Australian Government
Department of Finance

© Department of Finance 06/05/2025
This content is only accurate as at the date of printing or download. Refer to Home | Department of Finance to ensure you are viewing the latest version.

Implementing Australia’s AI Ethics Principles in government

The following practices are mapped to *Australia’s 8 AI Ethics Principle*s, demonstrating how governments can practically apply them to their assurance of AI.

Their application may differ according to jurisdictional specific governance and assurance protocols. Similarly, different use cases present different risks with some requiring a higher standard of assurance than others. Therefore, not all AI use cases will require the detailed application of all available practices to be considered safe and responsible.

These practices were developed by drawing extensively from the existing practices of the Australian, state and territory governments, as well as these publications:

- *NSW Artificial Intelligence Assurance Framework* (Digital NSW 2022)
- *Adoption of Artificial Intelligence in the Public Sector* (DTA 2023)
- *Safe and responsible AI in Australia consultation: Australian Government’s interim response* (DISR 2024)
- *Implementing Australia’s AI Ethics Principles* (Gradient Institute and CSIRO 2023)
- *Responsible AI Pattern Catalogue* (CSIRO 2023)
- *How might artificial intelligence affect the trustworthiness of public service delivery?*(PM&C 2023)

1. Human, societal and environmental wellbeing

Throughout their lifecycle, AI systems should benefit individuals, society and the environment.

1.1 Document intentions

Governments should define and document the purpose and objectives of a use case and the outcomes expected for people, society and the environment.

Document risks, consider whether the use of AI is preferable, whether there is a clear public benefit and what non-AI alternatives are available. Existing frameworks or policies for benefits realisation may assist.

1.2 Consult with stakeholders

Governments should identify and consult with stakeholders, including subject matter and legal experts, and impacted groups and their representatives.

Seek input from stakeholders early to allow for the early identification and mitigation of risks.

1.3 Assess impact

Governments should assess the likely impacts of an AI use case on people, communities, societal and environmental wellbeing to determine if benefits outweigh risks and manage said impacts appropriately.

Methods such as algorithmic and stakeholder impact assessments may assist.

2. Human-centred values

AI systems should respect human rights, diversity and the autonomy of individuals.

2.1 Comply with rights protections

Governments will ensure their use of AI complies with legal protections for human rights. This may include those protected under:

- legislation at all levels of government
- Australia’s international human rights obligations
- the Australian and state constitutions
- interpretation of common law.

Any use will also align with related obligations, policies and guidelines for the public sector, workplace health and safety, human rights, and diversity and inclusion.

Human rights impact assessments may assist to identify, assess and mitigate human rights risks. Where necessary seek advice from subject matter experts.

2.2 Incorporate diverse perspectives

Governments should involve people with different lived experiences, including marginalisation, throughout the lifecycles of a use case to gather informed perspectives, remove preconceptions and avoid overlooking important considerations.

This may include representation of:

- people living with disability
- multi-cultural communities
- religious communities
- people from different socio-economic backgrounds
- diverse genders and sexualities
- Aboriginal and Torres Strait Islander people.

2.3 Ensure digital inclusion

Governments should align to digital service and inclusion standards, and account for the needs, context and experience of individual users across an AI use case’s lifecycle.

Consider assistive technologies to support people who live with disability.

In focus: The CSIRO’s *Guidelines for Diversity and Inclusion in Artificial Intelligence*

The CSIRO’s *Guidelines for Diversity and Inclusion in Artificial Intelligence* (Zowghi D and da Rimini F 2023) address the evolving and holistic nature of AI technologies, the importance of diversity and inclusion consideration in the development and deployment of AI, and the potential consequences of neglecting it.

The guidelines emphasise the importance of a socio-technical perspective on diversity and inclusion in AI, highlighting the necessity of involving relevant stakeholders with diverse attributes, examining cultural dynamics and norms, and evaluating societal impacts.

Explore the guidelines on the CSIRO website.

3. Fairness

AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups.

3.1 Define fairness in context

Governments should consider the expected benefits and potential impacts of using AI, as well as vulnerabilities of impacted groups, to determine ‘fairness’ in a use case’s context.

3.2 Comply with anti-discrimination obligations

Governments will ensure their use of AI complies with relevant anti-discrimination legislation, policies and guidelines for protected attributes. These may include:

- age
- disability
- race
- religion
- sex
- intersex status
- gender identity
- sexual orientation.

Well trained and supported staff should be able to identify, report and resolve biased AI outputs. Where necessary, seek advice from subject matter experts.

