Office of the Victorian Information Commissioner

Use of enterprise Generative AI tools in the Victorian public sector

This guidance outlines the Office of the Victorian Information Commissioner's (**OVIC**) minimum expectations for Victorian public sector organisations (**VPS organisations**) using enterprise Generative Artificial Intelligence (**Generative AI**) tools.[1] It also sets out the key information privacy and information security obligations for VPS organisations to consider.

Generative AI is a type of Artificial Intelligence that creates content, such as text, images, music, audio and video. Generative AI tools do this by predicting the best response to the information entered, known as a prompt. Predictions, also known as outputs, are made based on the content of the prompt and the data on which the Generative AI tool has been trained.[2]

This guidance relates to Generative AI tools that are purchased at an enterprise level, and operate within a VPS organisation's securely managed environment. Enterprise Generative AI tools are integrated within an organisation's existing systems. Examples include tenanted versions of Microsoft 365 Copilot, Chat GPT Enterprise, Duet AI for Google Workspace, Zoom IQ and Slack GPT.[3]

The adoption and use of Generative AI tools may magnify existing information privacy and information security risks for VPS organisations, and create new ones (some examples are included below). VPS organisations should therefore take the steps outlined below to improve their information security maturity and privacy practices before adopting an enterprise Generative AI tool.[4]

The risks involved in using publicly available Generative AI tools are different to those involved in using enterprise Generative AI tools. For guidance on using publicly available Generative AI tools, please refer to OVIC's guidance on the use of personal information with publicly available Generative AI tools in the Victorian public sector.

[1] For the purpose of this guidance, VPS organisations are organisations subject to Parts 3 and 4 of the *Privacy and Data Protection Act 2014*. See sections 3 and 84 of that Act for a list of organisations covered by these Parts.

[2] For further information on Generative AI concepts, see the resource developed by the Centre of Excellence for Automated Decision-Making and Society, and OVIC: https://www.admscentre.org.au/genai-concepts/.

[3] This guidance replaces the guidance OVIC issued in its public statement on the use of Microsoft 365 Copilot in the Victorian public sector in October 2023.

[4] OVIC recommends that VPS organisations review OVIC's guidance on engaging third party providers prior to purchasing an enterprise Generative AI tool. The guidance outlines the steps VPS organisations should take to ensure they meet their information privacy and information security obligations: https://ovic.vic.gov.au/privacy/resources-for-organisations/engaging-contracted-service-providers/.

Disclaimer
The information in this document is general in nature and does not constitute legal advice.

OVIC's minimum expectations

This section sets out OVIC's minimum expectations for VPS organisations to work through prior to purchasing and integrating an enterprise Generative AI tool. VPS organisations should:

- Assess the maturity of their existing information security program,[5] including:
  - identifying existing information holdings and systems that may be impacted by the introduction of an integrated Generative AI tool, including due consideration of their security value
  - considering how any Generative AI outputs will be assessed, valued and securely managed, including applying appropriate protective markings for newly generated information assets
  - conducting a security risk assessment for the integration of an enterprise Generative AI tool, including due consideration of adjusted risk profiles and development of treatment plans to address relevant Generative AI features
  - implementing any updated treatment plans by rolling out new or changed controls, including consideration of the organisation's existing logical access controls
  - implementing a formal process for the ongoing monitoring and review of risks and controls. This is especially important given the dynamic development and release of enhancements and new features in Generative AI tools.
- Undertake privacy impact assessments (**PIAs**) to understand the ways in which the enterprise Generative AI tool will be utilised, the risks it presents to the privacy of individuals, whether and how those risks can be mitigated, and in light of any residual risks, the extent to which it is appropriate to proceed to purchase the tool.[6] There may be high-risk functions or specific business units within a VPS organisation for whom it is not appropriate to use a Generative AI tool.[7] Ensure that risks and controls are regularly reviewed and monitored.
- Ensure that enterprise Generative AI tools are not used to make decisions, undertake assessments, or for other administrative functions that may have consequences for individuals or cause them significant harm. It risks unfair decisions being made about them based on information created by Generative AI, that may be inaccurate or of a diminished quality.

[5] See the information security requirements section below.

[6] OVIC has a PIA template and accompanying guide available to assist organisations to undertake a PIA: https://ovic.vic.gov.au/privacy/resources-for-organisations/privacy-impact-assessment/.

[7] It likely will not be sufficient for one overarching PIA to be undertaken on the integration of the enterprise Generative AI tool. PIAs should be conducted for individual functions or uses of the enterprise Generative AI tool, or by business unit, where a business unit will use the Generative AI tool for a discrete, defined purpose.

