

Policy for the **responsible use of artificial intelligence** in the Tasmanian Government

Part 4 - Guidance on assessment of benefits and risks

July 2026 | Version 1.0

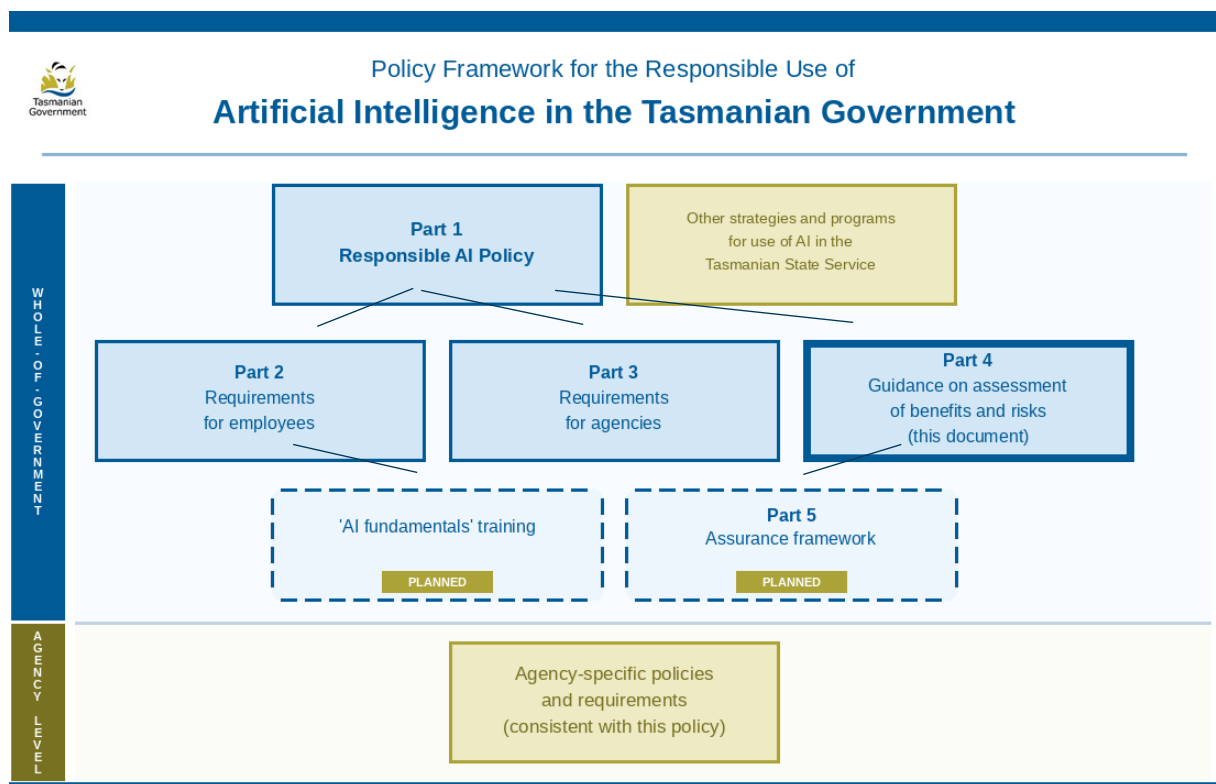
Purpose

This policy establishes a consistent ‘baseline’ approach to the use of AI by the Tasmanian Government, so that agencies can harness its benefits, while ensuring its use is **ethical**, **responsible**, and **safe**. This is essential for public trust in government use of AI.

Each government agency remains responsible for its own oversight, assurance, and implementation.

Elements

The policy framework is set out here:



Part 4 provides an optional framework and common terms for assessing the benefits and risks of AI. While using Part 4 is optional, agencies are still required to carry out their own risk assessments in accordance with Parts 1–3 of this policy.

Agencies may adopt the approach to rating risks for AI tools, as outlined in this Part, or adopt an equivalent approach using their agency’s existing risk framework. Additional guidance on assessing AI benefits and risks will be included in Part 5 – Assurance framework (planned).

Introduction

This guidance outlines some types of potential benefits from AI systems. However, identifying and realising these benefits is largely the responsibility of individual agencies, teams and employees.

This guidance focuses mainly on potential AI risks and how they can be managed.

Agencies are expected to manage these risks within their existing risk management frameworks.

This guidance is adapted from parts of the [NSW Artificial Intelligence Assessment Framework](#) (updated January 2026). Agencies are encouraged to review the full NSW framework, especially when considering higher risk AI applications.

