Office of the Victorian Information Commissioner

ARTIFICIAL INTELLIGENCE

UNDERSTANDING PRIVACY OBLIGATIONS

CONTENTS

INTRODUCTION

Many artificial intelligence (**AI**) technologies rely on enormous amounts of data – which may include personal information – in order to train and test algorithms. When Victorian public sector (**VPS**) organisations collect personal information to train an AI model, feed personal information into an AI system, or use AI to infer information about individuals, the Information Privacy Principles (**IPPs**) of the *Privacy and Data Protection Act 2014* (**PDP Act**) apply.

The purpose of this guidance is to assist VPS organisations to consider their privacy obligations when using or considering the use of personal information in AI systems or applications. It will cover the collection, use, handling and governance of personal information within this context. Organisations should also conduct a privacy impact assessment (**PIA**) when designing or implementing AI systems, to help identify potential privacy risks associated with the collection and use of personal information in the AI system. PIAs are discussed later in this guidance.

What is AI? Artificial intelligence, or '**AI**', is a way for computers to perform tasks that require abstraction and which would ordinarily be performed by humans. AI is used as an umbrella term to describe a collection of different techniques and technologies, including machine learning, speech recognition, natural language processing, robotics, and predictive analytics. AI is present in many of the day-to-day interactions in our personal lives – for example, when we give voice commands on our mobile phones, or the movie recommendations on streaming services.

The use of AI applications and systems is also growing in the public sector, enabled by the generation, availability and variety of sources of data accessible to government. Organisations are increasingly turning to AI to help carry out their functions, automate decision making processes, inform policy, and deliver services to the public. Common applications of AI include identifying objects, making predictions, translating language and processing very large amounts of information. For example, an increasingly common use of AI in the public sector is the use of chat bots to provide customer service and advice to individuals on a website.

While the use of an AI system to process personal information can deliver significant benefits, VPS organisations should consider whether the deployment of such a system is necessary to address an identified problem, and whether it is the best solution to that problem – AI systems should not necessarily be deployed simply because they are available.

This is particularly important given the potential risks associated with the use of an AI system, including the risk of discrimination, bias and inequality. The use of AI to inform decision making, for example, can have long lasting and significant impacts on individuals' lives and human rights, such as where AI is used to help assess the risk of a person recommitting an offence, or to assist in determining eligibility for welfare services or payments.

Ensuring that individuals' privacy rights are upheld in the context of an AI system is therefore crucial, in order to mitigate some of the potential risks that may arise.[1] VPS organisations using AI systems or applications are also obliged to consider and act compatibly with the *Charter of Human Rights and Responsibilities Act 2006* (Vic), which includes the right to privacy.

COLLECTION

Collection is the way in which an organisation comes into possession of personal information. Traditionally, this involves an organisation collecting information from individuals themselves (direct collection), or from a third party (indirect collection).

In the context of AI, organisations may collect personal information for different purposes – for example, using personal information to train an AI model, or feeding personal information into an AI system to produce an output, such as patterns or trends. AI systems can also generate personal information that did not previously exist, constituting a new collection of personal information. For example, an AI system may infer personal information about individuals (such as their age or gender) based on existing input data (such as movie preferences or spending habits). Under the PDP Act, whether personal information is true or not is irrelevant.[2] Organisations that infer personal information are, for the purposes of the IPPs, collecting that information indirectly. Inferring personal information from an AI system does not involve direct collection from an individual, and for the purposes of IPP 1, is therefore akin to collecting personal information from a third party who is a natural person.

There are several considerations that VPS organisations must bear in mind to ensure that the collection of personal information (inferred or otherwise) for the purposes of AI systems or applications complies with the IPPs.

Personal information must be necessary

AI can require large amounts of information (including personal information), for example for training machine learning models. While more data is often regarded as better in the context of AI, IPP 1.1 requires VPS organisations to only collect personal information that is necessary for one or more activities or functions. Organisations must therefore limit the collection and size of datasets containing personal information to what is actually necessary for the purpose or function that the AI system is designed to carry out. This also means that organisations cannot collect personal information just in case it could be useful for analytics or other AI applications in the future.

Additionally, before collecting personal information organisations should consider if de-identified or anonymous information could be used to achieve the same results. Alternatively, organisations could also consider whether synthetic data and datasets could be used instead of collecting data relating to real people. Synthetic data replicates the statistical components of real-world data, however it does not contain identifiers and is artificially generated. Machine learning models could also encode general patterns, rather than facts about specific training examples, which may contain personal information about individuals.