3.3 Ensure quality of data and design

Governments should ensure high-quality data and algorithmic design.

Audits of AI inputs and outputs for unfair biases, data quality statements and other data governance and management practices may assist to understand and mitigate bias in AI systems.

In focus: the Australian Human Rights Commission's *Using artificial intelligence to make decisions: Addressing the problem of algorithmic bias* • Technical Paper

This technical paper is a collaborative partnership between the Australian Human Rights Commission, Gradient Institute, Consumer Policy Research Centre, CHOICE and CSIRO's Data61.

It explores how the problem of algorithmic bias can arise in decision making that uses artificial intelligence and how this problem can produce unfair, and potentially unlawful, decisions as it may lead to a person being unfairly treated or even suffering unlawful discrimination based on characteristics such as race, age, sex or disability. It demonstrates how the risk of algorithmic bias can be identified and steps that can be taken to address or mitigate this problem.

This paper forms part of a AHRC's Human Rights and Technology Project. You can read the technical paper on the AHRC website.

4. Privacy protection and security

AI systems should respect and uphold privacy rights of individuals and ensure the protection of data.

4.1 Comply with privacy obligations

Governments will ensure their use of AI complies with legislation, policy and guidelines that govern consent, collection, storage, use, disclosure and retention of personal information.

This may include informing people when their personal information is being collected for an AI system or when personal information is used for a secondary purpose such as AI system training.

'Privacy by design' principles and privacy impact assessments may assist to identify, assess and mitigate privacy risks. Where necessary, seek advice from subject matter experts.

4.2 Minimise and protect personal information

Governments should assess whether the collection, use and disclosure of personal information is necessary, reasonable and proportionate for each AI use case.

Consider if similar outcomes can be achieved with privacy enhancing technologies.

Synthetic data, data anonymisation and deidentification, encryption, secure aggregation and other measures may assist to reduce privacy risks.

Sensitive information should always be managed with caution.

4.3 Secure systems and data

Governments should ensure each use case complies with security and data protection legislation, policies and guidelines, including through an AI system's supply chains.

Security considerations should be consistent with the cyber security strategies and polices of impacted jurisdictions.

Access to systems, applications and data repositories should be limited to authorised staff as required by their duties. Where necessary, seek advice from subject matter experts.

Governments should consider relevant security guidance and strategies including:

- 2023-2030 *Australian Cyber Security Strategy*(Home Affairs 2023)
- *Hosting Certification Framework* (Home Affairs n.d.)
- *Engaging with Artificial Intelligence* (ASD 2024)
- *Deploying AI Systems Securely*(ASD 2024)
- *Countering the Insider Threat: A guide for Australian Government* (Attorney- General's Department 2023)

In focus: Office of the Victorian Information Commissioner's *Artificial Intelligence – Understanding Privacy Obligations*

Published in April 2021, the Office of the Victorian Information Commissioner's *Artificial Intelligence – Understanding Privacy Obligations* (OVIC 2021) provides guidance to assist Victorian Public Service organisations consider their privacy obligations when using or considering the use of personal information in AI systems or applications.

It covers the collection, use, handling and governance of personal information within this context.

Organisations should conduct a privacy impact assessment 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.

5. Reliability and safety

Throughout their lifecycle, AI systems should reliably operate in accordance with their intended purpose.

5.1 Use appropriate datasets

Governments should ensure that, wherever practical, AI systems are trained and validated on accurate, representative, authenticated and reliable datasets that are suitable for the specific use case.

5.2 Conduct pilot studies

Governments should evaluate AI systems in small-scale pilot environments to identify and mitigate problems and iterate and scale the solution.

Consider the trade-offs between governance and effectiveness: a highly controlled environment may not accurately reflect the full risk and opportunity landscape, while a less controlled environment may pose governance challenges.

5.3 Test and verify

Governments should test and verify the performance of AI systems. Red teaming, conformity assessments, reinforcement from human feedback, metrics and performance testing, and other methods may assist.

5.4 Monitor and evaluate

Governments should ensure their use of AI is continuously monitored and evaluated to ensure its operation is safe, reliable and aligned to ethics principles.

This should encompass an AI system's performance, its use by people, and impacts on people, society and the environment, including feedback from those impacted by AI-influenced outcomes.

5.5 Be prepared to disengage

Governments should be prepared to quickly and safely disengage an AI system when an unresolvable problem is identified.

This could include a data breach, unauthorised access or system compromise. Consider such scenarios in business continuity, data breach and security response plans.

