Digital NSW

The NSW AI Assessment Framework

2024

Ministers Foreword

NSW commitment to safe and responsible use of AI

“Artificial intelligence offers the opportunity to create a safer and more productive world, and we must do so responsibly, safely, and ethically”

Artificial Intelligence (AI) is transforming how we work and the government services we provide for the people of NSW. It presents significant opportunities to enhance productivity, drive economic growth, and improve the way we live and work. While the NSW Government must be ready to embrace these opportunities, we must do so in a safe, ethical, and responsible way. Transparency about the information and approaches we use is also critical to maintaining public trust in government.

The NSW Government is a leader in the safe and responsible use of AI. On 1 July 2024, we released an update to our pioneering AI Assessment Framework. This improvement addresses new and emerging risks and opportunities, continuing to set the standard for managing AI. By setting appropriate guardrails for the design, deployment and use of AI, we can ensure we meet the expectations of our community and uphold the highest ethical standards.

Use of AI in the NSW Government is not new. We are already using AI to make our communities safer – for example, in bushfire intelligence, measuring the health of our environment via smart sensors, and enhancing the education of our youth through services like the NSW EduChat. The introduction of emerging technologies, such as generative AI, can enable us to deliver better, safer outcomes for the people of NSW.

Our efforts to harness the opportunities of innovation through emerging technologies must be underpinned with measures to appropriately manage risk. The Department of Customer Service has updated the NSW Digital Assurance Framework and the NSW AI Assessment Framework to incorporate the development of generative AI. These updates enhance the way we manage risk and are mandatory for NSW Government agencies. We are also making sure we have the right expertise to help us review high-risk projects through our AI Review Committee.

The NSW Government is committed to ensuring the safe, ethical, and responsible deployment of AI across NSW. As we continue to collaborate with industry, academia, and our Commonwealth, state and territory colleagues, I am confident that our approach will enable us to remain dedicated to deploying AI responsibly, and always with a view to delivering the best outcomes for our communities.

The Hon Jihad Dib MP

Minister for Customer Service and Digital Government
Minister for Emergency Services
Minister for Youth Justice
Member for Bankstown

Introduction

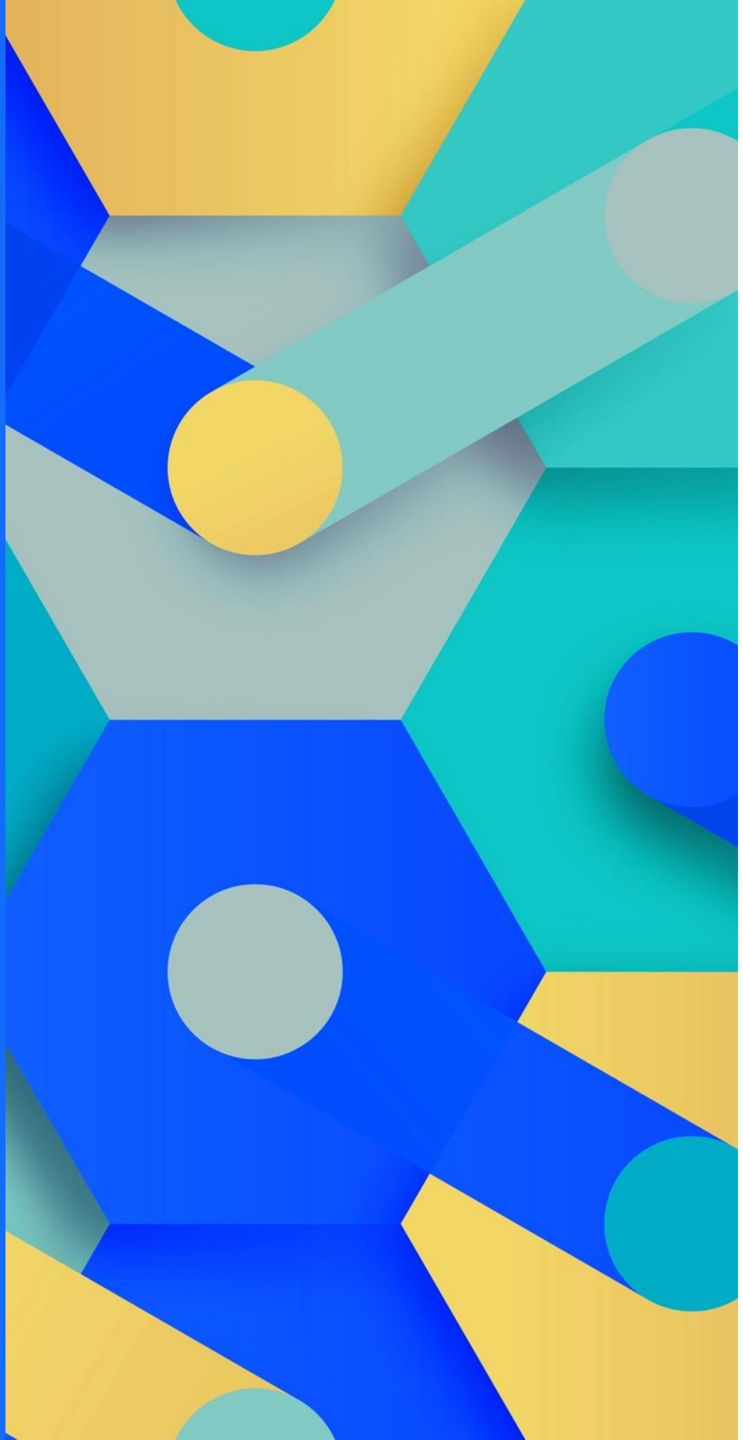
Artificial Intelligence (AI) is the ability of a computer system to perform tasks that would normally require human intelligence, such as learning, reasoning, and making decisions. AI encompasses various specialised domains that focus on different tasks and includes automation.

This AI Assessment framework is a self-assessment, intended to be applied during all phases of development, training and use of AI. Apply the Framework before you deploy your AI system, as well as after deployment to ensure appropriate monitoring of performance.

Systems with High levels of residual risk must be reviewed by the NSW AI Review Committee.

Contents

About the AI Assessment Framework



About the AI Assessment Framework

What is it?

The AI Assessment Framework, mandated since March 2022, guides responsible and safe AI usage in the NSW Government. It assists project teams and solution owners to analyse AI system risks, implement mitigation controls, and establish accountabilities.

Who should use it?

For all NSW Government Agencies, the Framework is necessary when designing, developing, deploying, procuring, or using systems containing AI components. It's applicable for project sponsors, executive business sponsors, technical leads, and data governance leads involved in business operations or projects utilising AI technology.

Elevated risk

The framework has a focus on what is referred to as elevated risk. Elevated risk involves systems influencing decisions with legal or similar level consequences, triggering significant actions, operating autonomously, using sensitive data, risking harm, and lacking explainability. All Generative AI solutions should be classified elevated risk. Elevated risk is referenced throughout the self-assessment with suggested mitigation.

Is applying the Framework everything I need to do?

The framework is not a complete list of all requirements for AI systems. Project teams and system owners should follow industry standards, agency-specific AI governance and assurance processes, and foster a positive AI risk culture.

When should you apply the framework?

The Framework should be used during all phases of the project and system lifecycle. Each time the framework is applied, it should build on the previous assessment.

ⓘ When you do not need to apply the framework

All NSW Agencies must ensure the safe and responsible use of AI, holding themselves accountable for risks outlined in this framework, tailored to each use case.

As a guide, you may not need to apply the framework to assess your product or service if you are using a widely available commercial application (which you are not training or customising) or conducting exploratory research that does not meet the criteria of elevated risk set on page 10.

For more guidance on when the Framework may not be needed, see page 12.

Alignment to NSW Ethics Policy

Mandatory principles

This AI Assessment Framework is structured in sections that align to the AI Ethics Principles defined in the NSW Ethics Policy. The NSW Ethics Policy and this framework are mandatory for all NSW Government Agencies using AI.

Community benefit

AI must prioritise community outcomes, ensuring alignment with laws, minimising harm, and maximising benefit.

Fairness

Use of AI will be fair, ensuring not to perpetuate bias and inequality by leveraging diverse representative datasets, monitoring datasets, monitoring performance, and using rigorous data governance.

Privacy and security

Ensure secure, transparent, compliant data use, and adhere to PPIP Act preserving public trust.

Transparency

The use of AI will be transparent, allowing concerns to be raised and addressed, GIPA Act compliant, cyber secure and ethical.

Accountability

Decision-making remains the responsibility of organisations and Responsible Officers.

Note the principles statements and descriptions may offer more detail than the current AI ethics policy if required to describe the detailed framework content.

How to conduct the self-assessment

Main Triggers

Change project stage: Exploring potential use of AI or changing project stage (.i.e., design to deploy) →

Systems update: Updating systems that include AI or will include AI. →

Ongoing: Existing AI system, unassessed or requiring periodic review (refer slide 13). →

Refer to slide 13 for more triggers

Scope of applying the framework →

Determine whether your system / project should use the AIAF.

Self-assessment readiness →

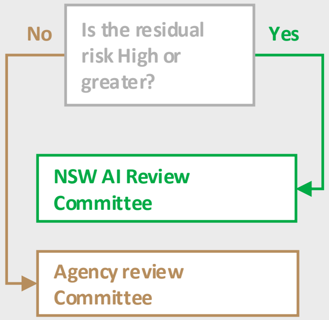
Recognise the importance of assessing AI benefits and risks, understand the structured self-assessment process, and identify responsible officers.

Self-assessment and risk summary →

Complete the self-assessment and review the summary risk.

Self-assessment mitigation and next steps →

Identify mitigations and controls to implement and next steps based on the highest residual risk.


Is the residual risk High or greater?

No

Yes

NSW AI Review Committee

Agency review Committee

Ongoing monitor & evaluate

Considerations for ongoing monitoring and evaluation of your system.

Primary focus of the AIAF, the Self-Assessment

High level guidance provided in previous step.

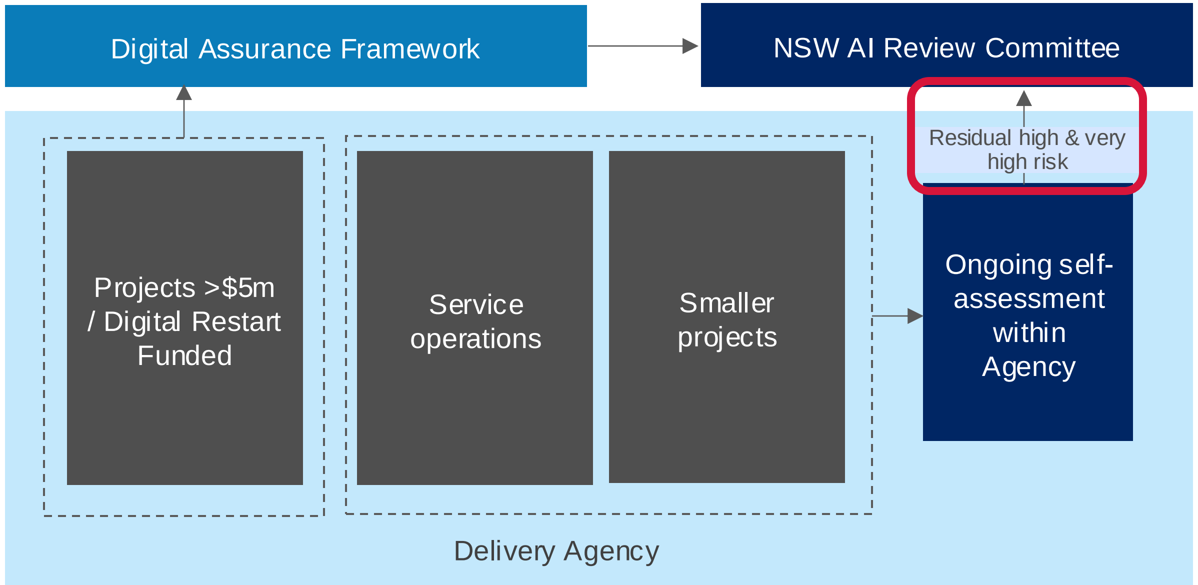
When to submit a self-assessment to the NSW AI Review Committee

After completing the self-assessment

Residual high and very high-risk projects/system must be submitted to the AI Review Committee. There are two ways to engage the AI Review Committee.

1. Guidance will be provided via the NSW Digital Assurance Framework when you register your project for: Projects >$5m, Digital Restart Funded projects

2. Direct via emailing the AI secretariate, for: Projects <$5m, Operational / ongoing system


Digital Assurance Framework
NSW AI Review Committee
Residual high & very high risk
Projects >$5m / Digital Restart Funded
Service operations
Smaller projects
Ongoing self-assessment within Agency
Delivery Agency

Recommendations from the NSW AI Review Committee

The AI Review Committee provides feedback and recommendations to improve the AI system. The Agency Responsible Officers remain responsible for implementing the mitigations, the impact and the outcomes.

AI Secretariate contact: AISecretariat@customerservice.nsw.gov.au

Completing the assessment

In all cases, the self-assessment is to be completed by (or the result confirmed with) the Responsible Officers and stored within agency records management system. Refer NSW State Records Act.

