National framework for the assurance of artificial intelligence in government

A joint approach to safe and responsible AI by the Australian, state and territory governments.

21 June 2024

Australian Government

NSW GOVERNMENT

VICTORIA State Government

Queensland Government

GOVERNMENT OF WESTERN AUSTRALIA

Government of South Australia

Tasmanian Government

ACT Government

NORTHERN TERRITORY GOVERNMENT

The Australian, state and territory governments acknowledge the Traditional Custodians of Country throughout Australia and recognise the continuing connection to lands, waters and communities.

We pay our respects to Aboriginal and Torres Strait Islander cultures and to Elders past and present.

Version 1.0, published 21 June 2024

To reference this document, use the author-date system as demonstrated in this example:

> Australian Government et al. (2024) *National framework for the assurance of artificial intelligence in government*, Australian Government, accessed DD Month YYYY.

Contents

Statement from Data and Digital Ministers

Artificial intelligence (AI), while not new, is a transformative technology undergoing accelerated development and adoption.

It presents great opportunities for all levels of government to transform public service delivery and enhance societal, economic and environmental wellbeing.

However, we know there are risks with governments' use of AI that require careful oversight, including legal, privacy, security and ethical risks such as bias and fairness. The importance of managing these risks has been outlined in the Australian Government's interim response to the safe and responsible use of AI consultation.

We recognise that public confidence and trust is essential to governments embracing the opportunities and realising the full potential of AI. To gain public confidence and trust, we commit to being exemplars in the safe and responsible use of AI. This requires a lawful, ethical approach that places the rights, wellbeing and interests of people first.

This national framework for the assurance of AI in government is a key step towards gaining public confidence and trust in the safe and responsible use of AI by Australia's governments.

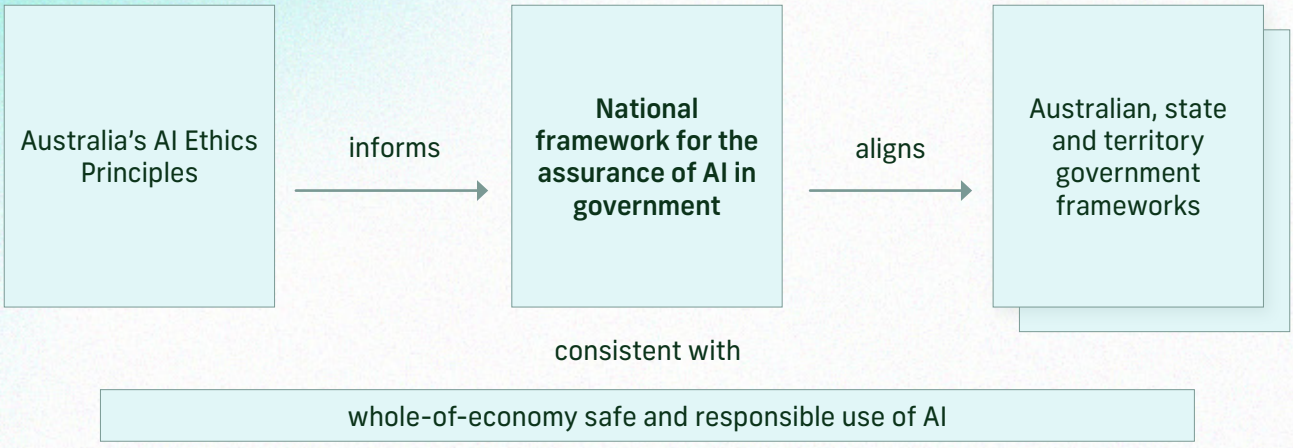
Based on *Australia's AI Ethics Principles*, it sets foundations for a nationally consistent approach to AI assurance, providing clear expectations, as well as consistency and certainty for our partners. It will assist governments to develop, procure and deploy AI in a safe and responsible way.

We recognise the scale and nature of AI developments can be uncertain. However, by embedding a principles-based approach in this national framework, we commit to flexibility, responsiveness, continuing collaboration and improvement of our AI assurance processes.

Our commitment to putting the rights, wellbeing and interests of people first remains steadfast and unchanged.

