information and privacy commission
new south wales

Guide – Privacy risks associated with the use of generative AI tools

May 2026

Who is this information for?	NSW public sector agencies seeking information on generative AI.
Why is this information important to them?	This guide is intended to support agencies in understanding key privacy risks that may arise through the adoption of generative AI tools, as well as suggested mitigations.

Introduction

The *Privacy and Personal Information Protection Act 1998* applies to the handling of personal information through generative artificial intelligence (**Gen AI**) tools. This guidance is intended to assist and support public sector agencies to understand and comply with their privacy obligations in their use and adoption of Gen AI.

Gen AI tools are increasingly being adopted for both private and business-related purposes. These Gen AI tools can offer real benefits: they can provide productivity and efficiency gains for employees and public sector agencies, as well as better outcomes for the broader community. However, these tools also present a range of privacy risks when used to handle personal information.

The purpose of this guide is to highlight key privacy risks and compliance considerations under the *Privacy and Personal Information Protection Act 1998* (**PPIP Act**) and *Health Records and Information Privacy Act 2002* (**HRIP Act**) associated with the adoption of Gen AI by agencies. This guide also suggests measures that agencies should consider implementing to manage these potential privacy risks on an ongoing basis.

Context for this guide

This guide should be read in conjunction with the IPC's Guide to undertaking Privacy Impact Assessments on AI systems and projects.

The guide should also be considered in the context of other NSW Government guidance and frameworks that examine broader information governance and cybersecurity matters, which agencies should refer to in order to holistically address AI-related risks. This includes Digital NSW's NSW Artificial Intelligence Assessment Framework, the Artificial Intelligence Ethics Policy, and guidance regarding the allocation of responsibilities with respect to AI practices.[1]

1. What is generative AI?

Generative AI tools can create text, images or other content in response to a user's instruction or prompt. There are a wide range of potential uses for these tools in a workplace setting, including but not limited to:

- drafting and editing content (such as emails, creative works, or computer code)
- analysis or summary of documents and data
- research and finding information
- translation and transcription, and
- the automation of processes.

Commonly used Gen AI Tools include ChatGPT, Gemini, Claude, Perplexity and Copilot.

[1] Digital NSW, NSW Artificial Intelligence Assessment Framework; Digital NSW, Artificial Intelligence Ethics Policy; Cyber Security NSW, Generative artificial intelligence end-user guidance; Digital NSW, Understanding Responsibilities in AI Practices.

Most Gen AI tools operate using large language models (**LLMs**) or other foundation models that can process both text and non-text-based data types.[2] These models are trained on extensive datasets of text, images and other content ('**training data**') which may comprise billions to trillions of data points.[3]

Once trained on this data, Gen AI tools can produce content in response to a user's prompt. In essence, Gen AI outputs are a prediction of what characters, words or content would best respond to the user's prompt, based on the model's training.[4] As a result, agencies should be mindful that AI systems can produce inaccurate content or reflect biases that might be contained in their training data, even though the outputs may seem authoritative or original (see further, '**What are the privacy risks?**', below).

Generative AI and personal information

As noted, the PPIP Act applies to the handling of personal information. Personal information is defined in the PPIP Act as information or an opinion (including as part of a database and whether or not recorded in a material form) about an individual whose identity is 'apparent' or 'can reasonably be ascertained' from the information or opinion.

Personal information may be handled both through the process of developing Gen AI models (i.e. by the AI developers), as well as through customers' use of the service. This guidance focuses on the use of available Gen AI tools by agencies, rather than the development or training of Gen AI tools.[5]

Different ways that personal information may be processed when an agency uses a Gen AI tool include:

- the prompt or query entered into the Gen AI tool may constitute personal or health information that identifies the employee or others, depending on the nature of the query
- files or data uploaded to the Gen AI tool for analysis may include personal or health information
- the Gen AI tool's output may include generated information that identifies the user or others.

Additionally, some Gen AI providers collect and use customer data (including prompts, uploaded files and outputs) as training data for the AI model, in order to further iterate and improve their products over time.

Generative AI service offerings

There is an increasing range of providers releasing Gen AI tools, including OpenAI, Google, Microsoft, Anthropic and DeepSeek. Most Gen AI providers have different service offerings targeted at either individual or enterprise users:

- **Individual accounts** – Most Gen AI providers offer both free and paid accounts for individual users, with paid accounts offering additional features and higher usage limits.

