Privacy Commissioner
Te Mana Mātāpono Matatapu

Artificial intelligence and the Information Privacy Principles

September 2023

Artificial intelligence and the Information Privacy Principles

1. Introduction: Privacy is a starting point for responsible use of AI tools

Thinking about privacy is vital if you're going to use AI tools well.

The uptake of these tools has been fast and focused on benefits, like having new ways to crunch data, do creative work, and make processes more efficient. However, there is also deep concern about the potential risks of making processes less transparent, reinforcing biases in data, and disconnecting people from important decisions.

This guidance is for all New Zealanders

If you're using, or considering using, AI tools in New Zealand then this guidance is for you. The Privacy Act 2020 applies whenever you collect, use, or share personal information.[1] As a rough guide, if you can say who information is about, it is personal information. That includes information like a name, address, contact details, or photographs of a person. It can also include technical metadata like map coordinates, Internet protocol addresses, or device identifiers related to a person.

Finally, personal information includes information about a person that is inaccurate or made up, including fake social profiles and deepfake images.

When working with personal information (including as part of using digital tools), you need to comply with the information privacy principles (IPPs).

The Privacy Act applies to business, government, and community organisations of all sizes, as well as to individuals.[2] Privacy law applies to the use of AI tools.

The Privacy Act applies to the use of AI. We want to help people understand this fast-moving area. In May 2023, the Privacy Commissioner issued and then updated a statement of expectations on privacy and generative AI.[3]

Now we're building on that work, looking at privacy and AI tools more broadly, and offering guidance on how you can apply the IPPs as part of your privacy thinking.

Our expectations for agencies using AI tools

- Have senior leadership approval based on full consideration of risks and mitigations.
- Review whether a generative AI tool is necessary and proportionate given potential privacy impacts and consider whether you could take a different approach.
- Conduct a privacy impact assessment before using these tools.
- Be transparent, telling people how, when, and why the tool is being used.
- Engage with Māori about potential risks and impacts to the taonga of their information.
- Develop procedures about accuracy and access by individuals to their information.
- Ensure human review prior to acting on AI outputs to reduce risks of inaccuracy and bias.
- Ensure that personal information is not retained or disclosed by the AI tool.

[1] With specific exceptions for news activity, the court system, and MPs other than Ministers.
[2] There is an exception for domestic use of personal information by individuals, but this does not apply where use would be highly offensive to a reasonable person.
[3] OPC, "Generative Artificial Intelligence – 15 June 2023 update" (15 June 2023) <privacy.org.nz>.

Think carefully about your use-case before relying on exciting new tools to solve it

There's a lot of excitement and urgency around new AI tools like image and text generators and best practices for these tools is still developing. However, privacy is a good starting point. To comply with privacy law, you need to be confident that you understand potential privacy risks and that you're upholding the IPPs. This is why we expect organisations to do a privacy impact assessment (PIA) before using AI tools.

A privacy impact assessment will help you use AI tools more effectively

To use AI tools effectively, you need some understanding about how they work. That includes what data sources they were trained on, and how relevant and reliable these sources are for your purposes. OPC offers guidance on writing a PIA, which is a good way to consider these questions, and to bring up and address potential risks for privacy, transparency, and trust. Doing a good PIA might also involve talking with people and communities who your work will impact.

We have tools on our website privacy.org.nz to help you develop a PIA.

Once you understand the potential risks, you can:

- Use privacy policies to govern your AI tools.
- Ensure privacy statements set clear expectations.

Considering Te Ao Māori perspectives

You need to consider Māori perspectives on privacy.[4] We recommend you're proactive in how you engage. We've heard specific concerns about Māori privacy and AI tools, including:

- Concerns about bias from systems developed overseas that do not work accurately for Māori.
- Collection of Māori information without work to build relationships of trust, leading to inaccurate representation of Māori taonga that fail to uphold tapu and tikanga.
- Exclusion from processes and decisions of building and adopting AI tools that affect Māori whānau, hapū, and iwi, including use of these tools by the Crown.

Next steps: we're keen to kōrero

We are continuing to monitor developments in the use of AI tools and their privacy impacts, and plan to update our guidance over time. If you'd like to talk with us about this work, please email us at ai@privacy.org.nz.