Introduction

The national framework for the assurance of AI in government provides for a nationally consistent approach for the assurance of artificial intelligence use in government.



Based on *Australia's AI Ethics Principles* (DISR 2019), and consistent with broader work on safe and responsible AI, the framework establishes cornerstones and practices of AI assurance. Instead of focusing on technical detail, the framework sets foundations across all aspects of government, with jurisdictions to develop specific policies and guidance considerate of their own legislative, policy and operational context.

Assurance is an essential part of the broader governance of how governments use AI, including its development, procurement and deployment. This enables governments to:

- understand the expected benefits of AI
- identify risks and apply mitigations
- ensure lawful use
- understand if AI is operating as expected
- demonstrate, through evidence, that the use of AI is safe and responsible.

As the Australian, state and territory governments continue to use AI, they will likely develop new or improved assurance practices based on their unique successes, vulnerabilities and impacts. These learnings will be shared and incorporated into future iterations of this framework.

Complementary initiatives

The national framework for the assurance of AI in government complements local and global initiatives on the safe and responsible use of AI, both by governments and in wider economies. The Australian, state and territory governments will consider these and other initiatives as they develop their unique assurance approaches.

Australia's AI Ethics Framework

First published in 2019 and developed by the CSIRO's Data61 and the Department of Industry, Science and Resources (DISR). *Australia' AI Ethics Framework* (DISR 2019) defines the ethics principles which inform the practices found in this national assurance framework.

Explore the ethics framework on the DISR website.

Safe and responsible AI in Australia

Following consultation initiated DISR in 2023, the Australian Government committed to ensuring the use of AI systems in high-risk settings is safe and reliable while use in low-risk settings can continue largely unimpeded.

As set out in the government's interim response to the consultation, this work will ensure AI is used safely and responsibly across the wider economy. A crucial element of this agenda is the role of government as an exemplar in the safe and responsible use of AI.

Read the Australian Government's interim response on the DISR website.

NSW Artificial Intelligence Assurance Framework

When published in 2022, the NSW Government became the world's first to mandate an assurance framework for the use of AI systems.

The *NSW Artificial Intelligence Assurance Framework* (Digital NSW 2022) assists project teams using AI to comprehensively analyse and document their projects' AI specific risks. It also assists teams to implement risk mitigation strategies and establish clear governance and accountability measures.

Access the *NSW Artificial Intelligence Assurance Framework* on the digital.nsw website.

OECD principles for responsible stewardship of trustworthy AI

Committed to by the Australian Government when first published by the Organisation for Economic Cooperation and Development in 2019. These principles aim to foster innovation and trust in AI by promoting the responsible stewardship of trustworthy AI while ensuring respect for human rights and democratic values.

Read the recommendation containing the principles on the OECD website.

Bletchley Declaration on AI safety

Agreed to by Australia, alongside another 27 countries and the European Union, at the UK Government's 2023 AI Safety Summit. The declaration affirms that AI should be designed, developed, deployed, and used in a manner that is safe, human-centric, trustworthy and responsible.

Read the text of the Bletchley Declaration as hosted on the DISR website.

Seoul Declaration for safe, innovative and inclusive AI

On 21 May 2024, the Australian Government agreed to 3 outcomes at the AI Seoul Summit, South Korea:

- Declaration for safe, innovative and inclusive AI
- Statement of Intent toward International Cooperation on AI Safety Science
- Ministerial Statement for advancing AI safety, innovation and inclusivity.

The agreements build on the Bletchley Declaration, confirming a shared understanding of opportunities and risks, and committing nations to deeper international cooperation and dialogue.

Read the text of the Seoul Declaration as hosted on the DISR website.

What is an AI system?

In November 2023, OECD member countries approved this revised definition of an AI system:

'A machine-based system 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. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.'

To avoid definitional complexities the Australian, state and territory governments should consider practical guidance for staff to identify when AI assurance processes apply, such as when:

- the team identifies that the project, product or service uses AI
- a vendor describes its product or service as using AI
- users, the public or other stakeholders believe the project, product or service uses AI.