The NSW framework is a recommended approach for managing AI risks under the [National framework for the assurance of artificial intelligence in government](#) (June 2024).

This guidance is an interim resource, pending development of a Tasmanian Government AI Assessment Framework.

This guidance covers the following topics:

- Potential benefits from AI systems
- Is your AI system a potential elevated risk?
- Guidance for AI systems with potential elevated risks
- AI risk levels
- Types of potential harm
- Could the AI system cause harms that are irreversible?
- Manage potential AI risks with good data governance
- Manage potential AI risks with employee and agency capabilities
- Resources

Potential benefits from AI systems

When assessing whether to use an AI system, employees and agencies are encouraged to consider how well it aligns with government policy and strategy, and whether it has the potential to:

- deliver a higher quality service or outcome, such as improved accuracy or client satisfaction
- reduce processing or delivery times
- improve the scale and speed of data analysis, and strengthen the evidence used for recommendations and decisions

- enhance monitoring and reporting
- reduce routine tasks for employees, allowing them to focus on higher value work
- generate financial efficiencies or cost savings
- deliver a new service or outcome
- support innovations to existing services or outcomes
- provide an AI capability that other agencies could also use or adapt.

AI risk levels

It is recommended that agency risk assessments focus on uses assessed as **Mid-Range Risk, or higher**:

None, negligible, or N/A Risk	Low Risk: Reversible with negligible consequences	Mid-range Risk: Reversible with moderate consequences	High Risk: Reversible with significant consequences	Very High Risk: Significant or irreversible consequences
Can't create harm	Can be reversed or corrected without causing harm.	Can be reversed or corrected but may cause inconvenience, disruption or harm.	Can be reversed but may cause significant damage or harm with potential lasting impact.	Cannot be reversed, may cause permanent or irreversible harms.
Examples <ul style="list-style-type: none"> Noise suppression on audio calls Image resolution enhancements Grammar and spell checking Text summarisation of non-sensitive content Search functions in browsers Analytics report 	Examples <ul style="list-style-type: none"> Anomaly detection software. Email spam filters Document classification and tagging Photo organising Non-critical content translation Voice assistance for basic tasks, i.e. Automated phone menu 	Examples <ul style="list-style-type: none"> Customer service chatbots Recommendation systems Language translation tools Content curation Predictive maintenance Natural language processing of gov documents. 	Examples <ul style="list-style-type: none"> Facial recognition systems. AI powered hiring and recruitment Autonomous emergency response system Autonomous tram with human oversight Healthcare decision support systems. Adaptive learning system 	Examples <ul style="list-style-type: none"> Autonomous benefits eligibility without human oversight Self-driving cars Predictive reoffending Medical diagnosis without oversight Autonomous AI systems on critical infrastructure (i.e. energy)

Source: New South Wales Government, Digital NSW - [NSW Artificial Intelligence Assessment Framework](#)

These higher-level risks are also the focus of agency AI Transparency Statements and the Register of AI Applications (see [Part 3 – Policy for agencies](#)).

Types of potential harm

These risk ratings consider the *severity* and *irreversibility* of potential AI harms.

The harms listed here are not exclusive to AI, but poorly controlled AI has potential to amplify these harms:

- physical or psychological harm (including trauma)
- environmental harm
- breach of privacy
- unauthorised use of health or other sensitive personal information
- unintended identification or misidentification of a person

- impacts on a right, privilege or entitlement
- breach of intellectual property rights
- breach of information security
- incorrect application of a fine or penalty
- other financial or commercial impacts
- incorrect advice or guidance
- inconvenience or delay
- economic disruption or impact
- lack of procedural fairness
- increased social inequality
- reduced trust in government policy or service delivery.

Could the AI system cause harms that are irreversible?

Irreversible harm refers to situations where it is not possible to return to the state that existed before the harm occurred.

You should ensure there is a way to overturn outcomes if harm occurs, or if the AI system makes incorrect decisions.

If potential harms are assessed as Mid-Range Risk or higher, it is recommended that you do not proceed until legal advice is obtained.

If the potential harms are unclear, it is recommended that you pause the project and consult your agency's Accountable Officer on how to resolve the issue.

If legal approval is provided, discuss the proposal with relevant stakeholders and consider whether additional impact assessments (such as ethics, privacy or human rights) are required.

Is your AI system a potential elevated risk?

Elevated risks include AI systems that influence decisions with legal or similar consequences; trigger significant actions; operate autonomously; use sensitive data; create a risk of harm; or have limited explainability.