1 Scope of applying the framework

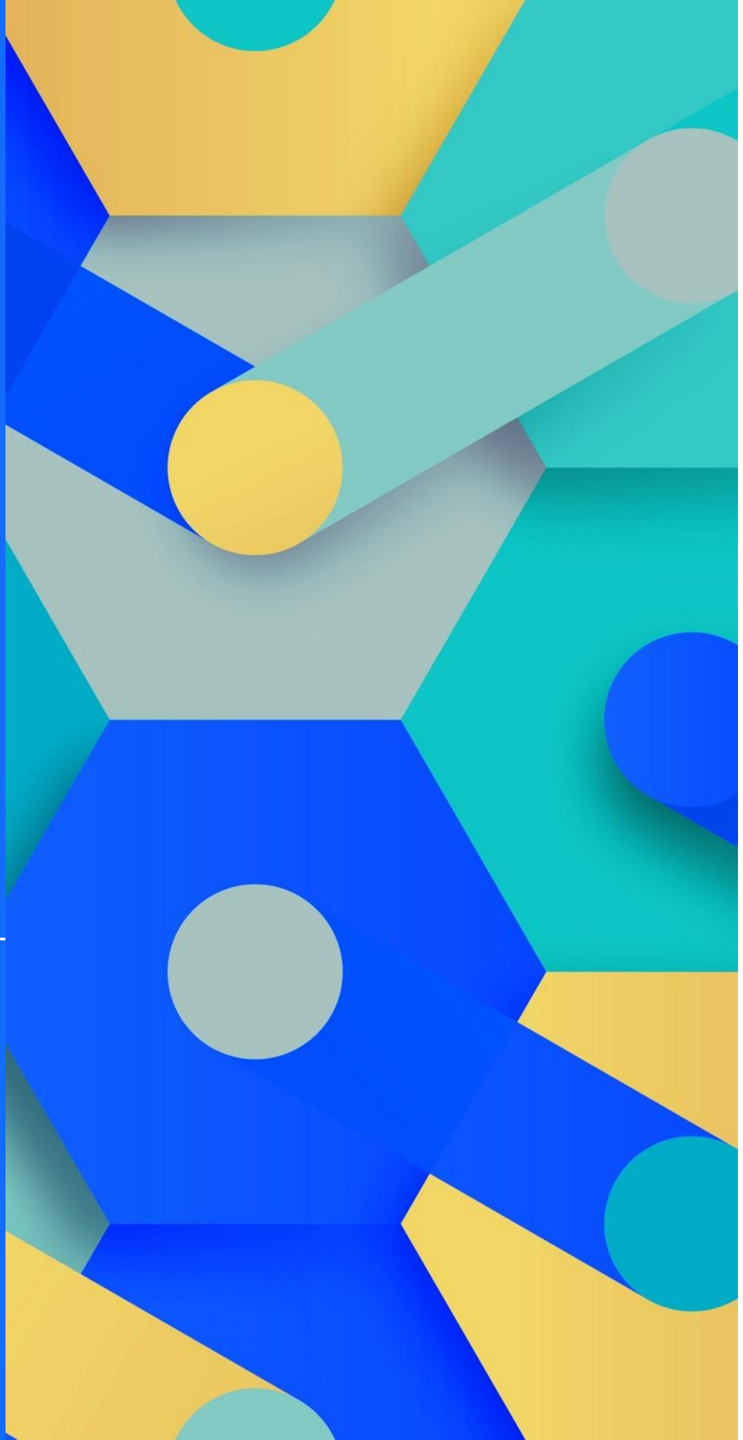
Determine whether your solution / project should use the AIAF and when.

1 Scope of applying the framework

2 Self-assessment readiness

3 Self-assessment & risk summary

4 Self-assessment mitigation & next steps



Is your system a potential elevated risk?

Evaluate potential elevated risk prior to starting the self-assessment as it is used though-out the self-assessment.

Operational impact

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

i.e, automating decisions on issuing infringements.

Operational Impact

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

i.e., an automated alerting system.

Autonomous

Does your system operate autonomously or have potential to produce harmful outputs independently of human action, without requiring manual initiation?

i.e., autonomous vehicles.

Data Sensitivity

Was any part of your system trained using sensitive information or can it produce outputs which contain sensitive information?

i.e. a biometric based face matching system.

Unintended harms

Is there a risk of system failure, misuse, or inappropriately deployed that could cause harm to an individual or group?

i.e., systems using unverifiable data inputs.

Explainability and Transparency

Does your system fail to provide explainability for generated content and decisions, hindering comprehension by laypeople and assessment by technical experts?

i.e., information informing policy development

→ Yes

Yes, to any questions means your use is potentially at elevated risk, and additional mitigation covered in this framework will apply.

→ No

No to all questions means you are not using AI in a manner which is potentially elevated risk.

ⓘ Elevated risk

To determine whether your use is potentially at elevated risk, you will need to make a judgement based on your specific use case. If you are unsure, assume that your use is potentially at elevated risk

Can I still use AI for a potentially elevated risk?

The range of considerations to ensure the appropriate use of AI will vary considerably across different use cases. Ultimately, it is a decision made within the Agency, considering all other alternatives and whether the use of the solution leads to better outcomes compared to taking no action at all.

Care should be taken to ensure independent evaluation and monitoring for potential harms at different stages of the system lifecycle. The level of independence in the review process should be heightened for elevated risks. Guidance is provided during the self-assessment process.

Language models and generative AI used for decision making, prioritisation or automation, require special care around output validation, ensuring a final decision is made by an appropriately authorised and qualified person.

For more information on considerations and risks associated with automated administrative decision making, see the NSW Ombudsman's Automated Decision-Making in the public sector resources.

Do I need to use the framework?

Does your system look like (or have elements of) the following?

Buy AI and use

Buying or using an off the shelf system. Used without modifying the algorithm or any risk mitigation tools, nor adding domain-specific content.

i.e. ChatGPT, or AI in Salesforce, SAP, etc.

Embed AI and/or co-train

Developing a product with embedded AI or purchasing an AI platform and augmented training data with domain-specific content.

i.e., integrating AI biometrics or developing a chatbot with augmented training.

Develop AI and/or train

Developing an AI tool in-house. Even if based on a standard platform, I am developing algorithms and supplying the training data.

i.e. Developing anomaly detection or training LLM with domain-specific content.

Automating decisions

Developing a tool in-house that uses AI and that automates at least one critical step in the decision-making process.

i.e. AI powered hiring and recruitment.

If the answer is Yes to <u>any</u> answer, consult the guidance to when the framework may not be needed.

→ Yes

→ No

If the answer is NO to <u>all</u> questions, there is no need to use the framework unless you have AI risk concerns.

ⓘ Guide: When the framework <u>may not</u> be needed

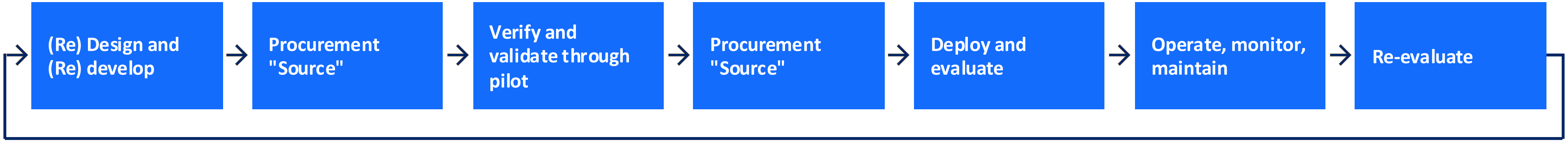
All NSW Agencies must ensure the safe and responsible use of AI. Like any digital solution and data usage, accountability lies with the agency.

This framework identifies key risks and harms to be mitigated. Its applicability hinges on agency specific use cases and nature of business.

As a guide, you may not need to apply the framework if you are using a widely available commercial application (which you are not training or customising) or conducting exploratory research that does not meet the criteria of elevated risk. If you are unsure, apply the Framework and always ensure you record your decisions in your records management system.

Note, Digital NSW is presently collaborating with agencies to devise compliance plans. This initiative aims to enhance transparency in the government's efforts to ensure the responsible and safe utilisation of AI, which is essential for building confidence and trust.

When you need to apply the framework


(Re) Design and (Re) develop
Procurement "Source"
Verify and validate through pilot
Procurement "Source"
Deploy and evaluate
Operate, monitor, maintain
Re-evaluate

Apply the AI Assessment Framework over the system lifecycle

After initially applying the framework, reassess your risk against it at each project phase and throughout the system lifecycle, following the frequencies recommended by your Agency Assurance function or the AI review Committee. All uses of AI should be piloted to verify correct operation, and then evaluated before being deployed at scale.

Aligning with Procurement

It's important to ensure that AI uses involving procurement of products or services identify and track associated risks and mitigations. This ensures they aren't overlooked during procurement, and suppliers' responsibilities are maintained throughout the lifecycle of the system. For more information contact: procurement at ICTServices@customerservice.nsw.gov.au.

Ongoing risk assessment beyond project phases

Beyond project phases, consider scenarios for when you should reassess your risk against the framework, including some examples:

- You are about to introduce the use of AI into an existing system
- You have discovered that you are using AI but haven't applied the AIAF.
- You or a supplier of your system are about to change the data, algorithm, model, or technology.
- You are about to change the purpose, use case, or intended use of the system.
- You are considering altering the level of human oversight or involvement.

2 Self-assessment readiness

Recognise the importance of assessing AI benefits and risks, understand the structured self-assessment process, and identify responsible officers.

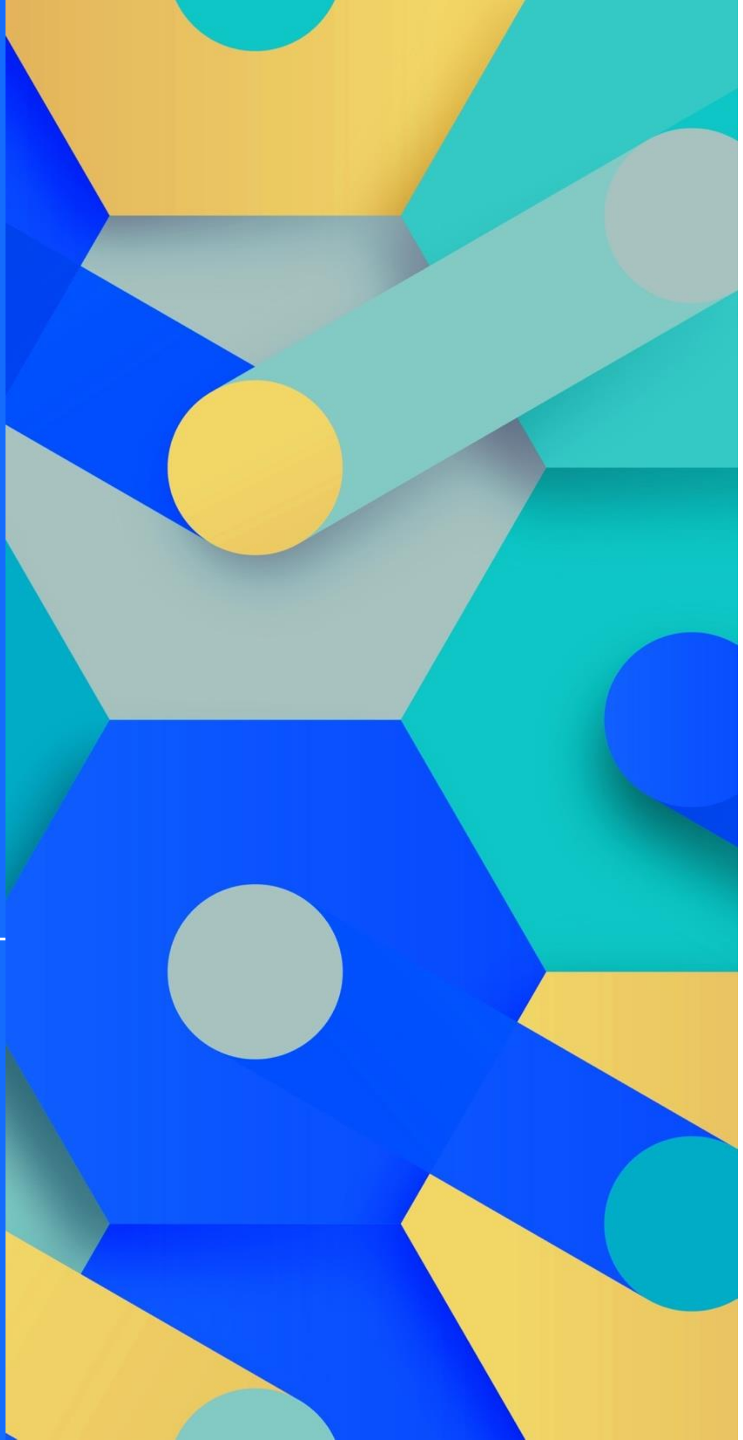
The start of the self-assessment

1 Scope of applying the framework

2 Self-assessment readiness

3 Self-assessment & risk summary

4 Self-assessment mitigation & next steps



Framework benefits and risks overview

Benefits and risks

AI can amplify existing risks; this means you need to carefully consider the risks and benefits.

Currently, we use AI tools to:

- deliver insights that improve services and lives
- help agencies work more quickly and accurately

While there are many areas where AI can benefit the work we do, we need to engage with risks early and throughout the solution lifecycle.

ⓘ Cannot answer some questions?

It is important to make a note of questions you cannot answer as you progress through the assessment. It may be because information is not available or can only be answered once a pilot is undertaken.

If the project/system proceeds, treat these unanswered questions as representing Mid-range risk, commence with a pilot phase and closely monitor for harms and establish controls.

Understanding the AI self-assessment process

This AI Assessment Framework is structured in sections that align to the AI Ethics Principles.

Each section contains questions for assessing specific risks and advise actions for mitigation. Proceed with completing the self-assessment and then implement required mitigations. Some controls may require pausing or stopping the project until necessary information is confirmed.

If your solution is already operational, complete the self-assessment and for project specific controls, consult responsible officers for an appropriate equivalent action.

At the conclusion of the self-assessment, you will confirm the highest residual risk rating for each of the five Ethics Principles. This rating will determine whether you need to submit your assessment to the AI Review Committee, proceed without changes, make changes, or halt the project.

ⓘ The balance of benefits and risks

Some Elevated risk uses of AI (for example within Health), are undertaken to improve existing processes, or because of a clear benefit to community. It is important to consider the risk of not using AI if it's the best solution available.

Ensuring commitment to Human Rights

Identify any potential risk to human rights

- Is the use of the AI system likely to restrict human rights? If so, is any such restriction publicly justifiable?
- Were possible trade-offs between the different principles and rights ascertained, documented, and evaluated?
- Does the AI system suggest actions or decisions to make, or outline choices to human users?
- Could the AI system inadvertently impact human users' autonomy by influencing and obstructing their decision-making?
- Did you evaluate whether the AI system should inform users that its outputs, content, recommendations, or results arise from an algorithmic decision?

ⓘ **Do I need a Human Rights Impact Assessment (HRIA)?**

The NSW AI Ethics Policy confirms that AI will not be used to make unilateral decisions that impact our citizens or their human rights.

If the questions provided may result in human rights being at risk, we recommend you conduct a human rights impact assessment (HRIA).

Ensure AI use complies with legal protections for human rights. Human rights are legally recognised and protected through:

- Laws at the federal and state and territory levels

- The Australian constitution

- The common law

Applicable federal and state laws that protect human rights include:

Australian Human Rights Commission Act 1986 (Cth)
Age Discrimination Act 2004 (Cth)
Disability Discrimination Act 1992 (Cth)
Racial Discrimination Act 1975 (Cth)
Sex Discrimination Act 1984 (Cth)
Anti-Discrimination Act 1977 (NSW)

Publicly available resources:

Australian Human Rights Commission

Public Sector Guidance Sheets

Identifying responsible officers

Reviewing the potential risks associated with AI requires individuals who are appropriately skilled, and qualified for the role. In the framework, these roles are referred to as responsible officers. The self-assessment is to be completed by (or the result confirmed with) the Responsible Officers. The roles cover the different elements of project leadership and those responsible for technical performance and data governance.