Cornerstones of assurance

In October 2023, the Department of Prime Minister and Cabinet (PM&C) published *How artificial intelligence might affect the trustworthiness of public service delivery?* (PM&C 2023).

The report identified that current trust in AI is low, and developing community trust would be a key enabler of government adoption of AI technology.

Alignment to *Australia's AI Ethics Principles*, developed by the CSIRO's Data61 and DISR, will ensure the trustworthy use of AI by governments in Australia. Each of its 8 ethics principles inform the assurance practices found in this framework and are also consistent with the Australian government's broader work on safe and responsible AI.

They will help governments demonstrate and achieve:

- safer, more reliable and fairer outcomes for all
- reduced risk of negative impact on those affected by AI
- the highest ethical standards when designing, developing and implementing AI.

To effectively apply the AI ethics principles, governments should also consider the following cornerstones for their assurance practices.

Governance

AI governance comprises the organisational structure, policies, processes, regulation, roles, responsibilities and risk management frameworks that ensures the safe and responsible use of AI in a way that is fit for the future.

The use of AI presents challenges that requires a combination of technical, social and legal capabilities and expertise. These cut across core government functions such as data and technology governance, privacy, human rights, diversity and inclusion, ethics, cyber security, audit, intellectual property, risk management, digital investment and procurement.

Implementation of AI should therefore be driven by business or policy areas and be supported by technologists.

Existing decision-making and accountability structures should be adapted and updated to govern the use of AI. This reflects the likely impacts upon a range of government functions, allows for diverse perspectives, designates lines of responsibility and provides clear sight to agency leaders of the AI uses they are accountable for.

Governance structures should be proportionate and adaptable to encourage innovation while maintaining ethical standards and protecting public interests.

At the agency level, leaders should commit to the safe and responsible use of AI and develop a positive AI risk culture to make open, proactive AI risk management an intrinsic part of everyday work.

They should provide the necessary information, training and resources for staff to have the knowledge and means to:

- align with the government's objectives
- use AI ethically and lawfully
- exercise discretion and judgement in using AI outputs
- identify, report and mitigate risks
- consider testing, transparency and accountability requirements
- support the community through changes to public service delivery
- clearly explain AI-influenced outcomes.

Data governance

The quality of an AI model's output is driven by the quality of its data.

It's therefore important to create, collect, manage, use and maintain datasets that are authenticated, reliable, accurate and representative, and maintain robust data governance practices that complies with relevant legislation.

Data governance comprises the policies, processes, structures, roles and responsibilities to achieve this and is as important as any other governance process. It ensures responsible parties understand their legislative and administrative obligations, see the value it adds to their work and their government's objectives.

Data governance is also an exercise in risk management because it allows governments to minimise risks around the data it holds, while gaining maximum value from it.

A risk-based approach

The use of AI should be assessed and managed on a case-by-case basis. This ensures safe and responsible development, procurement and deployment in high-risk settings, with minimal administrative burden in lower-risk settings.

The level of risk depends on the specifics of each case, including factors such as the business domain context and data characteristics. Self-assessment models, such as the *NSW Artificial Intelligence Assurance Framework*, help to identify, assess, document and manage these risks.

Risks should be managed throughout the AI system lifecycle, including reviews at transitions between lifecycle phases. The OECD defines the phases of an AI system as:

1. design, data and models - a context-dependent sequence encompassing planning and design, data collection, processing and model building.
2. verification and validation
3. deployment
4. operation and monitoring.

This AI system lifecycle may be embedded within the broader project management and procurement lifecycles, and risks may need re-evaluation where a significant change occurs at any phase.

During system development governments should exercise discretion, prioritising traceability for datasets, processes, and decisions based on the potential for harm. Monitoring and feedback loops should be established to address emerging risks, unintended consequences or performance issues. Plans should be made for risks presented by obsolete and legacy AI systems.

Governments should also consider oversight mechanisms for high-risk settings, including but not limited to external or internal review bodies, advisory bodies or AI risk committees, to provide consistent, expert advice and recommendations.

In focus: risk-based regulation

The Australian Government's 2023 'Safe and Responsible AI in Australia' consultation found strong public support for Australia to follow a risk-based approach to regulating AI.