Operational impacts

Does the system produce, or directly influence, an administrative decision (a government decision with legal or similar significant effect)?

- Example: automated systems used to issue infringements.

OFFICIAL

Does the system trigger a real-world action with more than a negligible effect (meaningful change to the environment, or to a system state)?

- Example: an automated alerting system.

Does your system produce or directly influence an administrative decision (government decision with legal or similar significant effect)?

- Example: automating decisions on issuing infringements.

Does your system trigger a real-world action with more than negligible potential effect (meaningful change to environment, or system state)?

- Example: automated alerting system.

Autonomy

Does the system operate autonomously or have the potential to produce harmful outputs without human intervention?

- Example: autonomous vehicles.

Sensitive data

Was the system trained on sensitive information, or can it produce outputs that contain sensitive information?

- Example: biometric-based face-matching system.

Unintended harms

Is there a risk that system failure, misuse, or inappropriate deployment could cause harm to an individual or group?

- Example: using data inputs that have not been verified.

Explainability and transparency

Does the system lack explainability for its outputs or decisions, making it difficult for non-experts to understand and for technical experts to assess?

- Example: limited traceability of data used to inform policy development.



If you answer **yes** to any of these questions, your AI use is likely to be at **elevated risk**, and you should consider the advice and mitigation measures provided below.

If you are unsure, assume that the AI use is potentially at elevated risk.

Guidance for AI systems with potential elevated risks

Some elevated-risk uses of AI (for example, in Health) may be necessary to improve existing processes, or because they deliver clear benefits to the community. It is important to consider the risks of not using AI if it is the best available option.

Independent evaluation and monitoring should be used to identify potential harms at different stages of the system lifecycle. For elevated risks, the level of independence in the review process should be increased.

Language models and generative AI used for decision-making, prioritisation or automation require particular care. Their outputs must be validated, and final decisions must always be made by an appropriately authorised and qualified person.

For elevated-risk AI uses, consider pausing the project until performance measures and targets (such as accuracy and fairness) are clearly defined.

For elevated-risk AI systems that continue to train and adapt, regular performance monitoring and calibration are strongly recommended. These systems can quickly move outside normal operating parameters. The higher the residual risk, the more frequently monitoring should occur.

Monitoring mechanisms may include red-teaming, conformity assessments, reinforcement from human feedback, checks for model drift, and metrics-based performance testing.

You should consult any affected community when designing or implementing the AI system. This is especially important for elevated-risk uses.

Provide clear and simple explanations of how AI is used to those who may be affected, particularly when risks are elevated.

Maintain strong human accountability and the ability to intervene. Avoid becoming overconfident or overly reliant on the system. Intervene if stakeholders raise concerns about insights, decisions, or outputs. Ensure clear processes for appeal and redress.

Operational AI systems with High or Very High residual risks should plan for regular, external, independent risk audits to assess the effectiveness of mitigation measures.

For elevated-risk AI uses (Mid-Range Risk or higher), the public benefits should be clear and documented before proceeding.

Manage potential AI risks with good data governance¹

Data governance is a key part of risk management because it helps government minimise risks associated with the data it holds, while maximising its value.

The quality of an AI model's output depends directly on the quality of its data.

It is therefore important to create, collect, manage, use and maintain datasets that are authenticated, reliable, accurate and representative, and to apply strong data-governance practices that comply with relevant legislation.

Data governance includes the policies, processes, structures, roles and responsibilities needed to achieve this, and is as important as any other governance process.

It helps ensure responsible staff understand their legislative and administrative obligations and recognise the value that good data practices bring to their work and to government objectives.

Manage potential AI risks with employee and agency capabilities

Many potential risks associated with AI systems can be managed by ensuring that Tasmanian Government employees and agencies develop appropriate capabilities.

For more information, see: [Part 2 – Policy for employees](#), and [Part 3 – Policy for agencies](#).

Useful resources

- Australian Government, Department of Finance - [National framework for the assurance of artificial intelligence in government](#), June 2024.
- New South Wales Government, Digital NSW - [NSW Artificial Intelligence Assessment Framework](#), July 2024.

References

¹ Commonwealth of Australia (2024), above, p.7.

AI Use and Readability Disclosure

This policy includes text revised with Microsoft 365 Copilot to about Year 9 readability. The target reflects Australian Government Style Manual guidance to write for lower-secondary readers and supports ISO 24495-1:2023 (Plain language) principles. All revisions underwent human review for accuracy.