These four roles have independent responsibilities and must not all be held by the same person*.

Project Sponsor name	Enter Project Sponsor name
Executive Business Sponsor name*	Enter Executive Business Sponsor name
Technical System owner name	Enter Technical System owner name
Data Governance owner name**	Enter Data Governance owner name

*The executive business sponsor may also be the project sponsor.

** The data governance role is equivalent to the "Accountable executive" and "responsible executive" defined within NSW data governance toolkit.

ⓘ Responsibility for managing AI

The responsibility for managing AI risks falls within the purview of the Agencies, who are obligated to ensure mitigation of the AI risk defined in the framework.

Understanding the risk levels

ⓘ Determine your risk levels

Given the absence of a standard for AI risk levels, use this foundation to build on. The self-assessment employs these concepts.

Consider the elevated risk factors (slide 10), reversibility, and your specific business use case when establishing your risk levels.

The examples provided may shift between risk levels based on this point.

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
Definition: AI systems or applications that have no, or extremely minimal risk associated with their use, or where the concept of risk is not applicable due to the nature of the AI system or its intended purpose.	Definition: AI systems or applications that, if they malfunction or produce unintended outcomes, can be easily reversed or corrected without causing any harm or damage.	Definition: AI systems or applications that, if they malfunction or produce unintended outcomes, can be reversed or corrected, but may cause moderate inconvenience, disruption, or harm.	Definition: AI systems or applications that, if they malfunction or produce unintended outcomes, can be reversed or corrected, but may cause significant financial losses, reputational damage, harm to the environment, individuals, or society.	Definition: AI systems or applications that, if they malfunction or produce unintended outcomes, may cause catastrophic, irreversible consequences for individuals, societies, or the environment.
Consequences: The potential consequences of an AI system in this category are either non-existent or so insignificant that they can be safely disregarded.	Consequences: The potential consequences of a low-risk AI system are minimal and do not cause any harm on individuals, organisations, or society.	Consequences: The potential consequences of a mid-range risk AI system are more noticeable and may have a temporary impact on individuals, organisations, or specific domains.	Consequences: The potential consequences of a high-risk AI system are substantial and may have a lasting impact on individuals, organisations, or entire industries.	Consequences: The potential consequences of a very high-risk AI system are severe and may have permanent and irreversible implications.

Understanding the risk levels (continued) - Examples

Note

Please use the following as a general guide only, examples can move between risk levels as they would be dependent on specific use cases.

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 • Noise suppression on audio calls • Image resolution enhancements • Grammer and spell checking • Text summarisation of non-sensitive content • Search functions in browsers • Analytics report	Examples • 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 • Customer service chatbots • Recommendation systems • Language translation tools • Content curation • Predictive maintenance • Natural language processing of gov documents.	Examples • 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 • Autonomous benefits eligibility without human oversight • Self-driving cars • Predictive reoffending • Medical diagnosis without oversight • Autonomous AI systems on critical infrastructure (i.e. energy)

3 Self-assessment & risk summary

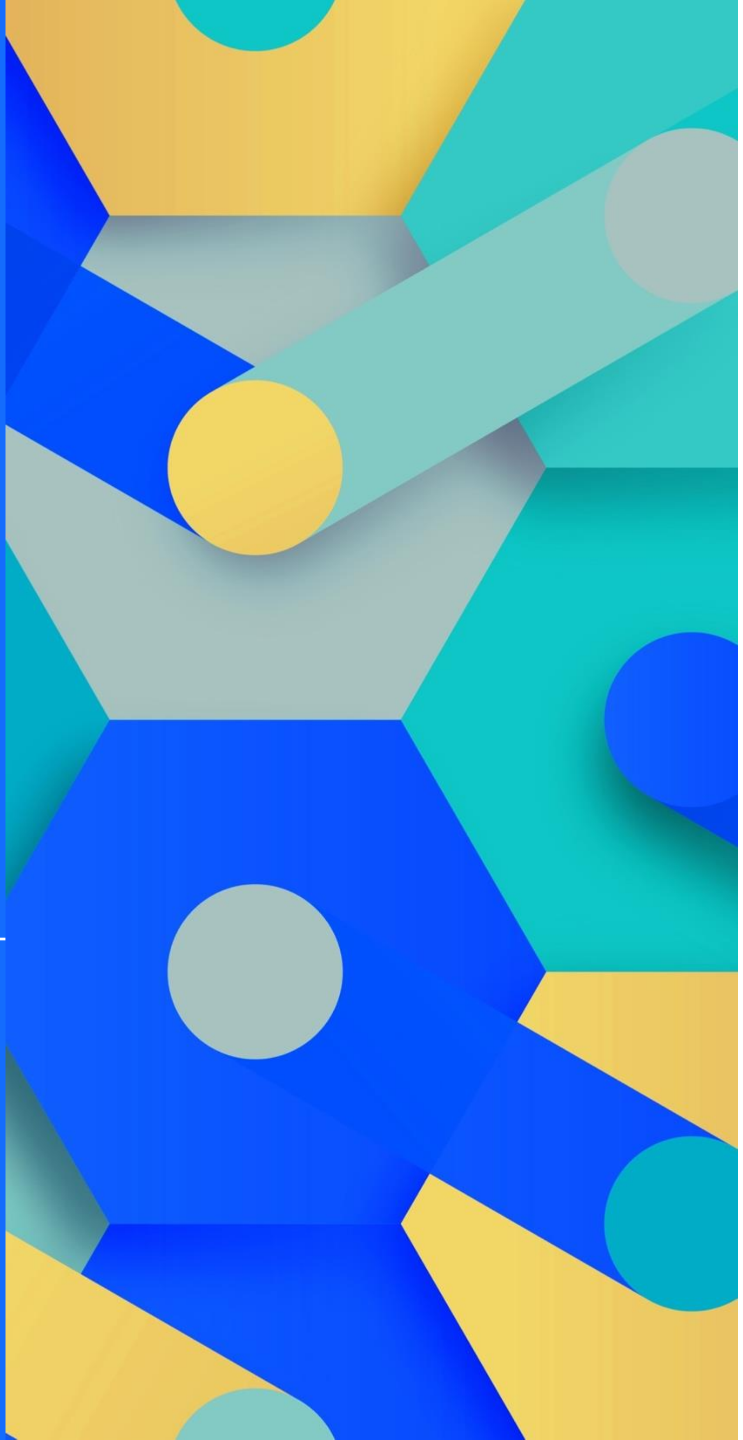
Complete the self-assessment and review the summary.

1 Scope of applying the framework

2 Self-assessment readiness

3 Self-assessment & risk summary

4 Self-assessment mitigation & next steps



Project/System information

Collaborate with team members to complete the self-assessment. The time this takes will vary based on the complexity of your AI system. Take your time to record responses in this document. Add pages if needed. Reach out to agency AI experts or NSW AI Review Committee secretariat for assistance at AISecretariat@customerservice.nsw.gov.au.

Project name / System name	**AI Universal Translator for Inclusive Government**
How is/was the system delivered?	Develop AI or co-train
What is the phase of the system?	Design and develop
Have you defined the responsible officers in the pre-assessment checklist?	No
List contributors to self-assessment, excluding responsible officers, with names and roles	Vinod Ralh
What is the next date/milestone that will trigger the next review?	N/A
System description	The universal translator is a tool designed to improve communication between diverse populations and local governments, focusing on inclusivity, particularly for individuals with disabilities. Its core features include: • Multilingual communication capabilities. • Simplification of complex government documents into plain language. • Accessibility features like text-to-speech and screen-reader compatibility. • Support for users with limited digital literacy or accessibility needs through personalized interfaces and audio guides. This tool aims to eliminate communication barriers, ensuring everyone can access essential information and services equitably .

Principle 1:
Community Benefit

Principle Statement

AI must prioritise community outcomes, ensuring alignment with laws, minimising harm, and maximising benefit.

Description

Government must prioritise the well-being and interests of the community. AI must be the most appropriate solution for a service delivery or policy problem, aligning with government priorities, complying with laws and regulations, and balancing risk against community value. Careful evaluation should assess benefit against potential harm to individuals, communities, and the environment, along with the degree of reversibility and impact

Note the principles statements and descriptions may offer more detail than the current AI ethics policy if required to describe the detailed framework content.

Section Instructions

General Benefits & Strategic alignment: This section is optional but encouraged if you haven't completed a project/system benefits analysis.

General Risk Factors: The initial high level risk assessment for your solution.

Questions with specific controls: Identifies questions where specific action may be required based on your response. Some controls may require pausing for essential information. Complete the self-assessment and implement necessary mitigations afterward.

Harms: Evaluate the risk of AI potential harms. If integrating into your risk framework, use these harms as the potential consequences.

General benefits assessment (optional)

Community benefit

What is your confidence level in achieving the following benefits?

This section is optional but encouraged if you haven't completed a benefits analysis. Think about the *potential* benefits and the likelihood of these benefits being *realised* in practice; as well the strength of available *evidence* supporting your assessment.

Consider the benefits associated with the AI project/system …	Very low or N/A	Low	Mid-range	High	Very high
Delivering a better-quality existing service or outcome (for example, accuracy or client satisfaction)					●
Reducing processing or delivery times				●	
Generating financial efficiencies or savings		●			
Providing an AI capability that could be used or adapted by other agencies					●
Delivering a *new* service or outcome (particularly if it cannot be done without using AI)				●	
Enabling *future* innovations to existing services, or new services or outcomes				●	
Comments					

← Move blue dots to your selection.

Strategic alignment (optional)

Questions with specific controls

Community benefit

This question is optional but encouraged if you haven't completed a benefits analysis

1. Will the AI system improve on existing approaches to deliver the outcomes described in: • **the Human Services Outcomes Framework** • **the Smart Places Outcomes Framework** • **NSW Treasury Budget Outcomes** • **Your Agency strategic Plans or** • **another relevant NSW Outcomes Framework?**	**Yes:** List which strategic outcomes the systems improve.	***Partially:** After your pilot, you must conduct a formal benefits review before scaling.	***Not sure:** Pause the project and prepare a Benefits Realisation Management Plan.	***No:** Do not proceed any further. Discuss this with the Responsible Officers.
	●			
Response – You must explain your answer.	"We're committed to increasing cultural participation, and helping businesses to innovate, adapt and grow." – **City of Sydney** "Australia is a prosperous, safe and united country. Our inclusive national identity is built around our shared values including democracy, freedom, equal opportunity and individual responsibility." – **Department of Home Affairs**			

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

ⓘ **Benefits:** All AI projects should have a benefits register that is kept up to date throughout the project. The benefits register should be maintained by the Responsible Officers.

General risk factor assessment

Community benefit

What is your initial risk assessment of this solution?

Consider the likelihood and potential consequences (harms) if the following risk were to eventuate. Factors to consider include familiarity of the solution and use case, reversibility, data sensitivity, and level of human intervention/oversight.

Consider the risks associated with …	Very low or N/A	Low risk	Mid-range risk	High risk	Very high risk
Whether this AI system is delivering a new or existing service		●			
The potential to cause discrimination from unintended bias			●		
Whether the AI system is a single point of failure for your service or policy		●			
If there is sufficient experienced human oversight of the AI system			●		
Over-reliance on the AI system or ignoring the system due to High rates of false alert			●		
Whether the linkage between operating the AI system and the policy outcome is clear	●				
The system's explainability and transparency regarding generated content and decisions			●		
Comments	Improving upon existing system information, which will continue to be provided in parallel, during a trial period.				

← Move blue dots to your selection.

Non-AI system consideration

Questions with specific controls

Community benefit

2. Were other, non-AI systems considered?	**Yes:**	***Informally:** After your pilot, you must conduct a formal benefits review before scaling.	***No:** Do not proceed any further. Review with the responsible officers on how to resolve.	**N/A**
		●		
Response – You must explain your answer.	Hackathon. Looking for innovative ideas.			

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Alternatives: For an AI system to be viable, AI must be the most appropriate system for your service delivery or policy problem. AI systems can come with more risk and cost than traditional tools. You should use an AI system when it the best system to maximise the benefit for the customer and for government.

Legal framework alignment

Questions with specific controls

Community benefit

3. Does this system and the use of data align with relevant legislation? **You must make sure your data use aligns with:** • Privacy and Personal Information Protection Act 1997 (NSW) (PPIPA) • NSW Anti-Discrimination Act 1977 • Government Information (Public Access) Act 2009 • State Records Act 1998 **Other relevant NSW or Commonwealth Acts including:** • **Public Interest Directions made under PIPPA (exemptions)** • **Health Records and Information Privacy Act 2002 (NSW) (HRIPA)** • **Health Public Interest Directions made under HRIPA (exemptions)** • **Public Health Act 2010** • **Relevant Acts for your Agency such as the Transport Administration Act 1988 (NSW) or the Police Act 1990 (NSW)**	**Yes:** If you have confirmed any other relevant acts, please list these in your response.	***Unclear:** Pause the project. Seek advice from an appropriate NSW legal source or the NSW Privacy Commissioner. You may need to redesign your project and or system.	***No:** Do not proceed any further unless you receive clear legal advice that allows you to proceed. Consider redesigning your project and or system.
			●
Response – You must explain your answer	As a hackathon we have not looked into relevant legislation.		

*If your solution is operational consult responsible officers for an appropriate equivalent action.

Move blue dot to your selection.

More information: You must always comply with privacy and information access laws, including when you are developing and using AI Systems.

Potential harms – negative consequences

Community benefit

This prompts early consideration to potential harms, revise after completing the self-assessment.

Think about how likely and serious a risk could be if it happens. Factors to think about include how familiar the solution and use case are, whether it can be reversed, how sensitive the data is, and how much human oversight is needed.

Consider the risks of ...	Very low or N/A	Low risk	Mid-range risk	High risk	Very high risk
Physical harms		●			
Psychological harms		●			
Environmental harms or harms to the broader community		●			
Unauthorised use of health or sensitive personal information (SIP)	●				
Impact on right, privilege or entitlement		●			
Unintended identification or misidentification of an individual		●			
Misapplication of a fine or penalty		●			
Other financial or commercial impact		●			

← Move blue dots to your selection.

Continued over page →

Potential harms – negative consequences

Community benefit

This prompts early consideration to potential harms, revise after completing the self-assessment.

Think about how likely and serious a risk could be if it happens. Factors to think about include how familiar the solution and use case are, whether it can be reversed, how sensitive the data is, and how much human oversight is needed.