As set out in the government's interim response, the government is now considering options for mandatory guardrails for organisations designing, developing and deploying AI systems in high-risk settings.

This work focuses on testing, transparency and accountability measures and is being informed by a temporary AI expert group.

Standards

Where practical, governments should align their approaches to relevant AI standards. Standards outline specifications, procedures, and guidelines to enable the safe, responsible, consistent, and effective implementation AI in a consistent and interoperable manner.

Some current AI governance and management standards include:

- *AS ISO/IEC 42001:2023 Information technology* - Artificial intelligence - Management system
- *AS ISO/IEC 23894:2023 Information technology* - Artificial intelligence - Guidance on risk management
- *AS ISO/IEC 38507:2022 Information technology* - Governance of IT - Governance implications of the use of artificial intelligence by organizations

Governments should regularly check the Standards Australia website for new AI related standards.

Procurement

Careful consideration must be applied to procurement documentation and contractual agreements when procuring AI systems or products. This may require consideration of:

- AI ethics principles
- clearly established accountabilities
- transparency of data
- access to relevant information assets
- proof of performance testing throughout an AI system's life cycle.

It is essential to remain mindful of the rapid pace of AI advancements and ensure contracts are adaptable to changes in technology.

Governments should also consider internal skills development and knowledge transfer between vendors and staff to ensure sufficient understanding of a system's operation and outputs, avoid vendor lock-in and ensure that vendors and staff fulfill their responsibilities.

Due diligence in procurement plays a critical role in managing new risks, such as transparency and explainability of 'black box' AI systems like foundation models. AI can also amplify existing risks, such as privacy and security. Governments must evaluate whether existing standard contractual clauses adequately cover these new and amplified risks.

Consideration should be made to a vendor's capability to support the review, ongoing monitoring or evaluation of a system's outputs in the event of an incident or a stakeholder raising concerns. This should include providing evidence and support for review mechanisms.

Governments may face trade-offs between a procured component's benefits and inherent assurance challenges, and resolutions will vary according to use case and tolerance threshold.

Ultimately, procurement should prioritise alignment with ethics principles alongside delivering on a government's desired outcomes.

In focus: responsible use of generative AI

Generative AI (also known as foundational models, large language models or LLMs) has garnered wide attention since the public release of ChatGPT in November 2022.

Whereas traditional AI has focused primarily on analysing data and subsequently making predictions, generative AI is able to create content across a wide range of mediums, including text, images, music and programming code, based on instructions or prompts provided by a user and informed by large datasets.

Recognising the potential and risk of generative AI, governments across Australia have released guidance for its use in the public service, including:

- *Interim guidance on government use of public generative AI tools* (DTA 2023)
- *Use of generative AI in Queensland Government* (Department of Transport and Main Roads, Queensland Government 2023)
- *Artificial Intelligence and public records* (Queensland State Archives 2024)
- *Public statement: Use of Microsoft Copilot for 365 in the Victorian public sector* (OVIC 2023)
- *Public Statement: Use of personal information with ChatGPT* (OVIC 2024)
- *Generative AI: basic guidance* (Department of Customer Service, NSW Government n.d.) and accompanying strategy, policy and practical resources
- *Guideline for the use of Large Language Model AI Tools and Utilities* (Department of Premier and Cabinet, Government of South Australia 2023).

Common across government guidance is focus on human oversight and human accountability for the use of content produced using generative AI to ensure compliance with policies, legal obligations and ethical principles.

This includes instructions on the use and protection of classified or sensitive information including personal information.

Implementing *Australia's AI Ethics Principles* in government

The following practices are mapped to Australia's 8 *AI Ethics Principles*, 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 center 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.