We'd be keen to talk about use-cases, about how people are using AI tools in New Zealand, as well as broader issues relating to AI and privacy.

2. Scope: What do we mean by AI tools?

Our guidance earlier this year focused on generative AI as the area seeing rapid change, but related privacy issues will also arise for other types of tools and uses of automation.

AI refers to computer systems doing tasks that seem like intelligent behaviour, such as finding patterns, putting items into categories, and triggering actions based on information.[5] People have been building computer systems to do some of these tasks for decades. Performance on these tasks

[4] The Commissioner must take account of cultural perspectives under s 21 of the Privacy Act 2020.
[5] J McCarthy, M Minsky, N Rochester, C Shannon (1955). "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence" <archive.org>

has now improved, with greater access to computing power, human expertise, funding, and training data. Interest in these systems and their impacts took off in the past year with the launch of consumer-facing image and text generation tools. This is still a developing area, where experts disagree on how capable current systems are and how this will develop over time.[6] Because of this uncertainty, we prefer to take a broad approach to AI systems and their potential privacy impacts.

Privacy impacts from AI can relate to:

- New ways of building computer systems (including collecting and using training data).
- New capabilities (for example, interpreting noisy data like faces in CCTV video feeds).
- New uses of computers to perform existing tasks (for example, automating decisions).

For our purposes in this paper, we are using the terms "artificial intelligence", "AI" and "AI tools" broadly to cover computer systems where one or more of the following applies:

- **machine learning systems** developed or refined by processing training data.
- **classifier systems** used to put information into categories (e.g., captioning images).
- **interpreter systems** that turn noisy input data into standardised outputs (e.g., deciding what words are present in speech or handwriting).
- **generative systems** used to create text, images, computer code, or something else.
- **automation** where computers take on tasks that people have done up until recently.

To help us consider potential privacy impacts at each stage, we're thinking in terms of a generalised set of steps summarising how AI tools are built and used, which includes:

- **gathering training data** as a resource to inform the behaviour of an AI tool
- **processing data** to create or refine a model to drive the behaviour of an AI tool
- **adopting an AI tool** in a business process, including providing it for public use
- **taking input**, including user-supplied prompts or other information
- **accepting output** and taking **decisions or actions** based on the use of an AI tool.

We are interested in how people in New Zealand are using and are impacted by AI tools. Let us know if you have a view on what features of AI tools give rise to specific privacy issues.

3. Guidance: How can you apply the IPPS to AI tools?

At the core of the Privacy Act are the 13 information privacy principles (IPPs) that set out how agencies must handle personal information. The IPPs govern the activities of collecting, using, and sharing personal information. They apply whether you're building your own AI tools, using AI tools to support decision making, or have team members who are informally using AI in their work. They also apply where overseas organisations provide AI tools for use in New Zealand. In each case, you need to think about how you're upholding your privacy obligations.

Why do AI tools create special issues for privacy?

Uptake of AI tools present some specific challenges for privacy. Privacy protections rely on people and organisations who can understand context and take responsibility for their actions. This may

[6] See e.g., M Mitchell, "Can Large Language Models Reason?" (Sep 2023), <aiguide.substack.com>

become harder as AI tools take on more tasks, because they enable new ways to gather and combine personal information, and because they can make it harder to see, understand, and explain how personal information is used. Even experts are often unable to explain how an AI tool arrived at a particular output or decision. This means it is important to think about where and how to use these tools before giving them our personal information.

How do AI tools relate to collecting, using, and sharing personal information?

The IPPs apply to collecting, using, and sharing personal information. In the context of AI tools, you should think about situations such as gathering training data, processing login details to provide access to a tool, accepting inputs such as user prompts, generating model outputs, and actions based on AI tools. Each potentially involves collecting, using, or sharing personal information.

Computer systems rely on good information to get good results. AI tools are built from training data that informs their pattern-matching and generative capabilities. To make good decisions about how to use these tools effectively, it's critical to understand what is in the training data, how relevant and reliable it is for your intended purpose, and whether it is gathered and processed in ways that comply with legal obligations and uphold your ethical values.[7]

The best time to start privacy work is at the beginning

By taking proactive steps early on you can check you're upholding your Privacy Act obligations. The best time to do privacy work is as soon as possible, especially for AI tools. At the start, you can take a privacy-by-design approach.[8] You can talk with people across your organisation and across relevant communities about their concerns and how to uphold good privacy practices. This is why our initial guidance said you should only use a generative AI tool after conducting a PIA and seeking feedback from impacted people and communities, including Māori, to understand and mitigate potential privacy risks. Skipping this privacy analysis at an early stage may create problems later.