Consider the risks of ...	Very low or N/A	Low risk	Mid-range risk	High risk	Very high risk
Incorrect advice or guidance			●		
Inconvenience or delay		●			
Erosion of trust		●			
Ethical implications			●		
Economic disruption / impact.		●			
Social equality		●			
Other harms		●			
Comments, ensure to include details of other harms if you selected this option.	The use of a verification agent and human intervention helps mitigate some of the risk. Data on more sensitive data classifications and more vulnerable user population could go through human in the loop workflows. A trial and finetuning could also help.				

← Move blue dots to your selection.

Reversible harms

Questions with specific controls

Community benefit

4. Could the AI system cause harms that are reversible?	**No**	***Yes: and mid-range or higher risk**, do not proceed until you receive legal advice. If you have legal approval: discuss this with all relevant stakeholders, you may need ethics approval, consider a Human Rights Impact Assessment.	**Yes: and low or very low risk**, explain below.	***Unclear:** Pause the project and review with the responsible officers on how to resolve.
			●	
Response – You must explain your answer	It will dependend on queries asked. Our solution put a human in the middle if there’s vulnerable groups and/or sensitive information. Also, there would be eval following a trial. To begin with we’re talking about style guide so very low risk. But citizenship website, and misinterpretation, could be more of an issue.			

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Reversible Vs Irreversible harms: Irreversible harm refers to a situation where it's impossible to revert to a previous condition before the harm occurred. For example, if an AI system makes an incorrect decision to deny somebody a pension without an option to have that overturned. You should ensure the ability to overturn outcomes if harm is caused or if the AI system makes incorrect decisions.

Significant harms

Questions with specific controls

Community benefit

5. Could the AI system cause significant or irreversible harms? Example: Autonomous AI systems on critical infrastructure (i.e. energy) For more information on when a Human Rights Impact Assessment is required. see https://humanrights.gov.au/	**No**	**Yes, but it's better than existing systems:** You must seek approval from an ethics committee. You must have clear legal advice that allows you to proceed. Consult with all relevant stakeholders. Consider a Human Rights Impact Assessment.	***Yes:** Do not proceed until you receive clear legal advice. If you have legal approval: discuss this with all relevant stakeholders, seek approval from an ethics committee, consider a Human Rights Impact Assessment.	***Unclear:** Pause the project and review with the responsible officers on how to resolve.
		●		
Response – You must explain your answer	Misinterpretation of policy or citizenship ship may cause issues. There is a human in the middle. Citing site of origination site may also help reduce risk.			

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Monitoring for possible harms: You must monitor your AI system closely for harms that it may cause. This includes monitoring outputs and testing results to ensure there are no unintended consequences. You should be able to quantify unintended consequences, secondary harms or benefits, and long-term impacts to the community, even during testing and pilot phases. Testing can still lead to harm if the system is making consequential decisions. You must consider and account for this possibility even if human testers are willing volunteers. Changing the context or environment in which the AI system is used can lead to unintended consequences. Planned changes in how the AI is used should be carefully considered and monitoring undertaken.

Possible secondary or cumulative harms

Questions with specific controls

Community benefit

6. Could the AI System result in secondary harms, or result in a cumulative harm from repeated application of the AI System? Example of a cumulative harm is a video system initially collecting and analysing data for security purposes, but over time, as more data is gathered and analysed, individual privacy could be at risk.	No	*Yes: and **mid-range or higher risk**, do not proceed until you receive legal advice. If you have legal approval: discuss this with all relevant stakeholders, you may need ethics approval, consider a Human Rights Impact Assessment.	Yes: and **low or very low risk**, explain below.	*Unclear: Pause the project and review with the responsible officers on how to resolve.
			●	
Response – You must explain your answer	Bias in training sets may result in unclear communication.			

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Secondary harms: Sometimes harms are felt by people who are not direct recipients of the product of service. We refer to these as secondary harms. Secondary harms include things like a loss of trust. You need to think deeply about everyone who might be impacted, well beyond the obvious end user.

Principle 2: Fairness

Principle Statement

Use of AI will be fair, ensuring not to perpetuate bias and inequality by leveraging diverse representative datasets, monitoring performance, and using rigorous data governance.

Description

The Fairness principle emphasises equitable AI, where decisions made by or with the assistance of AI do not perpetuate bias or inequality. It demands rigorous evaluation and management of data quality, advocating for diverse and representative datasets. AI systems must be designed to avoid unfairness, with strategies to detect and correct biases, ensuring fairness for all segments of society.

Note the principles statements and descriptions may offer more detail than the current AI ethics policy if required to describe the detailed framework content.

Section Instructions

Risk Factors: Evaluate likelihood and potential harm level for each risk factor and document the overall risk rating. If integrating into your own risk framework, consider these as risk events and the consequences being the harms listed under community benefits.

Questions with specific controls: Identifies questions where specific action may be required based on your response. Some controls may require pausing for essential information. Complete the self-assessment and implement necessary mitigations afterward.

Risk factors and ratings

Fairness

Fairness

Evaluate likelihood and potential harm level for each risk factor and document the overall risk rating.

Consider the risks associated with ...	Very low or N/A	Low risk	Mid-range risk	High risk	Very high risk
Using incomplete or inaccurate data		●			
Having poorly defined descriptions and indicators of "Fairness"		●			
Not ensuring ongoing monitoring of "Fairness indicators"		●			
Decisions to exclude outlier data	●				
Informal or inconsistent data cleansing and repair protocols and processes		●			
Using informal bias detection methods (best practice includes automated testing)		●			
The likelihood that re-running scenarios could produce different results (reproducibility)		●			
Inadvertently creating new associations when linking data and/or metadata	●				

← Move blue dots to your selection.

Continued over page →

Risk factors and ratings

Fairness

Fairness

Evaluate likelihood and potential harm level for each risk factor and document the overall risk rating.

Consider the risks associated with ...	Very low or N/A	Low risk	Mid-range risk	High risk	Very high risk
Differences in the data used for training compared to the data for intended use		●			
Comments	Using public, infrequently chaing website unstructured data.				

← Move blue dots to your selection.

Fairness

Questions with specific controls

Fairness

7. Can you explain why you selected the data you're using in your system?	**Yes**	**Unclear:** Consult with relevant stakeholders on data options or implement a data improvement strategy or redesign your project/system.	***No, but it's better than existing systems:** Document your reasons. Clearly demonstrate that you have consulted with all relevant stakeholders before proceeding.	***No:** Pause the project and review with the responsible officers on how to resolve.
	●			
Response – You must explain your answer	Selected based on challenges, data sets mandated and discussions with experienced mentors.			

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Data relevance and permission: Your AI system may draw on multiple datasets from different sources to find new patterns and insights. You need to determine if you can and should use the data for the AI system. This can be challenging for historical data that may have been collected for a different purpose. For a detailed considerations of Data Sharing and Use Controls see Appendix 4.

Fairness

Questions with specific controls

8. Is the data that you need for your system available and of appropriate quality given the potential harms identified? If your system is a data creation or data cleansing application, answer according to the availability of any existing data that is needed for the solution to succeed, for example, training datasets.	**Yes**	**Unclear:** Consult with relevant stakeholders to identify alternative data sources or implement a data improvement strategy or redesign your project/system.	***Partially, it's better than existing systems:** Document your reasons and details to demonstrate that you have consulted with all relevant stakeholders before proceeding.	***No:** Pause the project and discuss with responsible officers on how to resolve.
			●	
Response – You must explain your answer	There are inconsistent approaches to data taken in the describing "good government writing style". For example, some use			

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Data quality: Data quality is often described in terms of minimum requirements for accuracy, timeliness, completeness, and consistency. There are examples of data quality standards for AI in the appendices. Your AI system may be significantly impacted by poor quality data. It is important to understand how significant the impact is before relying on insights or decisions generated by the AI system. Absence of data may lead to unintended biases impacting insights generated by the AI system. Unbalanced data is a common problem when training AI systems (the situation where the distribution of classes or categories in the training dataset is not representative of the real-world scenario).

Fairness

Questions with specific controls

Fairness

9. Does your data reflect the population that will be impacted by your system?	**Yes**	***Partially, it's better than existing systems:** Consider seeking advice from an ethics committee. Document below how you have consulted with all relevant stakeholders before proceeding. Consider a Human Rights Impact Assessment.	***No or unclear:** Pause the project and review with the responsible officers on how to resolve.	**N/A:** Document your reasons as to why this does not apply, then go to next question.
	●			
Response – You must explain your answer	Public data sets are taken from sites used by public.			

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Fairness

Questions with specific controls

Fairness

10. Have you considered how your AI system will address issues of diversity and inclusion (including geographic diversity)? **11. Have you considered the impact with regard to gender and on minority groups including how the system might impact different individuals in minority groups when developing this AI system?** Minority groups may include: • those with a disability • LGBTQIA+ and gender fluid communities • people from CALD backgrounds • Aboriginal and Torres Strait Islanders • children and young people • People from varying socio-economic backgrounds	**Yes**	***Partially, it's better than existing systems:** Consider seeking advice from an ethics committee. Document below how you have consulted with all relevant stakeholders before proceeding. Consider a Human Rights Impact Assessment	***No or unclear:** Pause the project and discuss with responsible officers how to resolve.	**N/A:** Document your reasons as to why this does not apply, then go to next question.
	●			
Response – You must explain your answer	This solution has a focus to be inclusive of groups that are often missed. In the prototype we demonstrate Text to Speech for those who rely on voice rather than written work, text simplification for those who need a younger reading level, text translation to different languages and "plain language guide". These all help with inclusion and reaching diverse groups needs.			

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Diversity and inclusion, and the impact on minorities: AI often overlooks minority nuances, leading to biased outcomes. Considering cultural sensitivities and underrepresentation, it's vital to test AI outputs for fairness across all demographics, ensuring accurate representation and unbiased decisions. Think deeply about everyone who may be impacted.

Fairness

Questions with specific controls

12. Do you have appropriate performance measures and targets (including fairness ones) for your AI system, given the potential harms? Aspects of accuracy and precision are readily quantifiable for most systems which predict or classify outcomes. This performance can be absolute, or relative to existing systems. How would you characterise "Fairness" such as equity, respect, justice, in outcomes from an AI system? Which of these relate to, or are impacted by the use of AI?	**Yes**	***No or unclear:** For **elevated risk** uses of AI, pause the project until you have established performance measures and targets. For **non-elevated risk** projects or systems, results should be treated as indicative and not relied on. Document your reasons below.	**N/A:** Document your reasons as to why this does not apply, then go to next question.
	●		
Response – You must explain your answer	In the prototype, we intend to have the AI orchestrator generate AI principle measurements which go to the governance dashboard. If time permits we will have our agent verifier provide addition metrics on accuracy.		

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Measuring AI system performance: At the scoping stage, you will need to make important choices about what you measure. You should measure:

Accuracy: how close an answer is to the correct value

Precision: how specific or detailed an answer is

Sensitivity: the measure of how many actually positive results are correctly identified as such

Specificity: the measure of how many actually negative results are correctly identified by the AI system

Fairness objectives: whether the system is meeting the fairness objectives defined for the system (which could include for example that there aren't more prediction errors on some cohorts than others)

Fairness

Questions with specific controls

Fairness

13. Do you have a way to monitor and calibrate the performance (including fairness) of your AI system? Operational uses of AI which are continuously updated / trained can quickly move outside of performance thresholds. Supervisory systems can monitor system performance and alert when calibration is needed.	**Yes**	***No or unclear:** For **elevated risk** uses of AI, pause the project until you have established performance monitoring. For **non-elevated risk** projects or systems, results should be treated as indicative and not relied on. Document your reasons below.	**N/A:** Document your reasons as to why this does not apply, then go to next question.
	●		
Response – You must explain your answer	In our prototype, if time permits, we will look at training, finetuning and evaluating accuracy of our model based on generated data sets.		

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Measuring AI system performance: Elevated risk uses of AI should have clear performance monitoring and calibration schedules.

For Elevated risk uses of AI which are continuously training and adapting with moderate residual risks, **weekly** performance monitoring and calibration is recommended. For low risk, **monthly** evaluation and calibration is recommended.

For operational systems with High risk or Very High risk, a custom evaluation and calibration will be required.

Principle 3:
Privacy and security

Principle Statement

Ensure secure, transparent, and compliant data use to preserve public trust.

Description

NSW citizens need assurance of safe, secure and privacy-compliant data use. Transparent review mechanisms and community engagement are essential. Explicit consent, cybersecurity compliance, and privacy legislation adherence are vital. Any project outcome will be undermined if there is a risk of data breaches or compromised personal data, eroding public trust.

Note the principles statements and descriptions may offer more detail than the current AI ethics policy if required to describe the detailed framework content.

Section Instructions

Risk Factors: Evaluate likelihood and potential harm level for each risk factor and document the overall risk rating. If integrating into your own risk framework, consider these as risk events and the consequences being the harms listed under community benefits

Questions with specific controls: Identifies questions where specific action may be required based on your response. Some controls may require pausing for essential information. Complete the self-assessment and implement necessary mitigations afterward.

Risk factors and ratings

Privacy and security

It is critical to assess potential use of sensitive data. When the size of an identifiable cohort within the model training dataset is smaller, the likelihood of identification or re-identification increases, hence the higher risk.

Sensitive data including information on:	Identifiable cohort >50 or N/A	Identifiable cohort >20 and <50	Identifiable cohort >10 and <20	Identifiable cohort >5 and <10	Identifiable cohort <5
	Very low or N/A	Low risk	Mid-range risk	High risk	Very high risk
Children	●				
Religious individuals	●				
Racially or ethnically diverse individuals	●				
Individuals with political opinions or associations	●				
Individuals with trade union memberships or associations	●				
Gender and/or sexually diverse individuals	●				
Individuals with a criminal record	●				

← Move blue dots to your selection.

Continued over page →

Risk factors and ratings

Privacy and security

It is critical to assess potential use of sensitive data. When the size of an identifiable cohort within the model training dataset is smaller, the likelihood of identification or re-identification increases, hence the higher risk.

Sensitive data including information on:	Identifiable cohort >50 or N/A	Identifiable cohort >20 and <50	Identifiable cohort >10 and <20	Identifiable cohort >5 and <10	Identifiable cohort <5
	Very low or N/A	Low risk	Mid-range risk	High risk	Very high risk
Specific health or genetic information	●				
Personal biometric information	●				
Other sensitive person-centred data	●				
Comments	In the prototype this isn't an issue. In a real world scenario we could ask users to create a profile and provide optional data that can help personalise responses (including some PII or sensitive data), but this would be driven by the citizen. Many of the features could work anonymously regardless of approach.				

← Move blue dots to your selection.