Resources

ASD (Australian Signals Directorate) (n.d.) *Information Security Manual (ISM)*, ASD, Australian Government, accessed 25 March 2024. https://www.cyber.gov.au/resources-business-and-government/essential-cyber-security/ism

——(n.d.) *Legacy ICT management*, ASD website, accessed 22 April 2024. https://www.cyber.gov.au/resources-business-and-government/maintaining-devices-and-systems/system-hardening-and-administration/legacy-ict-management

——(2024) *Deploying AI Systems Securely*, ASD, Australian Government, accessed 25 March 2024. https://www.cyber.gov.au/resources-business-and-government/governance-and-user-education/artificial-intelligence/deploying-ai-systems-securely

——(2024) *Engaging with Artificial Intelligence*, ASD, Australian Government, accessed 25 March 2024. https://www.cyber.gov.au/resources-business-and-government/governance-and-user-education/governance/engaging-with-artificial-intelligence

——(2023) *Guidelines for secure AI system development*, ASD, Australian Government, accessed 25 March 2024. https://www.cyber.gov.au/about-us/view-all-content/advice-and-guidance/guidelines-secure-ai-system-development

——(2021) *Identifying Cyber Supply Chain Risks*, ASD, Australian Government, accessed 25 March 2024. https://www.cyber.gov.au/resources-business-and-government/maintaining-devices-and-systems/outsourcing-and-procurement/cyber-supply-chains/identifying-cyber-supply-chain-risks

Attorney-General's Department (n.d.) *Australia's anti-discrimination law*, Attorney-General's Department website, accessed 25 March 2024. https://www.ag.gov.au/rights-and-protections/human-rights-and-anti-discrimination/australias-anti-discrimination-law

——(n.d.) *Human rights protections*, Attorney-General's Department website, accessed 25 March 2024. https://www.ag.gov.au/rights-and-protections/human-rights-and-anti-discrimination/human-rights-protections

——(n.d.) *Public sector guidance sheets*, Attorney-General's Department website, accessed 25 March 2024. https://www.ag.gov.au/rights-and-protections/human-rights-and-anti-discrimination/human-rights-scrutiny/public-sector-guidance-sheets

——(2023) *Countering the Insider Threat: A guide for Australian Government*, Attorney-General's Department, Australian Government, accessed 25 March 2024. https://www.ag.gov.au/integrity/publications/countering-insider-threat-guide-australian-government

Australian Government (2024) *Framework for Governance of Indigenous Data*, National Indigenous Australians Agency, Australian Government, accessed 30 May 2024. https://www.niaa.gov.au/resource-centre/framework-governance-indigenous-data

Australian Human Rights Commission (n.d.) *Human Rights and Technology Project*, AHRC website, accessed 25 March 2024. https://humanrights.gov.au/our-work/technology-and-human-rights

——(2021) *Using artificial intelligence to make decisions: Addressing the problem of algorithmic bias • Technical Paper*, AHRC, Australian Government, accessed 25 March 2024. https://humanrights.gov.au/our-work/technology-and-human-rights/publications/technical-paper-addressing-algorithmic-bias

Commonwealth Ombudsman (2020) *Automated Decision-making Better Practice Guide* [PDF 571KB], Commonwealth Ombudsman, Australian Government, accessed 25 March 2024. https://www.ombudsman.gov.au/__data/assets/pdf_file/0029/288236/OMB1188-Automated-Decision-Making-Report_Final-A1898885.pdf

CSIRO (n.d.) *National Artificial Intelligence Centre*, CSIRO website, accessed 25 March 2024. https://www.csiro.au/en/work-with-us/industries/technology/National-AI-Centre

——(n.d.) *Responsible AI Network resources*, CSIRO website, accessed 25 March 2024. https://www.csiro.au/en/work-with-us/industries/technology/National-AI-Centre/Responsible-AI-Network/Responsible-AI-Network-resources-archive

——(2023) *Diversity and Inclusion in Artificial Intelligence*, CSIRO website, accessed 25 March 2024. https://research.csiro.au/ss/team/diai/

——(2023) *Responsible AI Pattern Catalogue, CSIRO*, Australian Government, accessed 25 March 2024. https://research.csiro.au/ss/science/projects/responsible-ai-pattern-catalogue/

——(2019) *Artificial Intelligence: Australia's Ethics Framework*, CSIRO, Australian Government, accessed 25 March 2024. https://www.csiro.au/en/research/technology-space/ai/ai-ethics-framework