You may want to build privacy thinking into work by:

- Having strategic conversations around your organisation's priorities, values, and approach to AI.
- Developing project plans to build or adopt AI tools into your organisation.
- Developing and updating key documents such as privacy policies, contracts with vendors and customers, and your processes for technology projects and information governance.

For smaller organisations, a great first step might be to tell your customers and community that you're looking at using AI tools, and to see if they're comfortable with that. If you're not sure, the safest approach is to avoid putting personal information into AI tools, and to make sure that is clear to everyone across your organisation. It's critical to understand how people in your organisation may already be using AI. These people will have useful context on how to think about use of AI tools but may be unknowingly putting people's privacy at risk.

… it does need to be meaningful

Complying with the Privacy Act may require meaningful steps to change how AI tools are used, and how associated personal information is governed. For example, while having a human in the loop can help, on its own this may not be enough to uphold good privacy practices. It is well-established that people overseeing computer systems can find it hard to notice errors and mistakes, suffering

[7] While our focus is on privacy, you may also want to consider other issues such as copyright, environmental impacts, use of click-workers and employment impacts.

[8] DIA, "Privacy by Design" <digital.govt.nz>.

"automation blindness".[9] If you think AI tools need human oversight, you may need to think about how to make this oversight effective.

...but it doesn't need to be difficult

While proactive work may be needed to comply with the Privacy Act, asking a few key questions and making some basic decisions about information governance can go a long way. Below we set out some key questions that agencies can consider to start this work.

Collection: Why are you collecting information and how should you do so?

Collection is where responsibilities under the Privacy Act start. IPPs 1-4 cover why and how personal information can be collected. In general, personal information can only be collected where it is necessary for a lawful purpose (IPP1). This means thinking carefully about why you need information, and not collecting more than is needed for that purpose.

The purpose of collection also limits the retention, use, and disclosure of personal information under other IPPs. Before you start collecting personal information you need to be clear what purpose you need it for, and why you need it for that purpose. There are some specific exceptions to the general rules for collection, but the nature of AI tools may make these difficult to rely on.

Any time you seek out or obtain information, you are collecting it. In relation to AI tools this could include asking a person directly for their information to feed into an AI system, using an AI tool to write a letter to a person or about a person based on prompts you give it, or asking an AI system a question that generates information about that person (including false information). If you have already collected the information in the past and are thinking of using it to feed into AI, you need to think about the purpose for which you originally collected the information, and whether feeding the information into AI is directly related to that purpose. Even if the AI model gives you information that you know to be far-fetched or inaccurate, that is personal information you have collected (we cover related IPP8 accuracy issues below).

Are you collecting information from the right person?

In general, agencies must get personal information directly from the person it is about (IPP2) and must be transparent about the information being collected and how it will be used (IPP3). Meeting these requirements may be difficult without a good understanding of the training data and processes used to develop an AI tool, particularly tools based on varied data sources.

> What's in the training data?
>
> AI tools work by reproducing patterns found in their training data. Popular image and text generation tools are based on data taken from a wide range of Internet sources, making it difficult to know whether personal information is reliable and collected in a fair way.[10]

Can you rely on an exception to the normal rules for collection?

There are some exceptions to the normal collection requirements. For example, there is an exception for "publicly available" information. However, it may be risky to rely on these exceptions as a path to upholding privacy obligations without a good understanding of the training process for an AI tool. Training data scraped from the Internet may include sources which require a login to access, such as social media profiles. This means you may need to be cautious about assuming this material is "publicly available" in a way you can rely on and might be outside the expectations people have on how this information would be used.

[9] See C Janssen, S Donker et al, "History and future use of human-automation interaction", International Journal of Human-Computer Studies 131 (2019) 99-107, <https://par.nsf.gov/servlets/purl/10186109> at 103.

[10] K Schaul, *Washington Post*, "Inside the secret list of websites that make AI like ChatGPT sound smart" (19 April 2023), <washingtonpost.com>.

