Office of the Victorian Information Commissioner

Use of personal information with ChatGPT

This is a public statement made by the Privacy and Data Protection Deputy Commissioner under section 8C(1)(f) of the *Privacy and Data Protection Act 2014* (Vic) (**PDP Act**).

This statement relates to the use of the Chat Generative Pre-Trained Transformer (**ChatGPT**) platform by Victorian public sector (**VPS**) organisations.

ChatGPT is an example of generative artificial intelligence (**genAI**). The considerations outlined in this statement should also be observed when utilising other forms of genAI.

Can personal information be used with ChatGPT?

VPS organisations must ensure staff and contracted service providers do not use personal information[1] with ChatGPT.

ChatGPT must not be used to formulate decisions, undertake assessments, or used for other administrative actions that may have consequences for individuals, for example, evaluations, assessments, or reviews. Doing so is a contravention of the Information Privacy Principles (**IPPs**), and may cause significant harm to individuals whose information is used with ChatGPT.

If an organisation becomes aware that personal information has been used with ChatGPT it should treat the occurrence as an information security incident and notify OVIC immediately.[2]

What is ChatGPT?

ChatGPT is a genAI platform developed by OpenAI that uses natural language processing to respond to a prompt from a user and generate human-like text, known as an 'output'.

It uses publicly available information to train its Large-Language Model (**LLM**) to detect patterns, context and meaning, and uses this to generate outputs to respond to a user's prompt.

Of particular concern, ChatGPT's LLM is also trained on any information that is entered by a user. This includes any prompts, messages, conversations, files, or documents uploaded to ChatGPT, or any feedback provided about the use of ChatGPT.

[1] 'Personal information' is defined in section 3 of the PDP Act as information or an opinion (including information or an opinion forming part of a database), 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.

[2] OVIC can be contacted by submitting an Incident Notification form through OVIC's website or via enquiries@ovic.vic.gov.au.

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

How does ChatGPT work?

ChatGPT generates conversational content following the direction of a prompt by a user. ChatGPT works by determining which words, phrases and sentences are associated with the prompt. ChatGPT's LLM then predicts what word is most likely to come next based off the LLM's training of the source materials' patterns, context, and meaning. ChatGPT then generates an output most likely to be related to the prompt, word by word.

What are the privacy risks when using ChatGPT?

Using personal information with ChatGPT raises significant privacy concerns and will contravene several IPPs:

- Using personal information with ChatGPT means disclosing that information to OpenAI. The information may then be subsequently used or accessed for unauthorised purposes by individuals outside of your organisation, outside of Victoria, and other third parties related to OpenAI, in contravention of IPPs 2.1, 4.1 and 9.

- Generating personal information with ChatGPT constitutes a new 'collection' of personal information, which may not be necessary, lawful, or fair and may result in inaccurate information, or opinions, being generated and subsequently used or disclosed, in contravention of IPPs 1.1, 1.2, 3.1 and 10.

- Personal information input into ChatGPT is indefinitely retained by OpenAI, in contravention of IPP 4.2 and an organisations' obligations under the *Public Records Act 1973*.

Further to this, the use of ChatGPT with personal information may cause significant harm to those individuals whose information has been used with ChatGPT. It risks unfair decisions being made about them based on information that may be inaccurate or of a diminished quality generated by ChatGPT.

What if input history and model training is disabled?

ChatGPT provides users with the ability to disable inputs from being recorded, and to opt-out of ChatGPT using this information to train OpenAI's LLM.

Any information entered prior to enabling these features will still be used to train OpenAI's LLM, and be retained by OpenAI indefinitely. Any information used after disabling these features will still be retained by OpenAI for 30 days, and may be reviewed by OpenAI for monitoring of inappropriate behaviour.

Reliance on these features requires users to be appropriately educated and trained on the use of ChatGPT, and requires active monitoring and assurance. For example, where a user disables input history and model training, these settings do not automatically sync across all browsers and devices.

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

Can other public sector information be used with ChatGPT?

VPS organisations should ensure their personnel (staff, contractors, volunteers, etc.) and contracted service providers limit the use of ChatGPT to public sector information that is already publicly known, or if disclosed would not cause any harm to an individual or damage to an organisation.

Scenarios

Scenario 1 – Inputting and generating content

A Manager is using ChatGPT to generate content for a recruitment selection report. They input the CVs and interview notes captured by the panel, and the reference reports of five candidates into ChatGPT. They prompt ChatGPT to generate six paragraphs recommending Candidate A for the role.

ChatGPT generates the output that it predicts will best respond to the prompt. Impressed by how quickly ChatGPT drafts the content for the selection report, the Manager quickly skims through the content, digitally signs it, and sends it for approval.

Trainers at OpenAI subsequently review the content and input used with ChatGPT by the Manager to improve the LLM, disclosing the personal information of five individuals in contravention of IPP 2.1 and 4.1. Further to this, the content is stored outside of Victoria, in contravention of IPP 9, and retained indefinitely by OpenAI in contravention of IPP 4.2.

Scenario 2 – Using ChatGPT for decision making

A VPS employee is writing a report evaluating whether a Prisoner should be granted parole, and uses ChatGPT to generate the content of the report, including the evaluation of risks. In doing so, they input the Prisoner's personal information.

ChatGPT generates an output in response to the prompt. The employee does not carry out their own independent evaluation and relies solely on the information contained in the output. The employee then submits the recommendation that the Prisoner's parole application is rejected to their supervisor.

The employee's supervisor reviews the report and notes that the reasoning for the opinion rejecting the parole application is flawed, and that the wrong recommendation had been reached. The supervisor rewrote the report themselves, undertaking analysis of the information to ensure the accuracy of the information relied upon and resubmitted this.

Disclaimer

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

If it had not been for the supervisor's intervention, the use of ChatGPT could have had severe negative consequences for the Prisoner. The use of the personal and sensitive information of the Prisoner would contravene IPPs 2.1 and 3.1 at a minimum.

Scenario 3 – Using ChatGPT in public schools

A teacher uses ChatGPT to generate an email to a student's parents expressing their concerns about the student's behaviour and academic performance. In doing so, the teacher inputs the notes they have taken about the student across the term to generate the output.

The teacher isn't happy with the tone of the output so regenerates it again. In doing so, ChatGPT generates a new output with a different tone and different opinion about the student. The teacher then sends the content in an email to the student's parents.

The parents receive the email and raise concerns with the teacher about the opinions formed about their child, and question who wrote the letter. The teacher explains they used ChatGPT to generate the content. The parents are not satisfied the generation of the content about their child was necessary, and consider the use of ChatGPT to be an unreasonably intrusive method of collecting information about their child.

As generating personal information using ChatGPT is considered a new 'collection', there has been a contravention of IPPs 1.1, 1.2 and 1.3.

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