- Ensure that settings can be configured to enhance information privacy and information security, such as disabling or limiting data sharing between certain platforms and between the VPS organisation and the company that owns the enterprise Generative AI tool.

- Develop organisational policies on the use of Generative AI tools, and communicate about the policies to staff, contracted service providers and other personnel.

- Develop and implement clear and tailored guidance and training for their staff, contracted service providers and other personnel on the use of the enterprise Generative AI tool, that:

    - covers how to safely and responsibly use the enterprise Generative AI tool, prompt engineering, risk assessment and management, human review of generated outputs, and active monitoring and reviews to identify potential misuse

    - ensures all users of the enterprise Generative AI tool are aware of and understand what information they are permitted to enter into the enterprise Generative AI tool and its permitted use cases

    - ensures all users of the enterprise Generative AI tool are aware of how the tool may generate, collect, use or disclose personal information and public sector information, and understand how to ensure the accuracy and security of the information.

- Develop an incident response plan for dealing with inadvertent or unauthorised disclosures, access or misuse of information through the enterprise Generative AI tool. VPS organisations are responsible for all actions and content generated by the enterprise Generative AI tool they use. Where an information security incident or interference with privacy occurs through the use of the tool, VPS organisations will not be able to simply say that the incident or interference was caused by AI.

- Notify OVIC of information privacy and information security incidents, in line with the Information Security Incident Notification Scheme.[8]

Information privacy requirements

This section applies to VPS organisations covered by Part 3 of the *Privacy and Data Protection Act 2014* (**PDP Act**).

[8] Incidents can be reported via OVIC's incident notification form: https://ovic.vic.gov.au/privacy/resources-for-organisations/information-security-and-privacy-incident-notification-form/.

Disclaimer
The information in this document is general in nature and does not constitute legal advice.

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 entered into, generated, or used by an enterprise Generative AI tool.[9]

While all 10 IPPs are relevant to the handling of personal information by an organisation, IPPs 1, 2, 3, 4 and 5 will be most relevant when integrating an enterprise Generative AI tool.[10]

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 to individuals about why the information is being collected, how it will be used and how individuals can access the information.

Organisations should understand that when an enterprise Generative AI tool 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.

Example – Collection of personal information via a collaboration platform

An organisation's HR team invites a Senior Operations Manager to a meeting via a collaboration platform 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 the enterprise Generative AI tool within the collaboration platform to 'follow' the meeting so they can receive a recap of what was discussed.

The HR team joins the meeting and see a message from the manager that they can no longer 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 use the time to instead discuss another matter, relating to a different staff member's request for family violence leave.

The next day, the manager sends a prompt to the enterprise Generative AI tool within the collaboration platform, asking "what were the main points discussed at the meeting?". The enterprise Generative AI tool responds with a summary of the details about the other staff member and their request for family violence leave.

[9] Personal information is defined in section 3 of the *Privacy and Data Protection Act 2014* as 'information or an opinion… that is recorded in any form and whether true or not, about an individual whose identity is apparent, or can reasonably be ascertained, from the information or opinion'. The definition of personal information excludes health information, which is covered by the *Health Records Act 2001*.

[10] For guidance on how to interpret and apply the IPPs, refer to OVIC's Guidelines to the Information Privacy Principles: https://ovic.vic.gov.au/privacy/resources-for-organisations/guidelines-to-the-information-privacy-principles/.

Disclaimer
The information in this document is general in nature and does not constitute legal advice.

The generation of the summary content by the enterprise Generative AI tool is a collection of personal information, which would likely contravene at least IPP 1.1 (as the manager had no need to know that information).

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 IPPs 2.1(a-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 an enterprise Generative AI tool could exacerbate this, given the increased speed at which personal information can be generated, used, and disclosed.

While users are in control of what they do with content generated by an enterprise Generative AI tool, 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 enterprise Generative AI tools use documents from an organisation's information holdings to generate content in response to a user prompt, they could provide the user with personal information with which they were previously unfamiliar (even if access controls are correctly configured and the user has a right to know this information). 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 an enterprise Generative AI tool

An organisation's Assessment Officer wants to write a letter rejecting Person A's application for a service. They prompt their enterprise Generative AI tool to draft this letter and base it off a letter they wrote two months ago rejecting Person B's application.

Impressed by how quickly the enterprise Generative AI tool drafts the six-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.

Disclaimer
The information in this document is general in nature and does not constitute legal advice.

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

The disclosure of Person B's personal information to Person A is a breach of IPP 2. The use of inaccurate personal information about Person A would also likely breach IPP 3, relating to accuracy of personal information (see discussion below). Additionally, this scenario raises concerns about the fairness of the decision made.