Are you sure that personal information was or will be collected in a way that is fair?

Agencies must ensure that the manner of collection of personal information is lawful, fair, and does not unreasonably intrude on personal affairs, particularly when collecting information from children or young people (IPP4). This may require special attention to situations where it may not be obvious to people that their information may be collected at all, or where the purpose for collecting or using information may not be obvious. What is fair in the circumstances may depend on the nature of your organisation and services, for example people cannot readily avoid using banking services, supermarkets, or telecommunications. This may mean more care is needed to ensure a fair manner of collection. Fairness may also require extra steps when collecting information from children and young people.[11]

AI tools are built on training data. If you're providing a service such as a chatbot or phone line, and want to collect information to train, refine, or use AI tools, you may need to take particular care to clearly explain this to people, and to consider offering people a chance to opt out from this use of information. It is a good idea to be transparent with people, set clear expectations, and avoid surprises. You may need to look at how third-party services you use are collecting data for training purposes and ensure this is clear to people too.

Key questions for collection

- Can you ensure that information about a person is collected from that person? If not, do any of the exceptions apply?
- What risks might be involved in using AI tools without understanding the training data? What steps are you taking to minimise risk? (e.g., no input of personal information)
- Can you choose to use AI tools based on responsibly collected training data?
- Are you clearly explaining to people the ways that you will use and disclose their information, including potential uses of AI tools?
- Are you engaging with communities who might have concerns or perspectives about the potential use of AI tools, including Māori?
- Is there a risk you may be using personal information obtained in a data breach? Can you do due diligence to make sure AI tools are not drawing on illegally obtained data?

[11] IPP4(b)

Use-case: Writing content and using chatbots

You might be interested in using AI tools to provide information to customers through a generative AI-based chatbot, or using an AI tool to write emails or letters. Try to avoid putting personal information into the AI tool. For example, a landlord might decide to use an online text generator to write letters and emails to tenants. However, under the Privacy Act they would be considered an agency and so the law applies to them; they should avoid putting personal information into AI tools.

To build their understanding and ensure they are upholding their privacy obligations, the landlord might want to ask questions like:

- What are the terms and conditions of this AI tool? Will the provider store information I put?
- Will it be used by the provider to train their AI systems?
- Have I talked with my tenant? Are they aware that I might put information about them into an AI tool? If so, are they comfortable with that?
- Can you use the tool in a way that minimises personal information in the input? If you are looking at an AI base chatbot the same questions apply.

If you are looking at an AI base chatbot the same questions apply. You may also want to ask:

- How will I know the chatbot is giving accurate information to people? What can I do to test the model is reliable and remains reliable over time?
- Do people have another way to get that information? Not everyone likes chatbots, so providing a website, an online form, and a phone number may still be important.

Security and retention

Agencies need to protect personal information against loss, unauthorised access, and other misuse (IPP5). This includes taking reasonable cybersecurity steps to protect information, for example adopting two-factor authentication to protect against risks from phishing.[12]

AI tools enable new ways to access and use information, and this creates new security risks. Use of AI tools within organisations may lead to more sharing of information with third party providers, as team members supply images, text prompts, and other data. Some AI tools can leak sensitive information that increases security risks. For example, Large Language Models-based programming tools have been known to suggest real passwords and API keys found in training data.[13]

AI tools can also make it easier for criminals to impersonate real people, create fake identities online, and automate hacking and phishing campaigns.[14]

These uses of AI tools add to existing privacy harms, making it easier to exploit illegally obtained information from data breaches, to re-identify people in published datasets, and to combine any sources of data to build detailed profiles of people which are more or less accurate.

[12] OPC, "Office of the Privacy Commissioner encourages two-factor authentication in war on cybercrime", (7 Jun 2023) <privacy.org.nz>
[13] See L Niu, S Mirza et al, "CodexLeaks: Privacy Leaks from Code Generation Language Models in GitHub Copilot" (August 2023), <usenix.org>
[14] See for example the OWASP top 10 for LLM <owasp.org>

You should proactively consider how to manage security risks to personal information, including risks related to your use of AI tools, and the use of AI tools by other people. You may need privacy breach response plans to ensure you can notice, contain, assess, and respond to privacy breaches quickly, to mitigate potential harm to individuals and comply with the Act.