Privacy and security

Questions with specific controls

14. Have you applied the "Privacy by Design" and "Security by Design" principles in your system?	**Yes:** Document any points to resolve, then go to next question. Consider contacting the Information and privacy commissioner or Cyber NSW for any points not resolved.	***Partially:** Pause the project, apply the principles before proceeding, document any points to resolve below then go to next question. Consider contacting the Information and privacy commissioner or Cyber NSW for any points not resolved.	***No or unclear:** Pause the project, apply the principles before proceeding, document any points to resolve below then go to next question. Consider contacting the Information and privacy commissioner or Cyber NSW for any points not resolved.
			●
Response – You must explain your answer	Not for the protype, in this would be required for a pilot or production implementation,		

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Privacy by design, security by design: Even small AI projects or systems may have privacy or security vulnerabilities. For example, an analytics system which stores commercially sensitive data in a non-secure environment unbeknown to the user.

The NSW Information Privacy Commissioner has prepared 7 Privacy by Design principles. These principles should be applied to your AI project and system.

If you are unsure how to apply these principles, you seek help from the Information and Privacy Commission.

NSW Government has also developed Security Principles which should also be applied to all digital projects and systems.

Privacy and security

Questions with specific controls

15. Have you completed a privacy impact assessment (either third party or self-assessed)?	**Yes:** Document the result below, then go to next question.	***No:** Pause the project until you have completed a privacy impact assessment.	**N/A:** Your system doesn't use or generate any sensitive information, confirmed with responsible officers, document below this confirmation.
			●
Response – You must explain your answer	Prototype based on public data sets.		

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Privacy impact assessment: Even systems not focussed on person-centred data may reveal information about a person, their relationships or preferences. For example, analysis of environmental or spatial data may reveal information about a land-holder's interaction with the local environment.

A Privacy Impact Assessment (PIA) can help you to identify and minimise privacy risks. A PIA can help you implement 'privacy by design' and demonstrate compliance with privacy laws.

The Information Privacy Commission has more information and templates.

Privacy and security

Questions with specific controls

16. If you are using information about individuals who are reasonably identifiable, have you sought consent from citizens about using their data for this particular purpose? See the NSW Privacy and Personal Information Protection Act (1998) for a definition of Personal Information. See the NSW Privacy Commissioner's fact sheet on Reasonably Ascertainable Identity.	**Yes**	***Authorised use:** For AI systems intended to operate under legislation which allows use of Identifiable Information, do not proceed unless you receive clear legal / independent privacy advice that allows you to proceed. The system should always be monitored for harms.	***Partially:** Pause the project until you have obtained consent or clear legal advice authorising use of this information	***No:** Pause the project until you have either consent or clear legal advice authorising use of this information.	**N/A:** Document your reasons below as to why this does not apply.
					●
Response – You must explain your answer	Not for the prototype.				

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Exceptions: You can ask the Privacy Commissioner to make a Public Interest Direction (PID) to waive the requirement to comply with an Information Protection Principle. These are only granted in circumstances where there are compelling public interests. For AI systems intended to operate under legislation which allows use Personally Identifiable Information, the public benefits must be clear before proceeding to pilot phase.

Governing use of Personally Identifiable Information: You must apply higher governance standards if you are managing Personally Identifiable Information. Refer to Appendix D. for examples of data sharing frameworks and controls.

Privacy and security

Questions with specific controls

17. Does your system adhere to the mandatory requirements in the NSW Cyber Security Policy? **Have you considered end-to-end Security Principles for your system?**	**Yes:** Provide information below that confirms you have done this and any key information to note for ongoing risk management.	*No or partially: Pause the project until you meet mandatory requirements.
		●
Response – You must explain your answer	Not for the prototype.	

*If your solution is operational consult responsible officers for an appropriate equivalent action.

Move blue dot to your selection.

Cyber security: AI can pose new cyber security risks, be vigilant.

You must comply with the mandatory requirements in the NSW Cyber Security Policy.

The NSW Government Chief Cyber Security Officer (CCSO) has responsibility for leading a coordinated government response to cyber security failures

Privacy and security

Questions with specific controls

Privacy and security

18. Does your dataset include using sensitive data subjects as described by section 19 of the NSW Privacy and Personal Information Protection Act 1998? If use of sensitive data is a must, ensure to leverage privacy enhancing technology such as use of synthetic data, data anonymisation and deidentification, encryption, secure aggregation and random noise generation.	**No:** Document how you have confirmed this.	**Yes:** Seek advice from an appropriate NSW legal source or the NSW Privacy Commissioner. Consider seeking approval from an ethics committee.	***Unclear:** Pause the project and review your data. Consider advice from an appropriate NSW legal source or the NSW Privacy Commissioner.
	●		
Response – You must explain your answer	Not for the prototype.		

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Sensitive data: The NSW Government Information Classification, Labelling and Handling Guidelines have been developed to help agencies correctly assess the sensitivity or security of information, so that the information can be labelled, used, handled, stored and disposed of correctly.

Governing Use of Sensitive Information: You must apply higher governance standards if you are managing Sensitive Information. Refer to Appendix D. For examples of data sharing frameworks and controls.

Principle 4: Transparency

Principle Statement

The use of AI will be transparent to the people it could impact, providing review mechanisms that allow concerns to be raised and addressed, privacy preserving, cyber secure and ethical.

Description

Transparency fosters public trust and accountability by ensuring community consultation, awareness of AI use, and the ability for individuals to seek explanations and challenge decisions that impact them directly, unless there is an overriding public interest in not doing so. The development of AI systems must be compliant with relevant legislation, cyber security policies and ethical.

Note the principles statements and descriptions may offer more detail than the current AI ethics policy if required to describe the detailed framework content.

Section Instructions

Risk Factors: Evaluate likelihood and potential harm level for each risk factor and document the overall risk rating. If integrating into your own risk framework, consider these as risk events and the consequences being the harms listed under community benefits

Questions with specific controls: Identifies questions where specific action may be required based on your response. Some controls may require pausing for essential information. Complete the self-assessment and implement necessary mitigations afterward.

Risk factors and ratings

Transparency

Evaluate likelihood and potential harm level for each risk factor and document the overall risk rating

Is a 'black box' AI system, like large language models, automatically High risk? Commercial AI systems' inner workings are often inaccessible and complex to interpret. Transparency risks exist when sourcing “black box” system components. Proactively consider human judgement in using 'unexplainable' insights or decisions.

Consider the risks associated with …	Very low or N/A	Low risk	Mid-range risk	High risk	Very high risk
Incomplete documentation of AI system design, or implementation, or operation		●			
No or limited access to model’s internal workings or source code (“Black Box”)		●			
Being unable to explain the output of a complex model		●			
A member of the public being unaware that they are interacting with an AI system	●				
No or low ability to incorporate user feedback into an AI system or model		●			

Move blue dots to your selection.

Continued over page →

Risk factors and ratings

Transparency

Transparency

Evaluate likelihood and potential harm level for each risk factor and document the overall risk rating

Is a 'black box' AI system, like large language models, automatically High risk? Commercial AI systems' inner workings are often inaccessible and complex to interpret. Transparency risks exist when sourcing "black box" system components. Proactively consider human judgment in using 'unexplainable' insights or decisions.

Consider the risks associated with …	Very low or N/A	Low risk	Mid-range risk	High risk	Very high risk
The inability to audit past decisions, where input from AI systems was used.		●			
Comments	For the prototype this is seen as low risk. Its likely that for a production implementation, a commercial LLM such as OpenAI or other would be used. These generally provide very little information as to their reasoning.				

← Move blue dots to your selection.

Transparency

Questions with specific controls

19. Have you consulted with the relevant community that will benefit from (or be impacted by) the system?	**Yes**	***Authorised use:** For AI systems intended to operate under legislation which allows use without community consultation, do not proceed unless you receive clear legal advice that allows you to proceed. The system should always be monitored for harms.	***Partially, it's better than existing systems:** Consider seeking advice from an ethics committee. Document below how you have consulted with all relevant stakeholders before proceeding.	***No:** Pause the project, develop a Community Engagement Plan and consult with the relevant community.	**N/A:** Document your reasons as to why this does not apply, then go to next question.
	●				
Response – You must explain your answer	As we are citizens of NSW we are potential users. We have consulted with stakeholders in Deptarment of Home Affairs.				

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Consultation: You must consult with the relevant community when you design your system. This is particularly important for Elevated risk uses of AI. Communities have the right to influence government decision-making where those decisions, and the data on which they are based, will have an impact on them. For AI intended to operate under legislation which allows use without community consultation, the public benefits must be clear before proceeding.

Transparency

Questions with specific controls

20. Are the scope and goals of the project publicly available, and have you communicated how safeguards have been put in place to mitigate any potential harms? Explore diverse approaches to instil confidence within communities regarding your AI utilisation. This may entail targeted communication strategies or maintaining public registers. Offer concise and straightforward explanations of your AI usage to those potentially affected, especially for elevated risk. Ensure these explanations foster trust without generating confusion.	Yes	*No: Make sure you communicate to relevant stakeholders and the community who are impacted before proceeding.	N/A: Document your reasons as to why this does not apply, then go to next question.
	●		
Response – You must explain your answer	The GovHack material provides a project overview that covers goals, scope and how governance layer provides some transparency.		

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Sharing project goals: The NSW AI Strategy recognises we have important work to do to encourage public trust in AI, by ensuring Government is transparent and accountable, and that AI delivers positive outcomes to citizens.

Transparency

Questions with specific controls

Transparency

21. Is there an easy and cost-effective way for people to appeal a decision that has been informed by your system? Individuals have the right to raise concerns or appeal decisions. Ensure the use of simple and easily understandable language to facilitate this process.	**Yes**	***No:** Pause your project, consult with relevant stakeholders and establish an appeals process.	**N/A:** Document your reasons as to why this does not apply, then go to next question.
	●		
Response – You must explain your answer	The intention is to build a capability that will allow users to comment on query results (thumbs up, down & comment)		

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

ⓘ **Right to appeal:** No person should ever lose a right, privilege or entitlement without right of appeal. A basic requirement of Transparency is for an individual affected by a relevant decision to understand the basis of the decision, and to be able to effectively challenge it on the merits and/or if the decision was unlawful. When planning your project/system, you must make sure no person could lose a right, privilege or entitlement without access to a review process or an effective way to challenge an AI generated or informed deci sion.

Transparency

Questions with specific controls

22. Does the AI system allow for transparent explanation of the factors leading to a decision or insight?	Yes	No, but a person makes the final decision: Consult with relevant stakeholders and establish a process to readily reverse any decision or action made by the AI system. Actively monitor for potential harms.	*No: Pause your project, consult with relevant stakeholders and establish a process to readily reverse any decision or action made by the AI system.	N/A: Document your reasons as to why this does not apply, then go to next question.
				●
Response – You must explain your answer	This is not designed into the prototype.			

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Clear explanations: As far as possible, you must have a way to clearly explain how a decision or outcome has been informed by AI. If the system is a "black box" due to lack of access to the inner workings or is too complex to reasonably explain the factors leading to the insight generation, it is essential to consider the role of human judgement in intervening before an AI generated insight is acted on. It is important to formalise and document this human oversight process. In low (or very low) risk environments, it may be sufficient to identify and document mechanisms to readily reverse any action arising from such an insight (for example, a person overriding an automated barrier).

Principle 5:
Accountability

Principle Statement

Decision-making remains the responsibility of organisations and individuals.

Description

Despite AI's autonomy, humans hold ultimate decision responsibility necessitating skilled operators with clear accountabilities. Establishing responsible parties across data, technology, models, and outcomes is crucial. Operators evaluate data inputs and outputs, understanding system limitations and model performance. The ability to identify and reverse AI decisions to prevent harm and apply human oversight to prevent over-reliance ensuring continuous review.

Note the principles statements and descriptions may offer more detail than the current AI ethics policy if required to describe the detailed framework content.

Section Instructions

Risk Factors: Evaluate likelihood and potential harm level for each risk factor and document the overall risk rating. If integrating into your own risk framework, consider these as risk events and the consequences being the harms listed under community benefits.

Questions with specific controls: Identifies questions where specific action may be required based on your response. Some controls may require pausing for essential information. Complete the self-assessment and implement necessary mitigations afterward.

Risk factors and ratings

Accountability

Evaluate likelihood and potential harm level for each risk factor and document the overall risk rating.

The skill and training for AI system operators is crucial. Automated systems pose the risk of over-reliance. Operators, including those exercising judgement over insights or alerts, must be well-trained. This includes the ability to critically evaluate insights and understand system limitations. Users must have confidence in their ability to identify, report, and resolve ethical concerns arising from AI-generated insights or decisions, or empower Responsible Officers to make decisions. Ensure consideration is given to training public servants delivering customer-facing services on how respond to inquiries from customers when AI is utilised, including guidance on who to direct such inquiries to.

Consider the risks associated with …	Very low or N/A	Low risk	Mid-range risk	High risk	Very high risk
Insufficient training of AI system operators		●			
Insufficient awareness of system limitations of Responsible Officers		●			
No or low documentation of performance targets or “Fairness” principles trade-offs		●			
No or limited mechanisms to record insight / AI System decision history		●			
The inability of third parties to accurately audit AI system insights / decisions		●			
Comments	This is a prototype built as-is by volunteers under creative commons licensing only.				

← Move blue dots to your selection.

Accountability

Questions with specific controls

23. Have you established who is responsible for: • use of the AI outputs, insights and decisions? • policy/outcomes associated with the AI system? • monitoring the performance of the AI system? • data governance? • technical solution governance? • appeal and redress processes?	**Yes:** Document who is responsible to each point within the question below.	*No or unclear: Pause the project while you identify who is responsible and make sure they are aware and capable of undertaking their responsibilities.
		●
Response – You must explain your answer	This has not been considered as part pf the prototype. However additional technical documentation provided does look at some of these aspects.	

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Responsible officers: This assessment is to be completed by, or the result confirmed with, the Responsible Officers. The Responsible Officer should be appropriately senior, skilled and qualified for the role.

Accountability

Questions with specific controls

Accountability

24. Have you established a clear processes to: • intervene if a relevant stakeholder finds concerns with insights, decisions or content generated (appeal and redress)? • ensure you do not get overconfident or over reliant on the AI system?	**Yes:** Document the details below.	***No:** Pause your project, consult with relevant stakeholders and establish appropriate processes.	**N/A:** Document your reasons as to why this does not apply, then go to next question.
			●
Response – You must explain your answer	Not for this prototype.		

*If your solution is operational consult responsible officers for an appropriate equivalent action.

← Move blue dot to your selection.

Human intervention and accountability: For elevated-risk applications, it's crucial to ensure human accountability and intervention capabilities. Consider updating your business continuity plans accordingly to reflect this. This principle may also be relevant for non-elevated risk uses of AI. Doing so will help build public confidence and control in your AI system.