Department of Customer Service (n.d.) Generative AI: basic guidance, Department of Customer Service, NSW Government, accessed 25 March 2024. https://www.digital.nsw.gov.au/policy/artificial-intelligence/generative-ai-basic-guidance

——(2022) *Artificial Intelligence (AI)*, Department of Customer Service website, accessed 25 March 2024. https://www.digital.nsw.gov.au/policy/artificial-intelligence

——(2022) *NSW Artificial Intelligence Assurance Framework*, Department of Customer Service, NSW Government, accessed 22 March 2024. https://www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-artificial-intelligence-assurance-framework

Department of Finance (2023) *Risk Management Toolkit*, Finance website, accessed 25 March 2024. https://www.finance.gov.au/government/comcover/risk-services/management/risk-management-toolkit

DISR (Department of Industry, Science and Resources) (2024) *Safe and responsible AI in Australia consultation: Australian Government's interim response*, DISR, Australian Government, accessed 22 March 2024. https://consult.industry.gov.au/supporting-responsible-ai

——(2024) *The Seoul Declaration by countries attending the AI Seoul Summit, 21-22 May 2024*, DISR, Australian Government, accessed 27 May 2024. https://www.industry.gov.au/publications/seoul-declaration-countries-attending-ai-seoul-summit-21-22-may-2024

——(2023) *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023*, DISR, Australian Government, accessed 22 March 2024. https://www.industry.gov.au/publications/bletchley-declaration-countries-attending-ai-safety-summit-1-2-november-2023

——(2019) *Australia's AI Ethics Principles*, DISR website, accessed 22 March 2024. https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles

——(2019) *Australia's Artificial Intelligence Ethics Framework*, DISR website, accessed 22 March 2024. https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework

DTA (Digital Transformation Agency) (2023) *Adoption of Artificial Intelligence in the Public Sector*, Australian Government Architecture website, accessed 25 March 2024. https://architecture.digital.gov.au/adoption-artificial-intelligence-public-sector-0

——(2019) *Interim guidance on government use of public generative AI tools*, Australian Government Architecture website, accessed 25 March 2024. https://architecture.digital.gov.au/generative-ai

Government of South Australia (2023) *Guideline for the use of Large Language Model AI Tools and Utilities*, Office of the Chief Information Officer, Department of the Premier and Cabinet, Government of South Australia, accessed 25 March 2024. https://www.dpc.sa.gov.au/responsibilities/ict-digital-cyber-security/policies-and-guidelines/artificial-intelligence

——(n.d.) *Online Accessibility Toolkit* [website], accessibility.sa.gov.au, accessed 25 March 2024.

Hiroshima AI Process (2023) *Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems*, Ministry of Internal Affairs and Communications, Government of Japan, accessed 25 March 2024. https://www.soumu.go.jp/hiroshimaaiprocess/en/documents.html

Home Affairs (Department of Home Affairs) (n.d.) *Hosting Certification Framework* [website], hostingcertification.gov.au, accessed 22 April 2024.

——(n.d.) *Protective Security Policy Framework (PSPF)* [website], protectivesecurity.gov.au, accessed 25 March 2024.

——(2023) *2023-2030 Australian Cyber Security Strategy*, Home Affairs, Australian Government, accessed 22 April 2024. https://www.homeaffairs.gov.au/about-us/our-portfolios/cyber-security/strategy/2023-2030-australian-cyber-security-strategy

ISO (International Organization for Standardization) (2024) *ISO/IEC DIS 42005 - Information technology — Artificial intelligence — AI system impact assessment*, ISO, accessed 25 March 2024. https://www.iso.org/standard/44545.html

——(2022) *Artificial intelligence*, ISO website, accessed 25 March 2024. https://www.iso.org/sectors/it-technologies/ai

NSW Ombudsman (2021) *Automated decision-making in the public sector*, NSW Ombudsman website, accessed 25 March 2024. https://www.ombo.nsw.gov.au/guidance-for-agencies/automated-decision-making-in-the-public-sector

OAIC (Office of the Australian Information Commissioner) (n.d.) *Privacy impact assessments*, OAIC website, accessed 25 March 2024. https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/privacy-impact-assessments