Accuracy of personal information

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

The use of an enterprise Generative AI tool presents an increased risk of organisations generating, using or disclosing inaccurate personal information. This is because Generative AI outputs will not always give the correct information. They simply make predictions based on the prompt they receive and the training they've had. Further, where a user re-generates the same prompt, the Generative AI tool likely won't reproduce the original output.

In choosing whether to implement an enterprise Generative AI tool, organisations should assess whether they will be able to mitigate risks associated with the generation of inaccurate, incomplete or outdated personal information.

Example – Summary of complaint documents present incomplete picture

A complainant makes a complaint to a regulator about the actions of a government department. In a 20-page document, the complainant sets out comprehensive details of their allegations, the serious harm they have experienced, and what they want the department to do to resolve their concerns, including payment of $20,000 in compensation.

The complaint is assigned to a Complaint Officer at the regulator's office. After reading the complaint once, the Officer writes a prompt to an enterprise Generative AI tool 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 the complaint.

The department responds, agreeing that the conduct occurred as alleged but insists that it was a trivial matter that had little impact on the complainant. The department offers $1,000 in compensation to resolve the complaint. The complainant rejects this, and the regulator closes the complaint on the basis that it cannot be resolved.

The complainant 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 the complainant with a copy.

The complainant is shocked when they read the summary. They believe it has omitted crucial elements of their allegations, and has minimised their description of the harm suffered. They believe it has created an inaccurate and incomplete description of the complaint, and has led to the department trivialising their experience and dismissing the complaint.

The collection and use of inaccurate personal information about the complainant would likely breach IPP 3, as the regulator has not taken reasonable steps to make sure that the personal information it collects, uses or discloses is accurate and complete.

Security of personal information

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

The adoption of an enterprise Generative AI tool could magnify existing risks to the security of personal information for organisations. One particular risk relates to unauthorised access to information, where access permissions have not been properly configured within other organisational systems that are then integrated with the enterprise Generative AI tool. Misconfiguration of access controls is one of the most common causes of data breaches reported to OVIC.[11] Organisations should review and update access controls for any platforms integrated with enterprise Generative AI tools.

Further information on steps organisations should take in relation to security controls is covered in the next section.

[11] 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/.

Disclaimer
The information in this document is general in nature and does not constitute legal advice.

Example – Unauthorised access due to misconfigured access controls

A manager is drafting an internal document for a new member of their team to explain the basics of performance development procedures. Looking for inspiration, the manager prompts an enterprise Generative AI tool, asking about any previous documents relating to performance development that other staff in the organisation have created.

Due to an access control misconfiguration, of which the manager was unaware, the Generative AI tool sends the manager a range of documents that include final copies of the Performance Development Plans of their colleagues.

The disclosure of the Performance Development Plans of other colleagues to the manager due to misconfigured access controls would likely be a breach of IPP 4.1.

Privacy policies

IPP 5 requires VPS organisations to have a privacy policy, outlining how they handle the personal information they hold. Where personal information is handled by an enterprise Generative AI tool, a VPS organisation's privacy policy should be clear about this. This is especially important where the outputs of an enterprise Generative AI tool are used to inform decision making about an individual, such as whether they are eligible for a service or a payment.

Information security requirements

Before VPS organisations implement any enterprise Generative AI tool, the public sector body head must assess the information security risks to the confidentiality, integrity and availability of public sector information and systems proposed to be used by the enterprise Generative AI tool,[12] and ensure adherence to the Victorian Protective Data Security Standards (**the Standards**).[13]

The Standards set out the mandatory requirements to protect public sector information across all security domains including governance, information, personnel, information communications technology and physical security.

[12] Public sector data, also known as public sector information, is defined in section 3 of the PDP Act as 'information (including personal information) obtained, received or held by an agency or body to which Part 4 applies'.

[13] The Victorian Protective Data Security Standards are available at: https://ovic.vic.gov.au/information-security/standards/.

Disclaimer
The information in this document is general in nature and does not constitute legal advice.

The PDP Act extends these information security obligations to contracted service providers (**CSPs**), requiring public sector body heads to ensure that a CSP does not do an act or engage in a practice that contravenes a Standard.[14]

VPS organisations therefore must ensure their security risk assessment includes an evaluation of CSPs relevant to the enterprise Generative AI tool.

By following this approach organisations:

- will be best placed to understand the risks of an enterprise Generative AI tool within the context of their organisation and risk profile
- will ensure that any changes to their operating environment (people, processes and technology) are incorporated into existing risk assessments
- provide the public sector body head with full visibility and understanding of current risks associated with an enterprise Generative AI tool following any changes.

[14] Section 88(2), PDP Act.

Disclaimer
The information in this document is general in nature and does not constitute legal advice.