- How might AI tools affect the risk of privacy breaches?
- Are you managing security risks?

Access and correction

When an agency holds information about them, a person can ask for that information (IPP6). The agency must generally give someone access to their information, unless there is a reason to withhold it.[15] A person can also ask you to correct information about them (IPP7).

Some uses of AI tools may make it hard to comply with the access and correction principles. Building AI tools involves processing training data to build models that do pattern-matching. The original training data, the pre-trained model, and the outputs may all potentially contain personal information while providing no practical way to access or correct it, and still posing a risk of disclosure or influencing outcomes for a person in response to particular prompts.

It is essential that you develop procedures for how your agency will respond to requests from individuals to access and correct their personal information. You also need to check you are providing information to the correct individual. AI tools may make it easier to realistically impersonate someone and someone and request access to their information.

Key questions for access and correction. At the procurement stage, before you put an AI tool into use, you might consider the following:

- Are you confident you can provide information about a person to them if they ask for it?
- Are you confident that you can correct personal information?
- How often are models you rely on updated? Can you correct AI outputs in a timely way?
- How will you verify the identity of an individual requesting their information?

Accuracy

Agencies that hold an individual's personal information must not use or disclose that information without taking reasonable steps to ensure that the information is accurate, up to date, complete, relevant, and not misleading (IPP8).

The accuracy principle protects people against harm and upholds the relevance and reliability of information about people. It is relevant at all stages from gathering training data, to training a model, to implementing a tool, accepting input, providing output, and taking actions as a result. To uphold this principle, it is important to consider what assurances you have that an AI tool will be accurate, relevant, and reliable at each stage, so you may need to investigate the training process behind an AI tool.

---

[15] Reasons to withhold information include that it does not exist or cannot be found, or that disclosure would have a negative impact on other interests such as another person's privacy, maintenance of the law, and defence. See Part 4 of the Act.

Generative AI tools produce confident errors of fact and logic, so you cannot rely on their apparent performance or self-declared capabilities as evidence of accuracy. You should also take a critical approach to accuracy claims by providers of AI tools. Most of the public-facing AI tools now available have been developed overseas and are based on training data that may not be relevant, reliable, and ethical for use in Aotearoa New Zealand.

While we are also part of the broader world, we have our own unique mix of cultural perspectives, demographics, and use of languages including English and Te Reo Māori. Adopting AI tools that are not designed for and in consultation with our communities may lead to inaccuracy in the form of bias, exclusion (particularly of poorly represented groups), and other privacy harms.

Before using AI tools, it's essential that agencies take reasonable steps to assure themselves that these tools will uphold the accuracy principle. Depending on the nature of the intended use and level of risk, this may require independent testing and auditing. Doing and updating a PIA will be an essential part of this process.

Use-case: Using AI tools to screen documents

You may be interested in using AI tools to make decisions based on text or documents, for example screening job applications to find people you want to interview, screening offensive comments online, or detecting text that has been generated by AI. The track record of AI tools in this area is not good, so you need to be very confident the system you want to use will be transparent, accurate, and fair before you ask anyone to rely on it.

You might want to ask:

- How can I find out about the reliability and accuracy of the AI tool for this use-case?
- Is there a risk of bias in the AI tool or the training data?
- Who can I talk with to ensure people are ok with me using this tool? Can I engage with experts? Can I engage with people and communities who might be affected?

Accuracy and automated decision making

One critical area for accuracy is the use of AI tools for automated decision making, where they may have direct impacts on outcomes for people. The higher risk of these situations means it will generally be necessary to develop processes for human review of decisions, and to empower and adequately resource the people doing this work. Simply having a "human in the loop" may not be enough to uphold the accuracy principle, given the well-known problem of automation blindness in people overseeing automated systems.[17] Overall, it may help to think in terms of using AI in ways that uphold accountability for people in your organisation, to your customers, and to the broader community. Where people have responsible roles, whether as financial managers, teachers, or public servants, any use of AI tools should maintain and complement the responsibilities people have in their roles. Talking with the people in these roles and the people they work will often be critical for good use of AI.

Key questions for accuracy

At the procurement stage, before you put an AI tool into use, you might consider the following.

- What steps can you take to ensure an AI tool is accurate and fair for your intended purpose?
- Can you investigate the training data? Is it reliable and relevant?