Risk Summary: Highest risks identified

Document the highest risk for each Principle area. The overall highest risk rating will determine your next steps, detailed in the following section “Self-assessment mitigation and next steps”.

N/A / Low / Mid-range / High / Very High

Community benefit AI must prioritise community outcomes, ensuring alignment with laws, minimising harm, and maximising benefit.	**Fairness** Use of AI will be fair, ensuring not to perpetuate bias and inequality by leveraging diverse representative datasets, monitoring performance, and using rigorous data governance	**Privacy and security** Ensure secure, transparent, compliant data use, and adhere to PPIP Act preserving public trust.	**Transparency** The use of AI will be transparent, allowing concerns to be raised and addressed, GIPA Act compliant, cyber secure and ethical.	**Accountability** Decision-making remains the responsibility of organisations and Responsible Officers.
Low risk	Mid range	Very low risk	Low risk	Low risk

NSW GOVERNMENT

4 Self-assessment mitigation & next steps

Identify mitigations and controls to implement and next steps based on the highest residual risk.

1 Scope of applying the framework

2 Self-assessment readiness

3 Self-assessment & risk summary

4 Self-assessment mitigation & next steps

Mitigation and risk plan summary

25. Review your self-assessment, list here the mitigations to be applied and the high-level steps you will take in ensuring these are included in your overall risk management plan. Record your decision, the self-assessment and any supporting information in your Records System.	Stakeholder co-design and consultation, an data science engagement, evaluation and trial and scaled rollout would typically be part of a In a production solution. The orchestrator could have additional safegurards – content filtering, additional verification prior to and post model invocation. Fine tuning and prompt engineering can take place, using feedback from users and results from testing.

If you're **procuring** any part of the solution, complete the procurement questions prior to finalising this section.

ⓘ **Monitoring ongoing performance:** For elevated-risk applications of AI, continuous performance monitoring is crucial. All AI systems should undergo ongoing evaluation, even those considered low-risk, as they could rapidly deviate from normal parameters of operation. Before scaling beyond the pilot phase, it's essential to identify mechanisms for monitoring and calibrating system performance. These mechanisms may include red teaming, conformity assessments, reinforcement from human feedback, monitoring for model drift, and metrics-based performance testing.

ⓘ **Monitoring ongoing risks:** Operational AI systems which progress with High and Very High risks must plan for regular external independent risk audits to cover among other things:

- the examination and documentation of the effectiveness of risk responses in dealing with identified risk and their root causes,
- the effectiveness of the risk management process.

For a more tools on considerations of risk, see "Useful Resources" (Appendix 2) and "Exploring risks, harms and mitigations" (Appendix 5).

Residual risk action

26. Is your project / system an elevated risk? If, after considering all mitigations provided within the self-assessment, Mid-range or higher residual risk(s) persist, this constitutes an Elevated risk use of AI. Use of a non-transparent, non-auditable algorithms or training data will likely be an elevated risk use of AI. They require protections limiting scope of use, or additional risk mitigations.	🟥	🟨	🟩	⬛
	***Yes, I have** **High or Very High-risk residual risk** Don't proceed without legal advice. If the project proceeds, pilot first with ongoing controls and monitoring. A formal review should be conducted after pilot phase. Conduct an independent risk audit, and your self-assessment needs to be reviewed by the NSW AI Review Committee	***Yes, I have** **Mid-Range residual risk.** Don't proceed without legal advice. If the project proceeds, pilot first with ongoing controls and monitoring, consider a review by the NSW AI Review committee and conduct an independent risk audit.	**No, I have** **low residual risk.** Proceed with appropriate controls and monitoring. Consider doing a pilot if there is any potential for the risk profile to increase.	**No, I have** **very low or N/A residual risk** Proceed with appropriate controls and monitoring.
				●

*If your solution is operational consult responsible officers for an appropriate equivalent action.

If you're **procuring** any part of the solution, complete the following slides, if not go to "what happens next" slide 72.

← Move blue dot to your selection.

The importance of documenting your assessment: You must make sure your answers, explanations and risk mitigating controls are recorded in your **Record Management system**. For Elevated risk uses of AI which include Mid-Range risks or higher, the public benefits must be clear and documented before proceeding.

Procurement considerations

Risk treatment in requirements & contract terms

27. When considering risks, did you identify treatments for these risks that were system requirements or contractual controls? In terms of a relative scope for control of potential risks: **Buy** AI and use has high supplier control, low agency control, **Embed** AI and/or co-train has shared supplier and agency control, **Develop** AI and/or train has no supplier control, full agency control. Ensure that supplier services are considered for providing skills development and knowledge transfer to help fulfill your responsibilities.	**Yes:** List the types of treatments that will be applied and categorise them against procurement controls mentioned below.	**No:** Proceed to next step.	**Unclear:** Pause the project and review with the responsible officers and your risk team.
		●	
Response:	No just prototyping for a hackaton.		

Skip this section if your solution has no products or services procured from market.

● ← Move blue dot to your selection.

Translating requirements into controls: Below are examples of translating the AI risk considerations from this framework into requirements and contractual controls

Data Governance: Procurements involving AI systems should establish explicit expectations and implement controls to assure high-quality data is maintained through security-by-design and privacy-by-design principles.

Monitoring ongoing performance: Regular performance evaluations and risk assessments for AI systems should be structured into the service agreement, ensuring that the supplier consistently maintains performance at various stages and checkpoints.

System updates: AI systems often receive updates and enhancements from third-party providers, which occur post-initial risk assessment. These updates necessitate robust control measures to manage any new risks that may be introduced.

Transparency, Explainability, and Auditing: Ensure that purchasers have sufficient transparency and explainability, along with access to third-party auditing. These measures are crucial for effective risk management, justifying decisions, and correctly assigning legal responsibilities to suppliers.

Procurement considerations

Sufficient contract clauses

28. Are the contractual clauses in your contract sufficient for the identified contractual controls?	**Yes:** Proceed to next step.	**No:** If your assessment is against Core&, you must use ICTA. Before proceeding to the next step, and you must reassess against ICTA. If using ICTA, draft appropriate additional conditions through the Order Form of the ICTA to satisfy the treatments.	**Unclear:** Pause the project and consult with either your legal team, responsible officers and risk teams (or both) to determine the status of the clauses and the path forward.
		●	
Response: Provide details regarding your assessment of the contractual controls that the chosen contract has against the inherent risks identified.	Just working on a hackathon.		

Skip this section if your solution has no products or services procured from market.

← Move blue dot to your selection.

ICT purchasing framework amendments: Core& vs ICTA

Core& is designed to be use as-is with no amendments or additions to the substantive terms. This is why it can only be used for procurements that are both low value and low risk.

The ICTA includes the concept of additional condition in the Order Form (Items 11 and 66) that allow for either party to the ICTA to include any terms or conditions that vary or are additional to the terms and conditions set out in the Core or Module terms of the ICTA.

Procurement considerations

Questions for suppliers/partners of services

29. When considering risks, were there any questions that you could not answer or could only partially answer due to supplier provided products or services?	**Yes:** Document the questions below that will require input from suppliers when you approach the market.	**No:** Proceed to next step.	**Unclear:** Pause the project and review with the responsible officers and your risk team.
		●	
Response:			

Skip this section if your solution has no products or services procured from market.

← Move blue dot to your selection.

RFx preparation: RFx documentation should outline supplier requirements to enable you to fulfil your responsibility such as privacy, security measures, algorithmic transparency, bias mitigation, and adherence to ethical principles. This aids in evaluating suppliers for AI selection prioritising safety, security, and ethics.

NOTE: It is critical that once you have the responses from your market engagement that you feed the fully answered questions back through the AI Assessment Framework to determine the outcome and appropriate action.

For further details on the treatment of risks and how/where to apply them in your procurement process, please refer to the NSW Government Procurement Guidance for AI.

Procurement considerations

Ensuring use of correct contracting framework

30. Are there any residual risk factors with a level above "Low"?	**Yes:** You must use the ICTA contract if you proceed.	**No:** You may use Core& or ICTA.	**Unclear:** Pause the project and consult with either your legal team, responsible officers and risk teams before proceeding.
		●	
Response: If your answer is "unclear", please provide further details here.			

Skip this section if your solution has no products or services procured from market.

← Move blue dot to your selection.

ICT Purchasing Framework Risk Levels: All NSW Government agencies must use the ICT Purchasing Framework when buying ICT-related goods and services. The ICT Purchasing Framework comprises:

- Core& contracting framework for ICT procurements that are low risk and up to $1 million (excluding GST)
- MICTA/ICTA contracting framework for ICT procurements that are High risk or over $1 million (excluding GST).

For further details on how to determine the right contracting framework in the context of AI risks, please refer to the NSW Government Procurement Guidance for AI.

For general guidance (i.e., broader than just AI) refer to the Guidelines for Assessing Risk in ICT/Digital Sourcing.

Procurement considerations

System requirements

31. Did you identify any treatments that are system requirements?	**Yes:** Draft Statement of Requirements and Evaluation Criteria adequately address the treatments. Document below the system requirements.	**No:** Proceed to next step.	**Unclear:** Pause the project and review with the responsible officers and your risk team to determine the status of the treatments and the path forward.
		●	
Response:			

Skip this section if your solution has no products or services procured from market.

← Move blue dot to your selection.

RFx preparation: The Statement of Requirements is a crucial document in the procurement process. It serves as a formal artefact that specifies what you, the buyer, aims to achieve, and how you need to achieve it with the system or solution that you are seeking to procure. Beyond defining the project scope, the Statement of Requirements also establishes the evaluation criteria, provides a common language, facilitates comparability, support effective project management and improves vendor compliance.

Procurement considerations

Procurement controls and mitigation

32. Do all risks have appropriate treatments, including the order in which the treatments are applied? Review the set of treatments and the accompanying residual risk to confirm that all risks are appropriately mitigated or controlled.	**Yes:** Document below the treatments and the order in which they are applied	**No:** Pause the project and consult with the appropriate subject matter experts to determine the risk treatment status.	**Unclear:** Pause the project and review with the responsible officers and your risk team to determine the risk treatment status.
		●	
Response:			

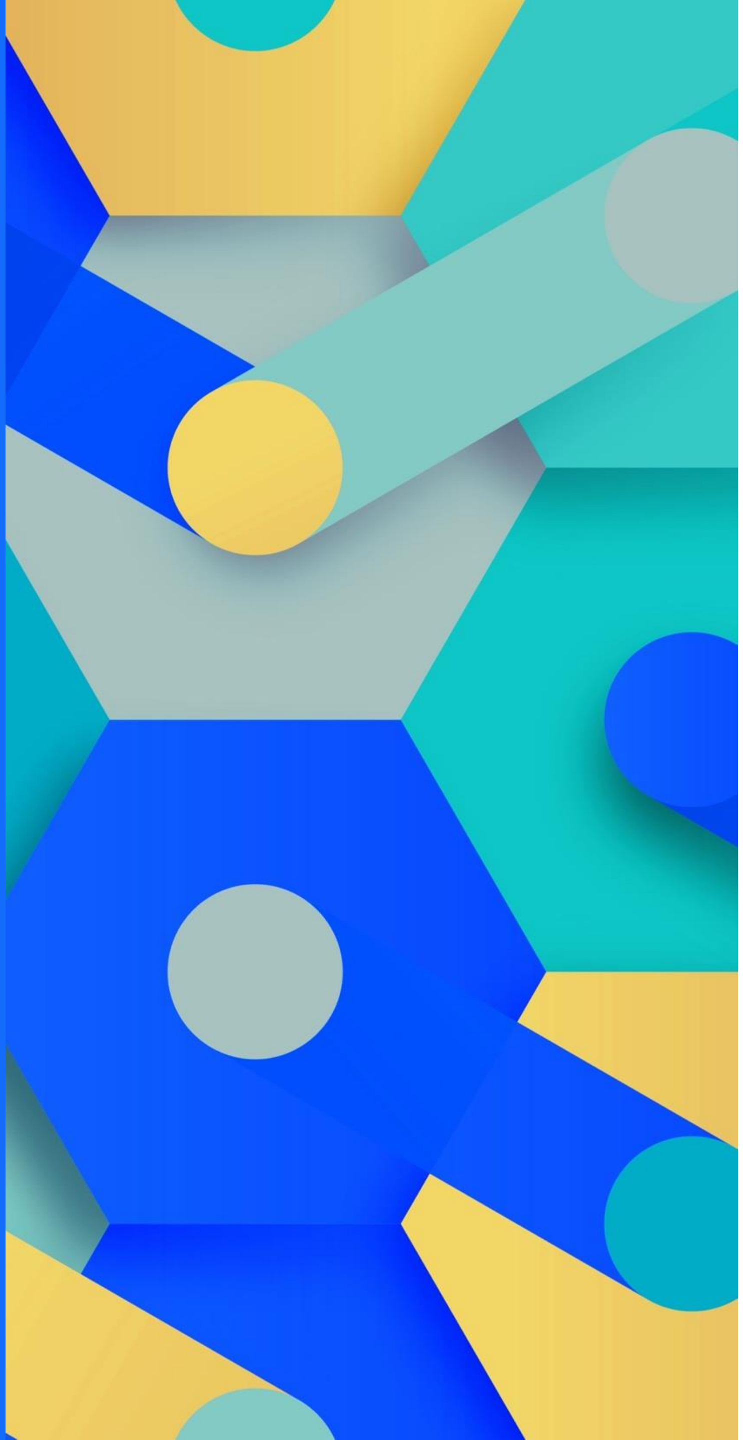
Skip this section if your solution has no products or services procured from market.

← Move blue dot to your selection.

Procurement approvals: After considering the provided Procurement Considerations, remember that AI is one of several factors requiring approval by various Agencies and functions. Approvals are needed from stakeholders in procurement, finance, legal, IT, senior management, and sometimes external parties. For guidance on obtaining these approvals, please contact ICTServices@customerservice.nsw.gov.au

5

End self-assessment stage



What to do next

- After you apply all mitigation, if your residual risk is **high or** greater you must engage the AI Review Committee via the channel specified on slide 8.

- Any residual risk **mid-range** and above you must run a pilot before scaling. Consider running a pilot if there's potential for **low** residual risk to increase.