---

[2] See Digital Platform Regulators Forum (DP-REG), Working Paper 2: Examination of technology – Large Language Models and Working Paper 3: Examination of technology –Multimodal Foundation Models.
[3] Ibid. See also, European Data Protection Supervisor, Guidance on Generative AI, strengthening data protection in a rapidly changing digital era. The training data used to build generative AI models is often sourced by developers from the internet or other publicly available sources using web scraping technologies. However, concerns have also been raised about web scraping from a copyright and privacy perspective. See for example, The Conversation, Generative AI is a minefield for copyright law.
[4] DP-REG, Working Paper 2: Examination of technology – Large Language Models.
[5] For further details about the handling of personal information by generative AI developers and associated privacy risks, refer to the following guidance from the Office of the Australian Information Commissioner (OAIC): Guidance on privacy and developing and training generative AI models.

Often, user data from individual accounts is used to further train the AI model, unless the user opts-out. This is particularly likely where free accounts are used.

- **Enterprise services** – Many Gen AI providers also offer products targeted towards businesses or large enterprises, which offer greater organisational controls to manage employee use of generative AI services. The use of customer data for model training purposes is often turned off by default for enterprise services, but access to internal resources may be broader.

Stronger organisational privacy and security controls are generally available for enterprise services. However, the offerings of each Gen AI provider will differ, and agencies should perform their own investigation and assessment of the benefits and risks of different services before adopting a Gen AI tool.

2. What are the potential privacy risks?

Before adopting any Gen AI tool, agencies need to consider their privacy obligations under the PPIP Act and HRIP Act. The PPIP Act and the HRIP Act contain Information Protection Principles (**IPPs**) and Health Privacy Principles (**HPPs**) that regulate how agencies are permitted to handle personal and health information, which extend to the processing of personal information by a Gen AI tool.[6]

The Gen AI tool related risks included below are not exhaustive and focus primarily on privacy-related risks. Agencies may also need to assess other types of risks that could arise from their use of Gen AI.[7]

Use and disclosure for unexpected purposes

Uploading personal information into a Gen AI tool can lead to the use or disclosure of personal information for purposes that are unrelated to the original purpose of collection, including uses or disclosures that individuals do not expect or are likely to object to. This may occur, for example, where individuals' personal information is included or uploaded in prompts or attachments and is then subsequently included in the AI developer's training data to further improve the AI model.

In addition to eroding public trust, this may lead to uses or disclosures of personal information that are inconsistent with the agency's purpose of collection:

- **IPP 10/HPP 10 (Limited use)** – Agencies must only use personal information for the purpose for which it was collected, with the individual's consent, or where another exception applies.
- **IPPs 11 and 12/HPP 11 (Limited disclosure)** – Agencies must not disclose personal information unless it is directly related to the purpose for collection and the individual would be unlikely to object, or another exception in the IPPs or HPPs applies. In the case of health information or certain sensitive forms of personal information,[8] this would generally be with the individual's consent or where the disclosure is necessary to prevent a serious and imminent threat to life or health.

[6] See, IPC, Fact Sheet - Information Protection Principles (IPPs) for agencies and IPC, Fact Sheet - Health Privacy Principles (HPPs) for agencies.
[7] For example, these could include safety risks, human rights risks (e.g. risk of discrimination), improper administrative decision-making, compliance with other legal requirements.
[8] Section 19(1) of the PPIP Act provides that public sector agencies can only disclose personal information relating to an individual's ethnic or racial origin, political opinions, religious or philosophical beliefs, trade union membership or sexual activities where the disclosure is necessary to prevent a serious and imminent threat to the life or health of the individual concerned or another person.

Cross-border disclosures of personal information

Depending on the Gen AI provider to be used, agencies may also need to consider the restrictions on the disclosure of personal and health information to entities outside New South Wales under section 19(2) of the PPIP Act and section 14 of the HRIP Act.[9] Under these provisions, public sector agencies must not disclose personal or health information to parties located outside New South Wales or to a Commonwealth agency unless:

- the public sector agency reasonably believes that the recipient is subject to a law, binding scheme or contract that effectively upholds principles for fair handling of the information that are substantially similar to the Information Protection Principles or the Health Privacy Principles, or
- the individual expressly consents to the disclosure, or
- another exception under section 19(2) of the PPIP Act or section 14 of the HRIP Act applies.