——(n.d.) *State and territory privacy legislation*, OAIC website, accessed 25 March 2024. https://www.oaic.gov.au/privacy/privacy-legislation/state-and-territory-privacy-legislation/state-and-territory-privacy-legislation

——(2018) *Guide to data analytics and the Australian Privacy Principles*, OAIC, Australian Government, accessed 25 March 2024. https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/more-guidance/guide-to-data-analytics-and-the-australian-privacy-principles

OECD (Organization for Economic Cooperation and Development) (2024) *Explanatory memorandum on the updated OECD definition of an AI system*, OECD, accessed 25 March 2024. https://www.oecd.org/publications/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system-623da898-en.htm

——(2023) *OECD AI Principles overview*, OECD website, accessed 22 March 2024. https://oecd.ai/en/ai-principles

——(2023) *Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449*, OECD, accessed 22 March 2024. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

OVIC (Office of the Victorian Information Commissioner) (2024) *Public Statement: Use of personal information with ChatGPT*, OVIC, State Government of Victoria, accessed 27 May 2024. https://ovic.vic.gov.au/privacy/resources-for-organisations/public-statement-use-of-personal-information-with-chatgpt/

——(2023) *Public Statement: Use of Microsoft 365 Copilot in the Victorian public sector*, OVIC, State Government of Victoria, accessed 25 March 2024. https://ovic.vic.gov.au/privacy/resources-for-organisations/vps-use-of-microsoft-365-copilot/

——(2018) *Artificial Intelligence – Understanding Privacy Obligations*, OVIC, State Government of Victoria, accessed 25 March 2024. https://ovic.vic.gov.au/privacy/resources-for-organisations/artificial-intelligence-understanding-privacy-obligations/

——(2018) *Artificial Intelligence and Privacy – Issues and Challenges*, OVIC, State Government of Victoria, accessed 25 March 2024. https://ovic.vic.gov.au/privacy/resources-for-organisations/artificial-intelligence-and-privacy-issues-and-challenges/

PM&C (Department of Prime Minister and Cabinet) (2023) *How might artificial intelligence affect the trustworthiness of public service delivery?*, PM&C, Australian Government, accessed 22 March 2024. https://www.pmc.gov.au/resources/long-term-insights-briefings/how-might-ai-affect-trust-public-service-delivery

Public Record Office Victoria (2024) *Artificial Intelligence*, PROV website, accessed 25 March 2024. https://prov.vic.gov.au/recordkeeping-government/a-z-topics/AI

——(2024) *Artificial Intelligence (AI) Technologies and Recordkeeping Policy*, PROV, State Government of Victoria, accessed 25 March 2024. https://prov.vic.gov.au/recordkeeping-government/document-library/ai-technologies-policy-ai-technologies-and-recordkeeping

Queensland Government Customer and Digital Group (2023) *Use of generative AI in Queensland Government*, Department of Transport and Main Roads, Queensland Government, accessed 25 March 2024. https://www.forgov.qld.gov.au/information-and-communication-technology/qgea-policies-standards-and-guidelines/use-of-generative-ai-in-queensland-government

Queensland State Archives (2024) *Artificial Intelligence and public records*, Queensland State Archives website, accessed 25 March 2024. https://www.forgov.qld.gov.au/information-and-communication-technology/recordkeeping-and-information-management/recordkeeping/resources-and-tools-for-records-management/artificial-intelligence-and-public-records

Reid A, O'Callaghan S and Lu Y (2023) *Implementing Australia's AI Ethics Principles: A selection of Responsible AI practices and resources*, Gradient Institute and CSIRO, accessed 25 March 2024. https://www.csiro.au/en/work-with-us/industries/technology/national-ai-centre/implementing-australias-ai-ethics-principles-report

Zowghi D and da Rimini F (2023) 'Diversity and Inclusion in Artificial Intelligence', in Lu Q et al. (eds) *Responsible AI: Best Practices for Creating Trustworthy AI Systems*, Addison-Wesley, Sydney. https://research.csiro.au/ss/guidelines-for-diversity-and-inclusion-in-artificial-intelligence/