This principle of necessity also applies to AI systems that use live data. If data is continuously fed into an AI system, such as one that is used to identify emergency situations on social media posts, the collection of personal information should be limited to what is necessary. This may mean excluding names or other identifiers.

Ultimately, applying IPP 1.1 in practice may be difficult, as it can be challenging to know what the outputs of an AI system will be before data is actually used, or to evaluate the quality or potential of data before it is collected, and whether it is indeed necessary. Where an organisation does collect personal information that later turns out is not necessary for the purposes of the AI system or application, or any other purpose, IPP 4.2 requires that reasonable steps be taken to destroy or permanently de-identify that information (subject to any recordkeeping obligations).

Collection must be lawful, fair and not unreasonably intrusive

IPP 1.2 requires organisations to have the legal authority to collect personal information, and to collect personal information in a manner that is fair and not unreasonably intrusive.

Inferring personal information is inherently less fair than collecting it directly. When personal information is collected from an individual, that individual is aware that the collection is taking place. Direct collection also means that details about the collection can be provided to the individual, such as the purposes for which the organisation will use their information. On the other hand, with inference and other methods of indirect collection, individuals may not always be aware that their personal information is being collected, and therefore do not have the opportunity to choose whether or not to provide it (where there is a choice to do so or not).

Indirect collection

Organisations inferring – and therefore indirectly collecting – personal information about individuals through AI systems must consider IPP 1.4. Under IPP 1.4, organisations are required to collect personal information only from the individual to whom it relates, where it is reasonable and practicable to do so. This means that organisations cannot infer personal information if it would be reasonable and practicable to collect it directly from the individual.

IPP 1.5 will also apply in this context where personal information is inferred by an AI. IPP 1.5 requires organisations to take reasonable steps to ensure that, when personal information is indirectly collected, individuals are aware of a range of matters listed in IPP 1.3:[3]

- the name of the organisation and how to contact it;
- the fact that the individual is able to gain access to the information;
- the purposes for which the information is collected;
- to whom (or the types of information or organisations to which) the organisation usually discloses information of that kind;
- any law that requires the particular information to be collected; and
- the main consequences (if any) for the individual if all or part of the information is not provided.

Sensitive information

IPP 10 places limitations around the circumstances in which sensitive information can be collected. Under the PDP Act, sensitive information includes information about individuals' racial or ethnic origin, philosophical beliefs, and political opinions.[4] Organisations generally cannot collect sensitive information without the consent of the individual to whom it relates. However, given that obtaining consent is not feasible when personal information is indirectly collected, organisations using an AI system or application that infers sensitive information about individuals must ensure that the collection is authorised by another exception under IPP 10.

Organisations will also still need to comply with IPP 1 when collecting sensitive information through inference – the sensitive information must be necessary for the organisation's activities or functions, and only the minimum amount of sensitive information needed should be collected.

Publicly available information

Organisations may use data from a wide range of sources for their AI systems and applications, including personal information that is publicly available (for example, a public register). Under section 12 of the PDP Act, 'generally available publications' are exempt from the IPPs.[5] This means that personal information found online may be exempt from the IPPs in some circumstances.

However, organisations should give careful consideration when relying on this exemption, as whether information that is publicly available can be considered a 'generally available publication' will depend upon the context in which the information appears. For example, a publicly available post on social media will not necessarily constitute a publicly available publication.[6]

USE

AI systems can involve the use of personal information in different ways, and for different purposes. The purposes for which an AI uses personal information – whether it is information already held by an organisation, or a new collection of information – can vary over time and across different stages over the lifetime of the AI system. For example, in the initial stages of implementing an AI system, personal information may be used to train the model, and an organisation may rely on one legal authority to use the information for this purpose. However, this legal authority may no longer be relevant or applicable at another stage, such as once the AI system has been deployed, or the AI uses personal information in new, unexpected ways. Where an AI system uses personal information for several purposes, VPS organisations should identify each purpose – across all stages of the AI system's lifecycle – and ensure that there is a lawful basis for each one under the PDP Act or other legislation.

Under IPP 2, personal information must only be used for the primary purpose of collection, unless an exception applies. However, the nature of AI means that it can be challenging for organisations to know ahead of time how an AI system will use information in the future, which can often be for new, unexpected or unanticipated purposes beyond the original purpose of collection.