Most major Gen AI providers are established in jurisdictions outside of New South Wales and may process personal information offshore. They may also be subject to different privacy and data protection laws. As a result, public sector agencies may be required to ensure that the cross-border data flow requirements in the PPIP and HRIP Acts are met for any personal or health information that is uploaded into a Gen AI service.

Unauthorised use or disclosure and risk of data breach

Information uploaded to Gen AI tools can resurface publicly in several different ways. For example:

- Gen AI tools can reveal extracts of its training data in response to prompts by other users, including any personal information contained within that training dataset. This could occur through routine use of a Gen AI service,[10] or in response to prompt attacks, through which malicious actors attempt to gain access to the information on which an AI system was trained.[11]
- There have been several instances where user chats have been publicly indexed on search engines.[12]

The unauthorised use or disclosure of personal information through these means could result in a range of privacy harms to the individuals concerned including safety risks, identity theft, or personal embarrassment, depending on the nature of the personal information involved.

In addition, this may constitute a notifiable data breach under Part 6A of the PPIP Act. For further details on the Mandatory Notification of Data Breaches Scheme, refer to the IPC's *Guide to managing data breaches in accordance with the PPIP Act*.[13]

---

[9] For further details, refer to IPC, Guideline - Transborder Disclosure Principle - section 19(2).
[10] ABC, AI chatbot blamed for psychosocial workplace training gaffe at Bunbury prison.
[11] See, Confederation of European Data Protection Organizations, Generative AI: The Data Protection Implications, 12; Information Commissioner's Office (UK), How should we assess security and data minimisation in AI?; Government Digital Service (UK), AI Insights: Prompt Risks.
[12] See for example, BBC, Hundreds of thousands of Grok chats exposed in Google results; The Telegraph, Private ChatGPT conversations leak onto Google search results.
[13] IPC, Guide - Mandatory Notification of Data Breach Scheme: Guide to managing data breaches in accordance with the PPIP Act.

Case Study 1 – Generative AI data breach

An employee is using a generative AI tool to prepare a report about how the public sector agency's services support individuals with mental health issues.

The employee wants to include a hypothetical case study that explains the standard customer journey when engaging with the agency. To inform the generation of case studies, the employee uploads real-life reports to the AI tool, which include information about individuals' mental health conditions. The employee uses their personal account on the generative AI service to do so. By default, the generative AI provider collects this information and includes it in the AI model's training data.

Months later, the generative AI model resurfaces these real-life reports to a researcher who is also using the service, including the specific details about the individuals' mental health conditions.

Extended retention of personal information

Uploading personal information to a Gen AI tool may lead to the extended retention of that information by the provider. Additionally, there may be practical challenges surrounding the erasure of personal information contained in the model itself. For example, the UK Information Commissioner's Office has observed that erasure of an individual's personal data from a model "may not be possible without re-training the model (either with the rectified data, or without the erased data) or deleting the model altogether."[14]

The extended retention of data may increase the likelihood that personal information is subject to unauthorised use or disclosure, as noted above.

Generation of new personal information

Generative AI may be used to create new personal information, such as where the AI tool is asked to perform analysis or generate inferences about an individual.

These inferences and outputs may fall within the definition of personal information to the extent that it constitutes "information or an opinion about an individual whose identity is apparent or can reasonably be ascertained from the information or opinion."[15]

The creation of this data would need to comply with the applicable IPPs or HPPs for the collection of personal information, including:

- **IPP 1/HPP 1 (Lawful purpose)** – Agencies must only collect personal information where it is reasonably necessary for a lawful purpose that is directly related to a function or activity of the agency
- **IPP 2/HPP 3 (Direct collection)** – Agencies should collect personal information directly from the individual concerned, unless an exception applies; and
- **IPP 4/HPP 2 (Relevant)** – Agencies should take reasonable steps to ensure that collected information is relevant, not excessive, accurate and does not intrude to an unreasonable extent on the personal affairs of the individual.