- Identify when to review risk next. Reassess risk at each project phase and throughout the system lifecycle, following recommended frequencies by your Agency Assurance function or the NSW AI Review Committee

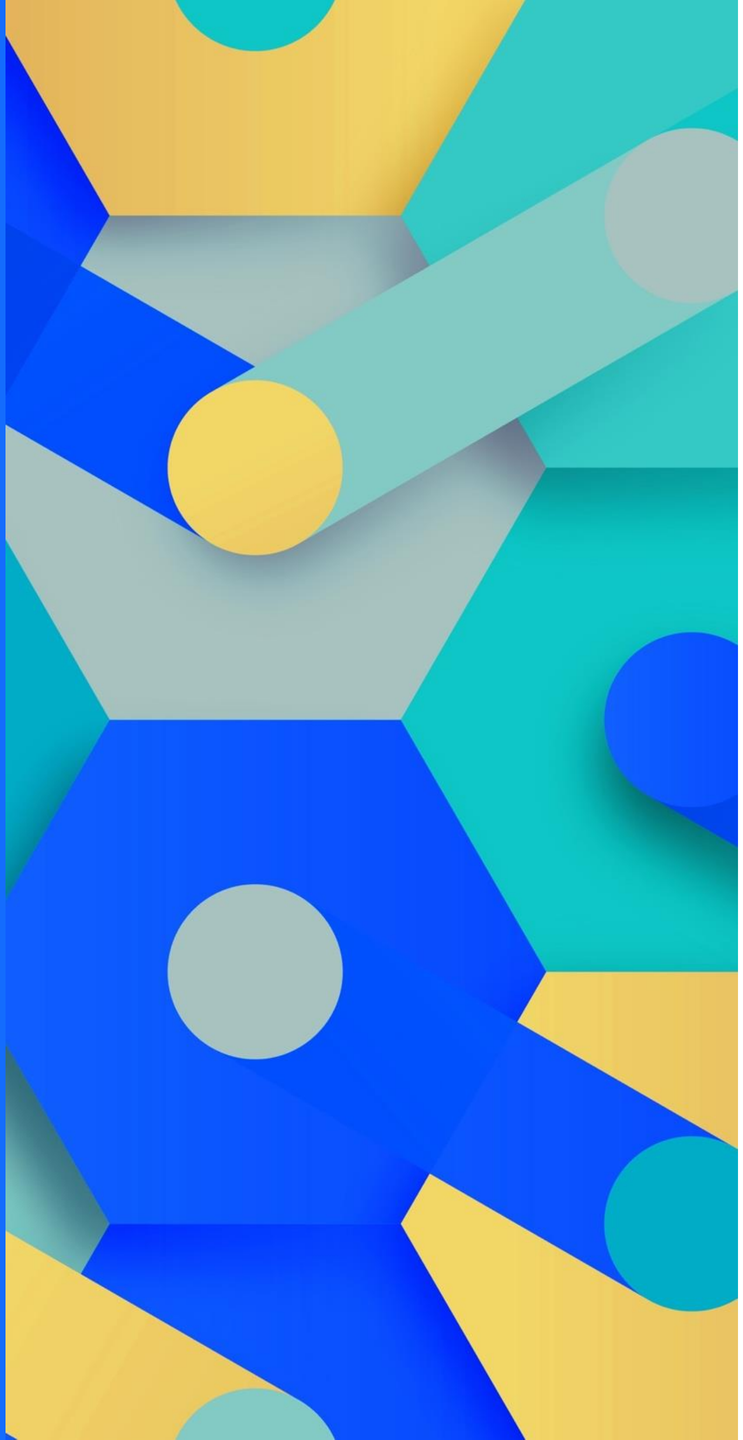
- Integrate identified risks and controls within your Agency's risk management framework.

- Record your self-assessment in your Records management system. Ensure all Responsible Officers approve and have access to the self-assessment record.

- Implement continuous monitoring and evaluations. When reassessing risk, compare it against the self-assessment, documenting any changes as an appendix. Update your risk management plan and records accordingly.

Appendix A

Glossary



Glossary

Administrative Decision Making. Administrative decisions are usually made under legislation and are directed towards a particular person (or organisation). They are different from contractual and commercial decisions and policy and political decisions.

Artificial Intelligence (AI) is the ability of a computer system to perform tasks that would normally require human intelligence, such as learning, reasoning, and making decisions. AI encompasses various specialised domains that focus on different tasks and includes automation.

Bias. In data, this means a systematic distortion in the sampled data that compromises its representativeness, in algorithms it describes systematic and repeatable errors in a computer system that create unfair outcomes, such as privileging one arbitrary group of users over others.

Data Governance. Implementation of a set of policies, processes, structures, roles and responsibilities to ensure that an agency's data is managed effectively and that it can meet its current and future business requirements.

Data Lifecycle. A data life cycle illustrates the stages of data management required over time, from the time of planning and creation to the time that data is either archived or destroyed.

Data Quality. Data quality is generally accepted as meaning "fitness for purpose". It is a term used to describe a documented agreement on the representation, format, and definition for data.

Data use sensitivity. Means risks or considerations associated with data subjects themselves or use of data.

Elevated risk. Elevated risk involves systems influencing decisions with legal or similar level consequences, triggering significant actions, operating autonomously, using sensitive data, risking harm, and lacking explainability.

Generative AI. Is artificial intelligence capable of generating text, images, or other media, using generative models. Generative AI models learn the patterns and structure of their input training data and then generate new data that has similar characteristics.

Hallucination. A hallucination or artificial hallucination is a response generated by an AI which contains false or misleading information.

Harm. Means any adverse effects experienced by an individual (or organisation) including those which are socially, physically, or financially damaging.

Glossary

Human Rights. Are rights inherent to all human beings, regardless of race, sex, nationality, ethnicity, language, religion, or any other status. Human rights include the right to life and liberty, freedom from slavery and torture, freedom of opinion and expression, the right to work and education, and many more. Everyone is entitled to these rights, without discrimination.

Large language model (LLM). A specialised type of artificial intelligence that has been trained on vast amounts of text to understand existing content and generate original content.

Operational AI. Systems that have a real-world effect. The purpose is to generate an action, either prompting a human to act, or the system acting by itself. Operational uses of AI often work in real time (or near real time) using a live environment for their source data.

Responsible Officer. These include the Officer who is responsible for: use of the AI insights / decisions; the outcomes from the project; the technical performance of the AI system; data governance.

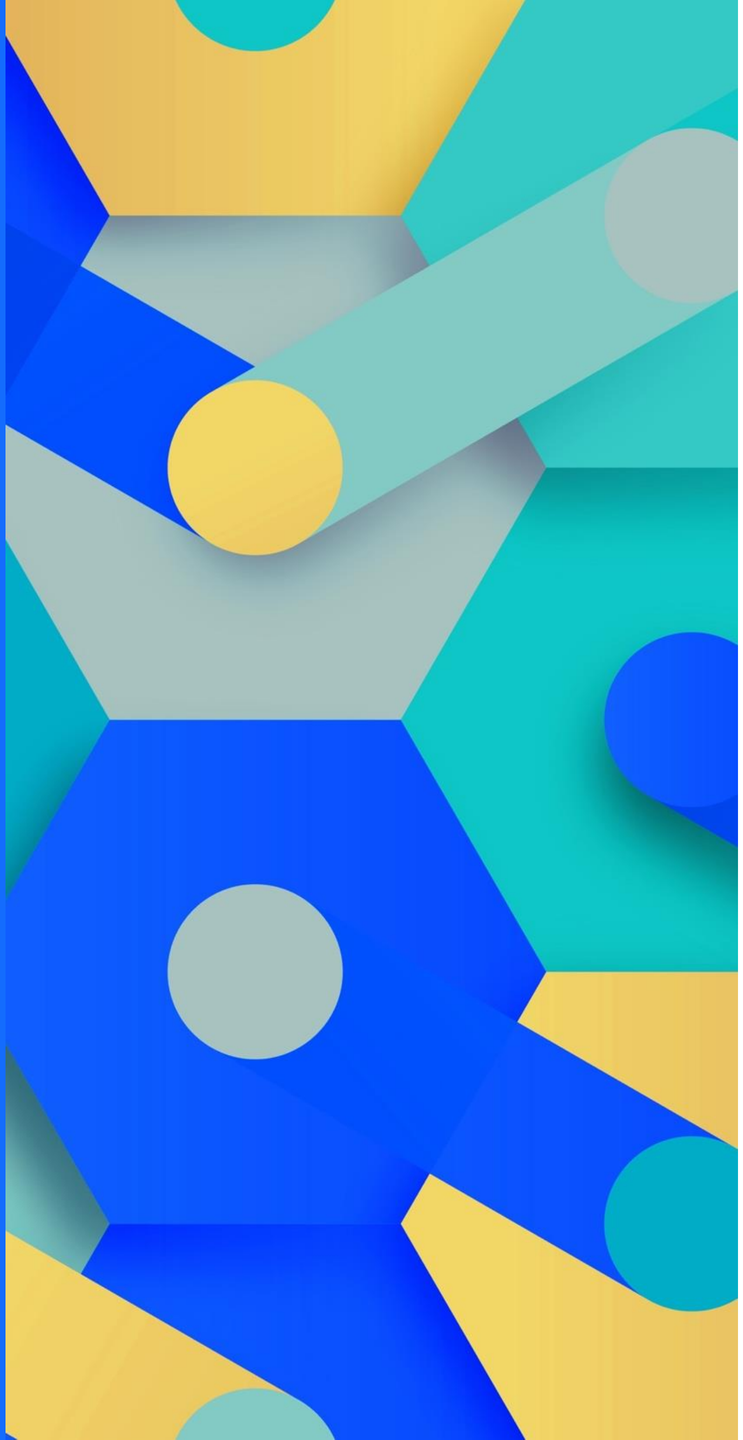
Reversible harm. Means an adverse effect that can be reversed with some level of effort, cost and time.

Secondary Harm. Means any adverse effects experienced by an individual (or organisation) not directly engaged with the AI system, or a subsequent harm identified after an initial harm is experienced by an individual (or organisation) engaged with the AI system.

Significant Harm. Always context specific, a harm which leads to significant concerns. Example from NSW Department of Communities and Justice – *"A child or young person is at risk of significant harm if the circumstances that are causing concern for the safety, welfare or wellbeing of the child or young person are present to a significant extent."*

Appendix B

Useful resources



Resource 1 – Policies, Guides and Frameworks

AI strategy and policy

AI Strategy

Digital Policy Landscape

AI Ethics Policy

General guidance on use of AI

Common AI Definitions

General Guidance on Use of Generative AI

Project governance

ICT Assurance

Benefits Realisation Management Framework

Digital Restart Fund

Privacy

NSW Information and Privacy Commission

Privacy by Design

Guide to NSW Privacy Impact Assessment

Data

NSW Data Strategy

NSW Data Policy

Cybersecurity

Cybersecurity Policy

Cybersecurity Guidelines for AI

Procurement guidance

Note: in development

Automated administrative decision making

For guidance on use of automated administrative decision making, see NSW Ombudsman website.

Benefits realisation framework

NSW Benefits Realisation Management Framework

Project planning and co-design resources

A guide to building co-design capability

General Resources

Gradient Institute (with the National AI Centre hosted by CSIRO) -Implementing Australia's AI Ethics Principles: A selection of Responsible AI practices and resources. This 2023 report provides practical advice to organisations wanting to move beyond talk of High -level AI ethics principles and helps them implement AI responsibly. It gives example practices to help implement the Australian Government’s 8 Ethical Principles and, points to specific online resources to use.

Data Sharing Frameworks The 2023 ACS report “Frameworks and Controls for Data Sharing” identifies the essential controls and methods to ensure that data is treated appropriately throughout its lifecycle, preserving the privacy of individuals while ensuring maximum possible value from data sharing practices.

Risk Rewards and Resilience: developed by ANU, this framework provides a way of analysing complex policy issues, including AI.

Resource 2 - Lean Canvas

Community Benefit from the Use of AI Systems

Community benefit

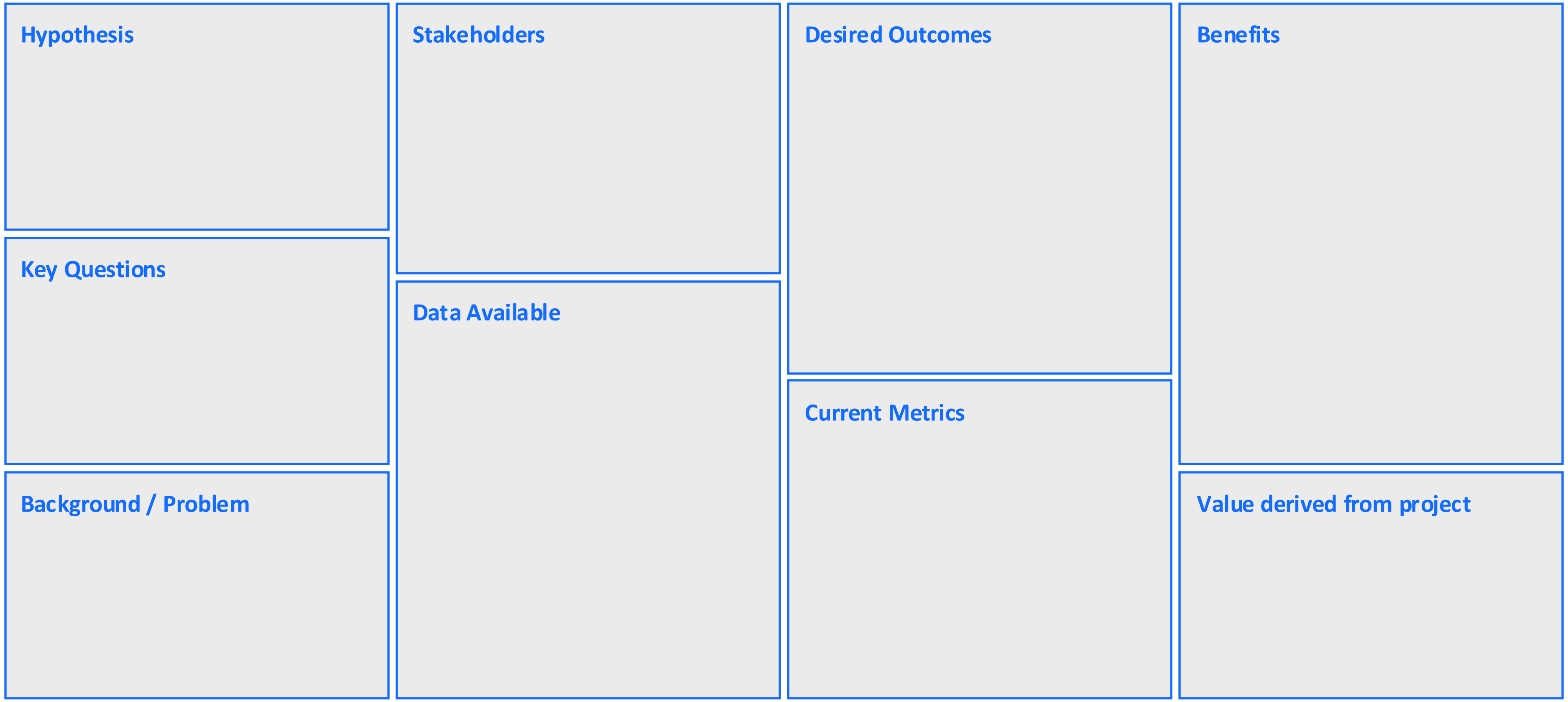
Overall costs and benefits for the project likely to be established by the business case.