In general, where AI systems and applications are not using personal information for the primary purpose, the use for a secondary purpose may be authorised where it is related to the primary purpose and would be reasonably expected by individuals (IPP 2.1(a)). For example, an AI may use personal information for the primary purpose of personalising a service that an organisation provides to individuals; however, the AI using that same information to develop a new service that is not related to the primary purpose, nor reasonably expected by individuals, would likely not be permitted under IPP 2.1(a).

IPP 2 also allows for secondary uses under other grounds, for example with the consent of an individual (IPP 2.1(b)), or where the use is required or authorised by or under law (IPP 2.1(f)). However, given the challenges of obtaining meaningful consent (discussed further below), organisations should consider the feasibility of using consent as the legal basis for secondary use of personal information in AI systems.

Organisations should refer to the relevant collection notice when designing or implementing a new way of using personal information. If an organisation uses personal information for a new purpose, such as AI analytics, the organisation should update any relevant collection notices and privacy policies to reflect the new use or uses of personal information. However, organisations should be mindful that collection notices do not provide the legal authority to use personal information for the purposes of AI systems or processing – the lawful basis must be established under IPP 2 or another legislation.

If practicable, organisations should also notify individuals of any new purposes. While this is not required under the IPPs, being transparent and upfront in how personal information is handled will help to maintain public trust in the organisation and minimise the risk of privacy complaints.

CONSENT

As noted above, consent is one legal basis that permits personal information to be used for a secondary purpose (under IPP 2.1(b)). Consent is also relevant to other IPPs such as IPP 10, where it is another legal basis to allow the collection of sensitive information.

Seeking individuals' consent for the processing of their personal information can have benefits; for example, it can help build trust and buy-in from individuals. However, there are many challenges with relying on consent, particularly in the context of AI. Meaningful consent requires individuals to be fully informed about how their personal information will be used. The transparency issues discussed above can make this difficult. If an individual does not understand how an AI system will operate or how their personal information will be used, their ability to give informed consent decreases.

Consent must also be voluntary, and therefore revocable. For example, if an individual consented to their personal information being used to train an AI model, and their information could be extracted from that model, then that individual should be able to withdraw their consent. This could then mean that the AI model may need to be discarded and retrained with a dataset that does not contain that individual's information.

In addition to being voluntary, meaningful consent must also be current, specific, and given by someone with capacity. This could also pose difficulties, such as where a minor's consent is required to process their personal information, or where consent is sought for a broad range of purposes (bundled consent) rather than for each specific use.[7]

Given the challenges with obtaining meaningful consent in the context of AI, organisations should try to avoid relying on consent as the legal basis for using personal information with AI systems.

GOVERNANCE

AI informed decision making

'AI informed decision making' refers to both AI systems that produce information that is then used by humans to make decisions, and AI systems that are capable of making decisions autonomously without human oversight.

However, not all government decisions can be legally made by AI. Some decisions stem from a legislative power in which only a specified person or delegate is authorised to make the decision.[8] It may not be possible for decision making power to be delegated to an AI system in all cases. Organisations seeking to utilise AI to make decisions autonomously must therefore consider whether legislative decision making power can be delegated to an AI system.

Further, providing individuals with mechanisms to contest the output or outcome (i.e. a decision) made by an AI is crucial to ensuring the system is fair. However, this relies on organisations being transparent about their use of AI and the AI system's decision making processes, which can be challenging in itself.

Transparency

The key underlying principle of IPP 5 is transparency, and requires organisations to promote clear and accountable information management practices through documents such as privacy policies. Organisations should be transparent when using AI to inform decision making, or where AI is used to make decisions autonomously without human intervention. Organisations should also be transparent in how an AI system reached a decision, particularly where it does so autonomously.

While there is no obligation in administrative law to provide reasons for decisions unless specified by relevant legislation,[9] providing explanations for how an AI informed decision was reached can assist in mitigating the risks to individuals' privacy and human rights,[10] provide recourse to adversely impacted individuals, and garner social license for the decision making process.

Given the complexity of AI systems and technologies, explanations for how an AI made a decision may not be accessible to the general public. Overly technical or complex explanations are unlikely to be accessible to most people and will not be helpful in assisting individuals understand how a decision was reached. Conversely, overly simplistic explanations may not provide enough information to determine if an AI system made a decision in a manner which contravened a law.