The generation of new personal information may also cause privacy harms where the created information is inaccurate, or where individuals are not provided with appropriate notice about the use of AI.

[14] Information Commissioner's Office (UK), How do we ensure individual rights in our AI systems?; OAIC, GenAI tools in the workplace: balancing protection of personal information and business efficiency.
[15] PPIP Act s 4.

Generation of inaccurate, unfair or discriminatory outputs

Agencies should be aware that AI-based technologies can be prone to producing inaccurate, out-of-date, unfair or discriminatory outputs. This may happen because:

- some models may have been built on incomplete training data or data that reflects systemic social inequalities or biases
- AI models generally produce outputs that are the statistically 'most correct' output based on its training data, rather than facts
- the accuracy and reliability of AI models may be susceptible to deterioration over time.[16]

Additionally, even when a model produces an error or false information, it often presents that information in way that appears highly credible.

This risk may be particularly high where agency employees use Gen AI tools to assist with making decisions that could have a significant effect on individuals, such as making hiring decisions or determining an individual's entitlement to services or support.

From a privacy perspective, this also raises compliance questions regarding the accuracy of personal information that is collected, generated or used through generative AI tools under **IPP 4/HPP 2 (Relevant)** and **IPP 9/HPP 9 (Accuracy).** These IPPs require agencies to take reasonable steps to ensure that personal information is relevant, accurate, up to date, complete and not misleading.

Case Study 2 – Use of irrelevant and inaccurate personal information

An agency employee wants to use a generative AI tool to assess applications for the agency's entrepreneurship grants. To do so, the employee uploads each individual's written application, as well as a spreadsheet containing the personal information of all applicants.

In their prompt, the employee asks the Gen AI tool to summarise each application and compare the strength of the applications to identify preferred candidates.

In preparing these summaries and recommendations, the Gen AI tool considers irrelevant matters and over-emphasises the strengths of certain demographic groups. Additionally, the system does not apply the selection criteria set out in the grants application process. The outputs of the system therefore contain irrelevant and inaccurate information.

The agency employee relies on the recommendations of the Gen AI tool to approve the grants and does not manually validate the accuracy of its outputs. As a result, several highly qualified applicants are denied the grants, causing them to miss out on potential financial benefits. In addition, the personal information of all applicants has now been disclosed to the generative AI provider without their knowledge.

Lack of transparency and loss of data subject control

Individuals may be unaware that their personal information is being handled through generative AI tools if this is not clearly outlined in the agency's privacy policy or collection notice. Lack of transparency can erode the public's trust in the agency as a responsible custodian of personal information. **IPP 3 and HPP 4** require agencies to inform individuals about the purposes for which their information is being collected, among other matters.

In a desktop review of NSW public sector agency Privacy Management Plans (**PMPs**), it was found that only 13% of PMPs provided a direct or indirect reference to the agency's use of AI or automated decision-making.[17]

[16] See relatedly, OAIC, Guidance on privacy and the use of commercially available AI products.
[17] See, IPC, Desktop Review of Documented AI or ADM Use within AIGs and PMPs.

Handling personal information through Gen AI tools may also affect individuals' ability to exercise control over their information, particularly where their information is incorporated into the AI model's training data or where AI is used to generate new inferences about them.

3. Recommended measures

There are a range of measures that agencies should consider in order to mitigate privacy risks that may arise through the adoption of generative AI tools.

The following privacy risk mitigations should be considered within the broader framework of AI and information governance and cybersecurity controls. These measures should be applied holistically to address AI-related risks, including effective records classification and management, and clear allocation of responsibilities for AI practices within the agency.[18]