16 "OpenAI discontinues its AI writing detector due to "low rate of accuracy"", *ArsTechnica* (7 July 2023), <arstechnica.com>
17 C Doctorow, "Pluralistic: Supervised AI isn't", (23 August 2023), < https://pluralistic.net/2023/08/23/automation-blindness/>

- Who can help you think about issues of accuracy, fairness, and bias? Can you engage with:
    - Experts who can offer an independent perspective?
    - People and communities who will bear risks from biased or inaccurate information?
- What processes can you put in place to ensure an AI is fit-for purpose? Can you:
    - Run a trial to test how reliable and accurate it is?
    - Run regular audits of accuracy?
- What processes will ensure that information is accurate:
    - For any personal information you hold?
    - For any personal information your AI provider holds?
    - And ensure you are deleting or correcting inaccurate data?
- How often are models you rely on updated? Can you correct AI outputs in a timely way?
- How will you verify the identity of an individual requesting their information?

Use and disclosure

The ways you can use and disclose personal information depends on the purpose of collection. Agencies need to clearly identify the purpose or purposes for collecting personal information, and to limit use and disclosure to those purposes or a directly related purpose (IPPs 10 and 11). In general, you cannot repurpose existing information for a new purpose unless it is a directly related purpose, or when authorised by law, or when a specific exception applies.

The development and use of AI tools opens new ways to use and disclose information. Some of these may be directly related to the purpose for which it was collected, but others may not be. For example, if you want to use personal information to train an AI tool, you need to make sure that is clear at the time you collect it. Using an AI tool may require you to log in and to supply a prompt. Sharing existing information about a person in this way may go beyond the purpose for which you collected that information. In general, providers of AI tools may not ring-fence information you provide from other information they hold. If you share personal information that is then used for future training purposes, it may later be leaked or re-identified. This may be a reason to prefer AI tools that you can run offline, without sharing back data, or AI tools where the contractual terms stop the provider from using input data for training.

Key questions for use and disclosure

- At the procurement stage, before you put an AI tool into use, you might consider the following.
- What was the purpose for which information was collected? How are you tracking this?
- If you share personal information with third party suppliers, might they be using it for training AI tools?
- Do your supplier contracts and customer communications set clear expectations with the people whose information is involved about how it will be used and kept secure?

Overseas disclosure

IPP12 states that personal information must not be disclosed overseas unless certain conditions are met. Using offshore technology providers to store or process your data is not treated as a disclosure under IPP12, so long as they are not using that information for their own purposes.
Whether or not IPP12 applies in your situation may depend on the details of AI tool being used. You should check whether information you share to technology providers is used for purposes beyond those you collected it for. This may be worth looking at when you choose providers, and it is important you tell people clearly if this may affect their information.

It is important to remember that if you decide to use a third-party system to handle personal information (this could be an AI tool), you remain responsible for protecting that information. This responsibility should not stop you from using third-party services. You simply need to assess the risks and make sure you deal with them appropriately. A PIA will be important for assessing this risk.

Unique identifiers

The Privacy Act imposes restrictions on how agencies can assign and use a 'unique identifier' (IPP13).

A unique identifier is an identifier, other than the individual's name, that uniquely identifies an individual (for example, a Tax File Number or customer number). In general, agencies are required to assign a unique identifier only where a person's identity is clearly established, to avoid using the same unique identifiers as another agency, to minimise the risk of misuse (for example by truncating account numbers on correspondence), and to only require an individual to disclose a unique identifier in connection with the purpose for which it was assigned.

AI systems are powerful pattern matching tools that learn from very large data sets. This may lead to AI tools finding patterns in a person's behaviour that qualify as a unique identifier, even if this is not an intended outcome of using these tools. IPP13 limits the use the same identifier by another agency, to avoid the creation of profiles on people.

Key questions for use and disclosure:

At the procurement stage, before you put an AI tool into use, you might consider the following.

- Might the use of an AI tool lead to assigning unique identifiers to individuals? If so, how are you managing these unique identifiers?
- Is the AI assigned unique identifier the primary way your agency will identity the individual?
- If it is the primary identifier, do you know if another agency is using the same identifier?
- Does training data include unique identifiers that can be revealed in outputs?