Explanations are most beneficial when they are meaningful and accessible. Organisations may wish to endeavour to provide both kind of explanations: simple ones that can be easily understood, and complex ones that can be interpreted by technical experts.[11] In some cases, however, organisations may find it challenging to be transparent about how an AI made a decision. Where an organisation does not understand itself how an AI reached a decision, providing any kind of explanation may not be feasible at all. Nonetheless, where an explanation is possible, organisations should aim to provide one in the interest of transparency.

Access and correction

Where organisations use AI to infer personal information about individuals, they should ensure there are mechanisms in place for individuals to access that inferred information and request its correction if necessary. Several avenues for access may be relevant, depending on the circumstances. For example, individuals may need to make a request under the Freedom of Information Act 1982 (Vic), or IPP 6 may apply to give individuals the right to access or correct their personal information used or inferred by an AI.

Correction may require that information to be manually changed, as opposed to attempting to re-infer the information until it is correct.

Outsourcing and third party AI systems

In some cases, organisations may decide to buy AI solutions from a third party, or outsource the use of an AI system altogether. Where a VPS organisation enters into an outsourcing arrangement for an AI system, the outsourcing organisation is primarily responsible for ensuring those arrangements are consistent with the PDP Act.

Before outsourcing or implementing a third party AI solution, organisations should conduct a PIA and security risk assessment. These tools will help to identify the privacy and security risks arising from the outsourcing arrangement, and consider the potential trade-offs of the outsourced or third party solution. Additionally, these arrangements should be regularly reviewed, as new privacy and security risks may arise throughout the duration of the arrangement.[12] PIAs are discussed further below.

Organisations should also consider whether the outsourcing arrangement involves the transfer of personal information outside Victoria, as IPP 9 places limitations around when this can occur. Transborder data flows are discussed further below.

HANDLING

Data quality

Organisations must ensure that the personal information they collect, use and disclose is accurate, complete and up to date, in line with IPP 3. In the context of AI this applies to, for example, data used to train AI models, data fed into AI systems, and inferences that such systems make.

AI systems used for broad analytics or predictions can tolerate some degree of inaccuracy due to the large volumes of required data and outputs that do not relate to any specific individual. For AI systems used to infer information or make decisions about a specific individual, it is significantly more important that personal information is accurate.

Incorrect, incomplete, or out of date information is likely to result in inaccurate inferences. However, accurate information can also lead to inaccurate inferences. Organisations should not assume that inferred personal information is accurate and should take steps to verify it.

This is particularly important if inferred information will be used to make a decision that will affect an individual's life. For example, inferred information about a person's characteristics such as age or personal habits could result in making them ineligible for certain services, or being offered incorrect products.

Care should be taken when using historical data with AI. Historical data often contains bias reflective of the period and circumstances in which it was collected, and may lead to AI systems possessing conscious or unconscious bias.[13]

Data security and retention

The use of AI systems can carry potentially significant risks to the security of personal information, given the large amounts of data processed by, and often required to train, an AI system.

Under IPP 4.1, organisations must take reasonable steps to protect the personal information they hold from misuse, loss, and unauthorised access, modification, or disclosure. VPS organisations subject to Parts 4 and 5 of the PDP Act must also adhere to the Victorian Protective Data Security Framework 2.0 (**VPDSF**) and accompanying Victorian Protective Data Security Standards (**VPDSS**), which cover measures relating to governance and the four security domains: information security, personnel security, ICT security and physical security.[14]

What is considered 'reasonable' in a security context will depend on the circumstances and a number of different factors, such as the type and sensitivity of the information used in the AI system, the potential impact of a possible security breach of the system, and the potential harm to an individual if the security of their personal information is compromised. VPS organisations can also refer to the VPDSF and VPDSS as a guide to determining what constitutes 'reasonable steps'.

Destruction and de-identification

IPP 4.2 requires organisations to take reasonable steps to destroy or permanently de-identify personal information if it is no longer needed for any purpose. In the context of AI systems this could include, for example, personal information once used to train an AI model but is no longer needed for any purpose, or personal information that has been inferred by an AI that is similarly no longer necessary. However, keeping the data in case the model needs to be retrained may be considered an appropriate purpose.

IPP 4.2 also applies where the output of an AI system contains personal information – if that information is no longer needed for any of an organisation's purposes, it should be destroyed or permanently de-identified. Importantly, organisations need to be mindful of the limitations of de-identification. De-identification of a dataset can be difficult to perform such that it would remain permanently de-identified, particularly where the dataset is released in a public context.[15] Nonetheless, while only removing direct identifiers is generally not sufficient to de-identify a dataset, direct identifiers should be removed if deleting them would not affect the quality of the results.