Privacy impact assessments	If a generative AI system will be used within the agency to handle personal information, a privacy impact assessment (**PIA**) should be undertaken. PIAs are an important means by which agencies can build privacy into the design of their processes and projects (also known as a 'privacy by design' approach) and should identify: • potential positive and adverse privacy impacts, including alignment with community expectations • compliance with privacy laws and other relevant legislation • measures to reduce any identified risks to privacy. For further details, agencies should refer to the IPC's Guide to undertaking Privacy Impact Assessments on AI systems and projects.
Review privacy management plan and privacy policies and processes (e.g. data breach response plans)	Agencies should also review their privacy management plans and update them to account for any agency use of AI, as needed. Under section 33 of the PPIP Act, agencies must maintain privacy management plans that explain the agency's policies and practices for complying with the PPIP Act and HRIP Act, among other matters. Agencies are encouraged to review their PMPs on an annual basis. For further details, agencies should refer to the IPC's Guide to making privacy management plans. Agencies should also review and update their key organisational privacy policies and processes to ensure that they explicitly account for AI-related scenarios. For example, agencies' data breach response plans and policies should account for the inadvertent or unauthorised disclosure, access or loss of personal information through AI tools. Finally, senior leadership within the agency should consider the importance of setting a strong strategic direction towards the responsible and privacy-protective use of GenAI tools within the agency. This should include necessary uplift programs to achieve this, as well as the setting of responsibilities in relation to responsible AI adoption.

[18] Resources to support broader AI governance and risk management include Digital NSW, NSW Artificial Intelligence Assessment Framework; Digital NSW, Artificial Intelligence Ethics Policy; Cyber Security NSW, Generative artificial intelligence end-user guidance; Digital NSW, Understanding Responsibilities in AI Practices.

Transparency measures and consent	Agencies should ensure that their privacy policies or notices to the public are clear regarding any handling of personal information through AI tools, as well as how individuals can access human review or intervention. Agencies could also consider the use of 'just-in-time' notices to deliver this information at the time that information is collected from the individual.[19] Agencies should also ensure that consent is obtained from individuals if AI tools will be used to handle their personal information, where required under the PPIP or HRIP.
Introduce an AI acceptable use policy	Agencies should consider introducing policies regarding acceptable employee use of generative AI tools at work, either within or in addition to their existing IT Acceptable Use Policies. For example, these policies could govern employee use of generative AI at work in the following ways: • **Approved tools** – Restricting use to agency-approved generative AI tools, using authorised work accounts (as opposed to an employee's personal or individual account) • **Prohibited uses** – Prohibiting the uploading of personal or health information to generative AI tools by staff without prior review and approval by the agency's privacy team. Employees should use only approved agency tools. Agencies should be clear and specific about the use of shadow tools within their environments. Prior to approving a particular tool or use of generative AI within the agency, a PIA should generally be performed to understand the potential risks to individuals' personal information and to ensure that those risks are managed (see above). Finally, agencies should consider the role of effective information governance practices, including data classification and labelling, in order for IT teams to effectively enforce prohibitions in an AI acceptable use policy (e.g. prohibiting the uploading of personal information).
Staff training	Agencies should also implement necessary training and guidance for employees regarding appropriate use of generative AI tools, and relevant risks so that they are aware of associated privacy risks and the agency's policies.
Pre-deployment testing	Prior to deployment, agencies should assess whether the AI tool will be suitable for its intended use cases. Agencies should ensure that they understand how the product works, any limitations of the AI tool, and the accuracy of its outputs in the context in which it will operate. Consideration should also be given to testing for bias and discrimination. These matters may be considered as part of the agency's PIA for the proposed project.

[19] For further details, see Information Commissioner's Office (UK), What methods can we use to provide privacy information?

Third party assessments	Where agencies intend to use a third-party generative AI tool, they should assess the vendor's privacy controls, contractual safeguards and overall alignment with the standard of protection afforded by the PPIP Act and HRIP Act. This assessment should provide clarity regarding the data storage arrangements and data flows that take place when using the third party's services.
	
	Agencies should ensure that generative AI providers offer:
	
	• enterprise-grade generative AI services and associated organisational controls
	
	• contractual terms which:
	
	  ○ clearly prohibit the vendor's use of customer data (including personal information) to further train the AI model or for their own commercial purposes
	
	  ○ bind the vendor to comply with the PPIP Act or a similar level of privacy protection, including data breach notification obligations that are aligned with Part 6A of the PPIP Act
	
	• defined data retention and deletion timeframes
	
	• audit and right to terminate clauses where privacy obligations are not met
	
	• residency of customer data within Australia, where possible.

