with a Microsoft Accredited partner and ensure training is targeted to specific roles.

Standard 6 – Information Security Incident Management

Standard 6 requires an organisation to establish and maintain an information security management process and plan relevant to its size, resources, and risk posture.

An organisation should revise its incident management process to ensure it captures the organisation's use of Copilot, before integrating Copilot.

Standard 8 – Third Party Arrangements

Standard 8 requires an organisation that engages a third party to ensure that third party securely collects, holds, manages, uses, discloses or transfers public sector information.

Where an organisation has an awareness that a third party it engages with uses Copilot, it should ensure appropriate controls and information security practices are in place with that third party to protect the organisation from any risks from the third party using Copilot. This is particularly important where a user may have access beyond their organisation into another organisation's systems (for example, cross-department collaboration or information sharing arrangements).

Standard 11 – Information Communications Technology Security

Standard 11 covers the need for organisations to establish, implement and maintain Information Communications Technology (ICT) security controls to protect public sector information and systems.

Before an organisation integrates Copilot, it should:

- conduct a security assessment;
- verify Microsoft's security claims;
- manage and implement appropriate ICT controls for the usage of Copilot; and
- actively monitor Copilot use to detect potential security issues (for example, prompt injection attacks).