Organisations should also consider other recordkeeping obligations – such as the *Public Records Act 1973* or organisational policies – that may apply in the context of AI, for example to personal information processed by AI, or decisions made by an AI system.

Transborder data flows

IPP 9 permits organisations to transfer personal information outside Victoria only under limited circumstances, including with the consent of the individual to whom the information relates (IPP 9.1(b)), and where the organisation believes the recipient of the information is subject to a law, binding scheme or contract that upholds information handling principles similar to the IPPs.

Cloud computing used to train AI models or run AI systems may require personal information to be sent overseas. Organisations transferring personal information outside Victoria as part of using AI systems must ensure that the transfer is permitted under IPP 9.

In some instances, an AI model itself may not necessarily contain personal information, even if personal information was used to train it – meaning that it could potentially be transferred outside Victoria without restrictions.

PRIVACY IMPACT ASSESSMENTS

A PIA is an important tool to assist organisations to understand and evaluate a program or project's compliance with the IPPs under the PDP Act. PIAs can help to identify potential privacy risks posed by the use of an AI system, and develop risk mitigation strategies to address these risks. Organisations seeking to adopt a new AI system should undertake a PIA in the early stages, before its implementation. PIAs should be considered a living document that is regularly updated as the scope, risks, purpose, context and nature of an AI system changes.

---

[1] For more information about the potential for discrimination, bias and inequality in AI, see OVIC's *Closer to the Machine: Technical, social and legal aspects of AI*, available at https://ovic.vic.gov.au/wp-content/uploads/2019/08/closer-to-the-machine-web.pdf.

[2] Section 5 of the PDP Act defines personal information 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 be reasonably ascertained, from the information or opinion'.

[3] Unless doing so would pose a serious threat to the life or health of any individual (IPP 1.5).

[4] See Schedule 1 of the PDP Act for the full definition of sensitive information.

[5] Section 3 of the PDP Act defines a generally available publication as 'a publication (whether in paper or electronic form) that is generally available to members of the public and includes information held on a public register'.
[6] *Jurecek v Director, Transport Safety Victoria [2016] VSC 285 (11 October 2016)* at [84], available at: https://www.austlii.edu.au/cgi-bin/viewdoc/au/cases/vic/VSC/2016/285.html.
[7] For more information about the elements of meaningful consent, see OVIC's *Guidelines to the Information Privacy Principles*, available at https://ovic.vic.gov.au/privacy/privacy-guidance-for-organisations/.
[8] In the Australian public sector, a Department of Finance guideline makes the Secretary of each department responsible for decisions made by an automated system.
[9] Examples of such a right can be found in international law, for example Recital 71 of the European Union's *General Data Protection Regulation*.
[10] For more information see the chapter 'Algorithmic transparency and decision-making accountability: thoughts for buying machine learning algorithms' of *Closer to the Machine: Technical, social, and legal aspects of AI,* available at: https://ovic.vic.gov.au/closer-to-the-machine-ai-publication; see also Australian Human Rights Commission, *Human Rights and Technology Discussion Paper* (2019).
[11] For more information on explaining AI decisions, see guidance by the UK Information Commissioner's Office and The Alan Turing Institute '*Explaining decisions made with AI*', available at https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-ai/.
[12] More information about outsourcing is available at https://ovic.vic.gov.au/privacy/privacy-guidance-for-organisations/.
[13] For more information about bias, see the chapter 'A matter of perspective: discrimination, bias and inequality in AI' of *Closer to the Machine: Technical, social, and legal aspects of AI.*
[14] For more information refer to the Victorian Protective Data Security Framework resources, available on the OVIC website at https://ovic.vic.gov.au/data-protection/framework-vpdsf/.
[15] See OVIC's *Protecting unit-record level personal information: The limitations of de-identification and the implications for the Privacy and Data Protection Act 2014*, available at https://ovic.vic.gov.au/privacy/privacy-guidance-for-organisations/. See also the Victorian Centre for Data Insight's (**VCDI**) *De-identification Guideline*, February 2018, available at https://www.vic.gov.au/sites/default/files/2019-03/Victorian-Data-Sharing-Act-2017-De-identification-guidelines.pdf.