Human in the loop	During use, agencies should ensure that there is access to a person to review or discuss the outputs of Gen AI tools. Embedding a human-in-the-loop for these processes is crucial to provide accountability for decisions and to mitigate against the risk of inaccuracy or bias. Human review is particularly important in high-risk settings, such as where the outputs of an AI tool inform decisions that an agency makes about individuals (e.g. decisions with legal consequences, or that affect an individual's access to a service, opportunity or benefit). In such cases, agencies should proceed cautiously and consider whether the potential risks outweigh the benefits.
Ongoing monitoring and reinforcement	Agencies should not take a 'set and forget' approach to these measures. Instead, they should continuously monitor and review the use of generative AI tools within the agency, culture and governance, and any emerging privacy risks and adjust their policies and procedures accordingly. Agencies should also consider regularly reinforcing the importance of these measures to employees, including the potential privacy harms that may occur from improper use of GenAI, to ensure that a strong culture of privacy-protective use of Gen AI tools is established.

4. Glossary

The following definitions may assist agencies to understand and assess the privacy risks associated with the use of generative AI:

Term	Definition
Artificial intelligence ('AI') system	A machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments
Data breach	A data breach occurs when information held by an agency is subject to unauthorised access, unauthorised disclosure or is lost in circumstances where the loss is likely to result in unauthorised access or unauthorised disclosure.
Generative AI	A wide-ranging term that refers to any form of AI system that is capable of generating new content, including text, images, video, audio, or code. Easily accessible examples of Gen AI include ChatGPT (OpenAI), BardAI (Google), MidJourney, and CoPilot (Microsoft). These tools allow individuals to input text and receive AI-generated content. They offer functionalities such as summarising lengthy articles, providing concise answers to questions, or generating code snippets for described functions.
Human-in-the-loop	A process in which a human plays an active and integral role in decision-making, monitoring, and control over an AI system that produces outcomes.
Large language models ('LLMs')	A subset of generative AI models that specialise in generating human-like text.
Machine learning	A subset of AI that allows computers to autonomously learn and improve without being explicitly programmed. Machine learning algorithms are trained on data to make predictions or decisions.
Personal information	Information or an opinion (including information or an opinion forming part of a database and whether or not recorded in a material form) about an individual whose identity is apparent or can reasonably be ascertained from the information or opinion. Personal information includes such things as an individual's fingerprints, retina prints, body samples or genetic characteristics. See PPIP Act, Section 4.
Prompt	An instruction provided to a generative AI tool by the user of the service.
Training data	The extensive datasets that AI developers use to train models, so that generative AI systems are able to predict what characters and words would best answer a prompt, or the content that must be generated.

Further information on commonly used AI terms can be found on the following Digital NSW website: A common understanding: simplified AI definitions from leading standards.

5. Additional resources

For more information on the privacy risks relating to AI tools and recommended measures to manage them, agencies should refer to the following:

- IPC, Guide to undertaking Privacy Impact Assessments on AI systems and projects
- IPC, Desktop Review of Documented AI or ADM Use within AIGs and PMPs
- IPC, Guide to Privacy Impact Assessments in NSW
- Digital NSW, NSW Artificial Intelligence Assessment Framework
- Digital NSW, Artificial Intelligence Ethics Policy
- Cyber Security NSW, Generative artificial intelligence end-user guidance
- Digital NSW, Understanding Responsibilities in AI Practices
- Department of Industry, Science and Resources, Voluntary AI Safety Standard
- Digital Platform Regulators Forum (DP-REG), Working Paper 2: Examination of technology – Large Language Models
- DP-REG, Working Paper 3: Examination of technology –Multimodal Foundation Models

For more information

Contact the Information and Privacy Commission NSW (IPC):

Freecall: 1800 472 679
Email: ipcinfo@ipc.nsw.gov.au
Website: http://www.ipc.nsw.gov.au/

NOTE: The information in this guideline / guide / report is to be used as a guide only. Legal advice should be sought in relation to individual circumstances.

Document information

Identifier/Title:	Privacy Risks associated with the use of generative AI tools
Business Unit:	IPC
Author:	Privacy Commissioner
Approver:	Privacy Commissioner
Date of Effect:	14 May 2026
Next Review Date:	14 May 2028
EDRMS File Reference:	D26/010786/DJ
Key Words:	Generative AI, GenAI, Privacy, risks, personal information