6. Transparency and explainability

There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them.

6.1 Disclose the use of AI

Governments should ensure their use of AI is disclosed to users or people who may be impacted by it. Governments should maintain a register of when it uses AI, its purpose, intended uses, and limitations.

6.2 Maintain reliable data and information assets

Governments should comply with legislation, policies and standards for maintaining reliable records of decisions, testing, and the information and data assets used in an AI system. This will enable internal and external scrutiny, continuity of knowledge and accountability.

6.3 Provide clear explanations

Governments should provide clear, simple explanations for how an AI system reaches an outcome. This includes:

- inputs and variables and how these have influenced the reliability of the system
- the results of testing including technical and human validation
- the implementation of human oversight.

When explainability is limited, governments should weigh the benefits of AI use against explainability limitations. Where a decision is made to proceed with AI use, document reasons and apply heightened levels of oversight and control.

When an AI system influences or is used as part of administrative decision making, decisions should be explainable, and humans accountable.

6.4 Support and enable frontline staff

Governments should ensure staff at frontline agencies are well-trained and supported to clearly explain AI-influenced outcomes to users and people.

Consider the importance of human-to-human relationships for a range of people, including vulnerable people or groups, people facing complex needs and those uncomfortable with government's use of AI.

In focus: Public Record Office Victoria's *AI Technologies and Recordkeeping Policy*

Released in March 2024, Victoria's *Artificial Intelligence (AI) Technologies and Recordkeeping Policy* (PROV 2024) was designed to address transparency and accountability concerns in relation to AI implementation and use and to enable explainable AI use.

This includes the production of full and accurate records/data, as well as the appropriate management of those records/data in accordance with the PROV Recordkeeping Standards framework.

7. Contestability

When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system.

7.1 Understand legal obligations

Governments will ensure their use of AI in administrative decision-making complies with law, policy and guidelines that regulate such processes.

This includes principles of legality, fairness, rationality and transparency, and access to reviews, dispute resolutions and investigations.

Where necessary, governments should seek legal advice as to their legal obligations and proposed use of AI.

7.2 Communicate rights and protections clearly

Governments should clearly communicate the rights and protections of those impacted by each AI use case and create an avenue to voice concerns and objections and seek recourse and redress.

This includes clearly communicating the channels and processes to challenge the use or outcomes of an AI system.

Feedback and response mechanisms should be clear and transparent, ensure timely human review and exist across the use case's lifecycles.

In focus: the Commonwealth Ombudsman's *Automated Decision-making Better Practice Guide*

Released in March 2020, the *Automated Decision-making Better Practice Guide* [PDF 571KB] (Commonwealth Ombudsman 2020) recognises the significant role automation plays in administrative decision-making. The key message of the guide is that people must be at the centre of service delivery.

It provides specific guidance on administrative law, privacy, governance and design, transparency and accountability, and monitoring and evaluation of automated decision-making systems including those that contain AI.

It also provides practical tools for agencies, including a checklist designed to assist managers and project officers during the design and implementation of new automated systems, and ongoing assurance processes for once a system is operational.

Similarly, the NSW Ombudsman has released guidance on automated decision making in the public sector.

8. Accountability

Those responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.

8.1 Establish clear roles and responsibilities

Governments should ensure their use of AI is overseen by clearly identified roles and lines of accountability. Governments should consider:

the role of senior leadership and area-specific responsibilities

security, data governance, privacy and other obligations

integration with existing governance and risk management frameworks.

8.2 Train staff and embed capability

Governments should establish policies, procedures, and training to ensure all staff understand their duties and responsibilities, understand system limitations and implement AI assurance practices.

8.3 Embed a positive risk culture

Governments should ensure a positive risk culture, promoting open, proactive AI risk management as an intrinsic part of everyday practice.

This fosters open discussion of uncertainties and opportunities, encourages staff to express their concerns and maintains processes to escalate to the appropriate accountable parties.

8.4 Avoid overreliance

Governments remain responsible for all outputs generated by AI systems and must ensure incorrect outputs are flagged and addressed.

Governments should therefore consider the level of reliance on their use of AI and its potential risk and accountability challenges. Overreliance can lead to the acceptance of incorrect or biased outputs, and risks to business continuity.

Home →

Statement from Data and Digital Ministers →

Introduction →

Cornerstones of AI assurance →

Resources →

Contact

✉ Data and Digital Ministers Meeting

Content Focus

Audience

Government

Topic

Public Data

Updated: 21 June 2024