Community benefit in the use of AI to be set out:

- Were alternatives to AI considered and why were they discounted?
- How will the use of AI result in improved customer and service delivery outcomes and efficiencies?

Lean Business Canvas: Title of Project

Project Sponsor Name: Name of Sponsor



Hypothesis

Key Questions

Background / Problem

Stakeholders

Data Available

Desired Outcomes

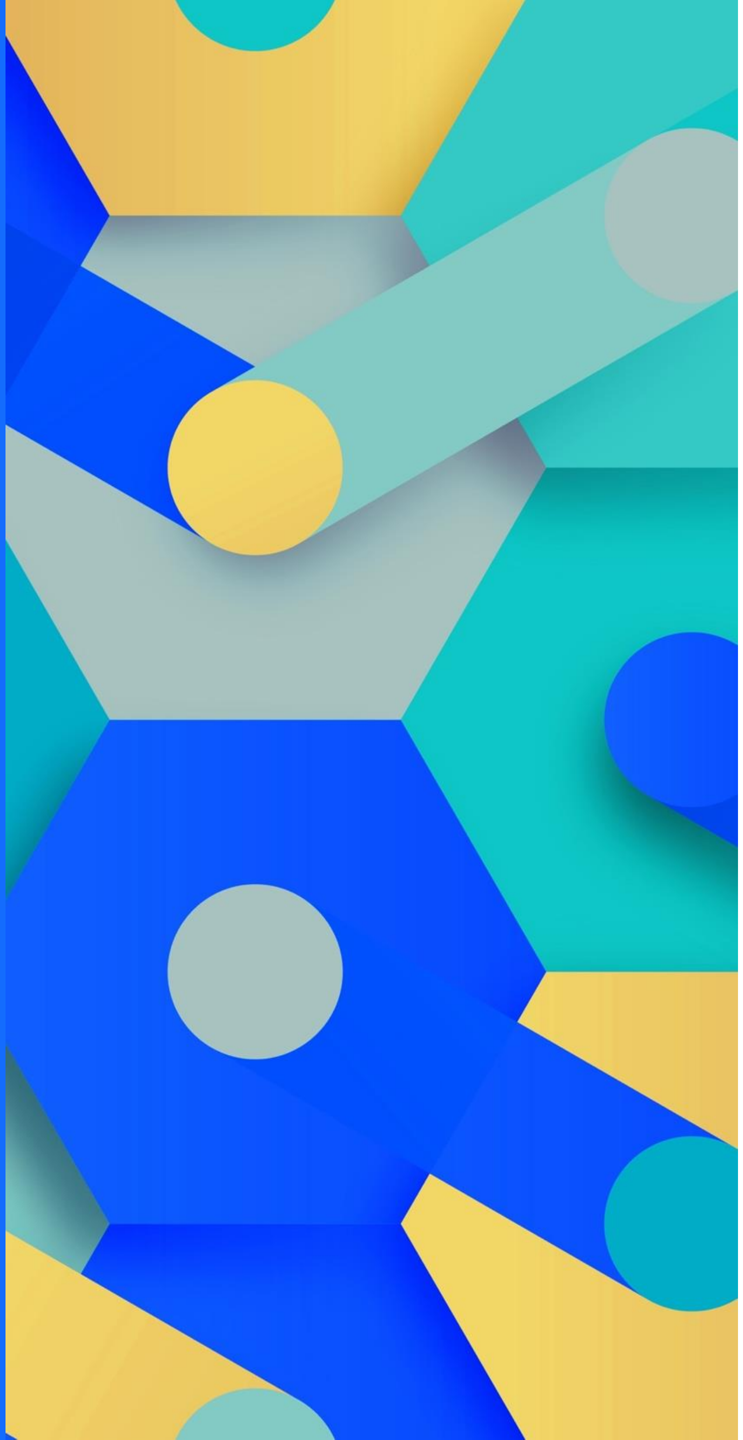
Current Metrics

Benefits

Value derived from project

Appendix C

Some relevant standards



Existing and Developing Standards Families relevant to AI

Where possible adopt and use standards. Standards are continually evolving. The following section provides a snapshot of standards as of the time of this document's release.

Standards Australia provide the Data and Digital Dashboard, where you can interact with and see existing and emerging standards.

The Data and Digital standards landscape June 2022 provides a view of existing and emerging standards. This includes recent standards for data quality AI and ML (ISO/IEC AWI 5259-x standard)

The most recent standard related to AI management systems is the AS ISO/IEC 42001:2023

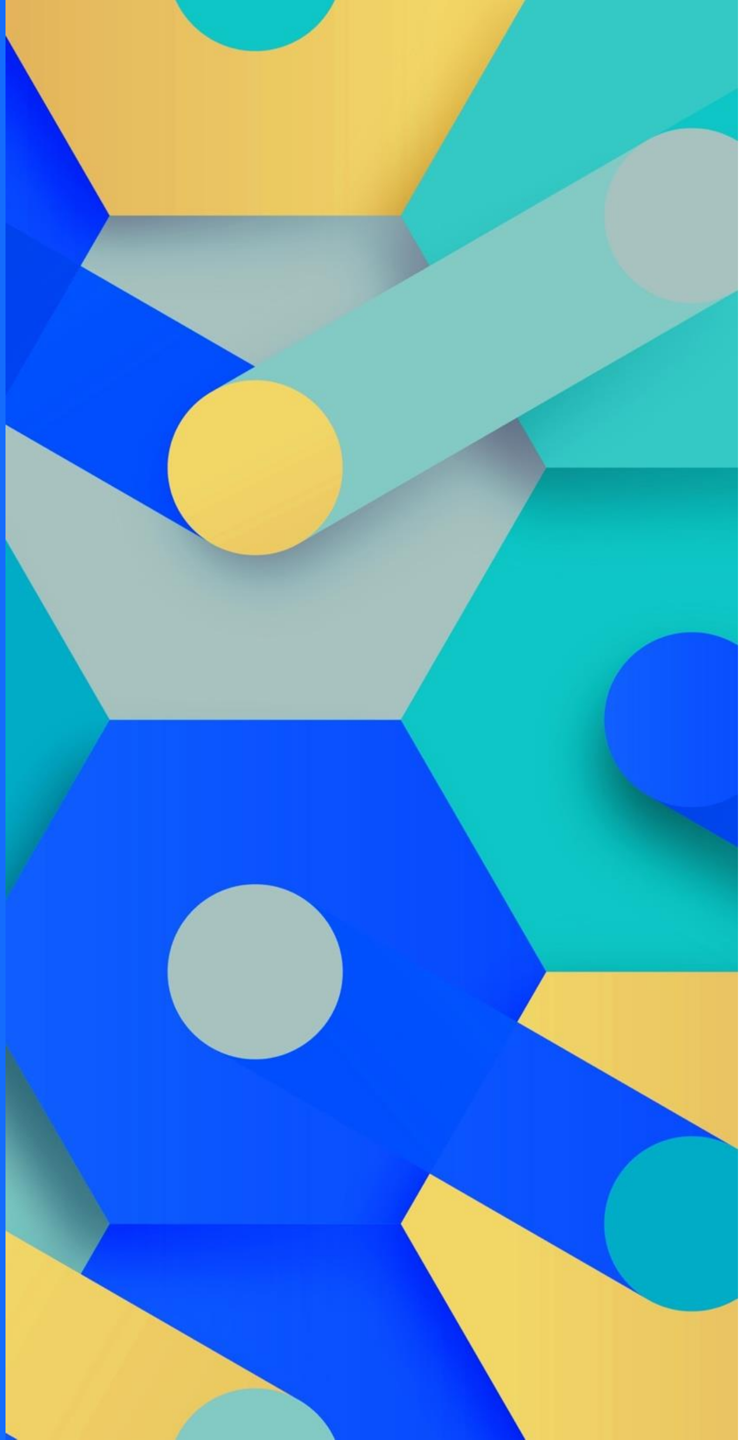
The most relevant groups within the IEC/ISO/JTC1 family include subcommittees (SC) for data sharing and use include:

- SC 27 - Information Security, Cybersecurity and Privacy Protection
- SC 32 - Data Management and Interchange - Within SC 32, Working Group 6 (WG6) on Data Usage
- SC 38 - Cloud Computing and Distributed Platforms
- SC 40 - IT Service Management and IT Governance
- SC 41 – Internet of Things and Digital Twin
- SC 42 - Artificial Intelligence

The ISO/IEC25000 series standards focus on system and software data quality models and requirements.

Appendix D

Example data sharing frameworks



Data sharing frameworks

AI is a "use" of data identifying appropriate frameworks and controls. The ACS report "Frameworks and Controls for Data Sharing" identifies the example controls and methods to ensure that data is treated appropriately throughout its lifecycle, preserving the privacy of individuals while ensuring maximum possible value from data sharing practices.

Taking a simple perspective, AI is a tool that operates on data. Many concerns about the appropriate use of AI relate to considerations across the lifecycle of data leading to its utilisation by AI. This includes the lack of controls associated with the products created from the use of data by AI, as well as concerns about the future use of the original data itself. The range of data products created by AI is very wide, from insights and alerts, to decisions or synthesised material. All however can be treated in similar ways within appropriately considered data sharing and use frameworks.

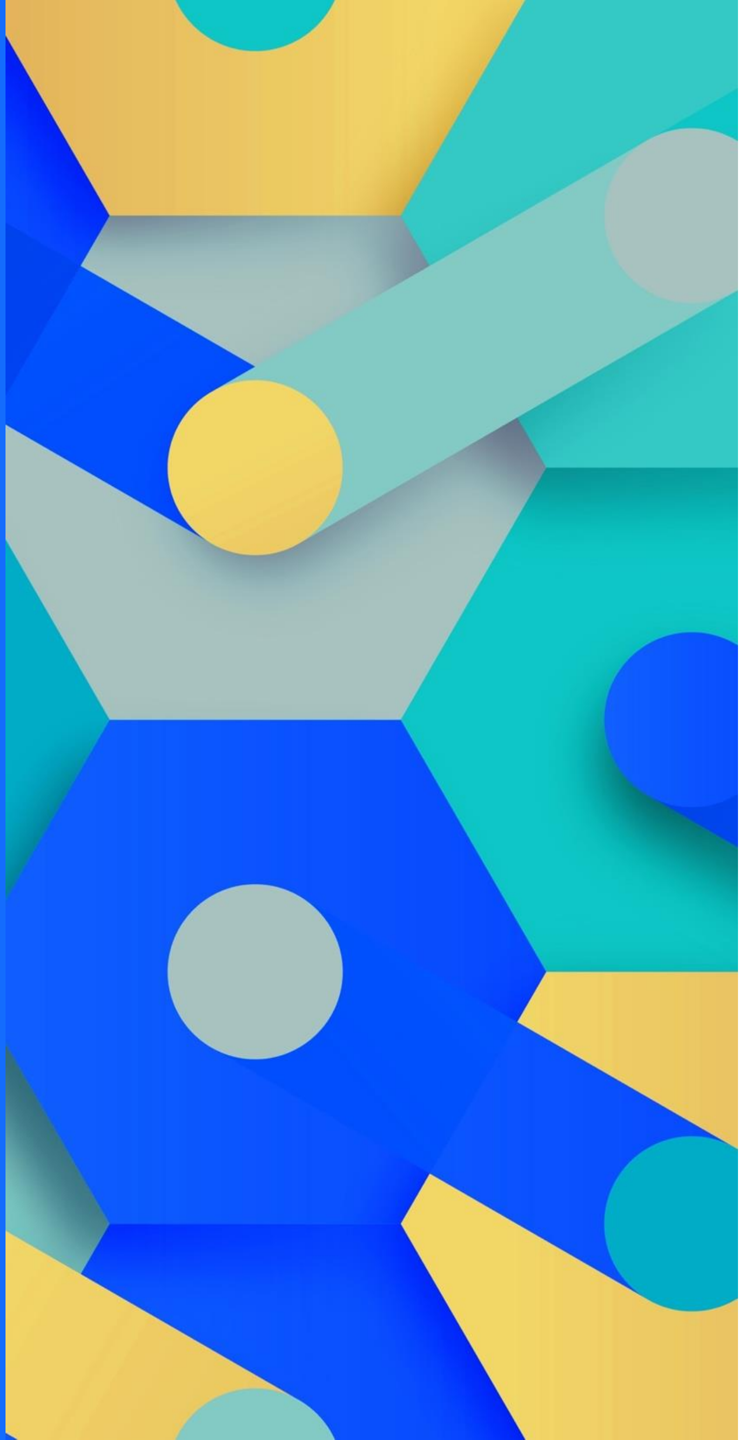
The ACS reports takes several simplifying lenses on

- Data lifecycle
- Identification of "sensitivities" associated with the use of data and data products created by AI
- Different levels of access to data and data products
- Focus areas for governance
- Characterising layers of control for differing levels of sensitivity, levels of personal information and domain expertise required for appropriate use.

acs
Frameworks and Controls for Data Sharing
February 2023
D - data
M - metadata
P - provenance
I - data product
Transmit
Use
Archive/remove
Create/collect
Store
Share
D, M, P and I are potentially modified at each stage of the data life cycle
Metadata can track/capture:
chain of governance
authorising framework
data provenance.
Metadata can be used to assess:
When data/data products are
Raw data
Governance focused on sensitivity
Governance focused on privacy and sensitivity
(e.g. open data)
Governance focused on privacy
Personal information factor
No control
Control in environment
Data life cycle stage
Raw data
Insights

Appendix E

Exploring risks, harms and mitigations



Risks, harms and mitigations

There are many resources available to better understand risks, harms, and mitigations for AI. NSW is committed to collaborating with other states and jurisdictions to share knowledge and lessons learned, making this content available to NSW public servants. While most publicly available resources are dense in content, we encourage everyone to continue learning, applying, and sharing. Some publicly available resources that may assist in your learning and development:

CSIRO AI resources. A valuable asset and resource for Australia, CSIRO has many AI-related resources. The link provided is to their published risk management tool

OECD.AI resources. OECD.AI combines resources from across the OECD, partners and stakeholder groups to create a one-stop-shop for AI policymakers. There site includes a tools and metrics section for AI.

World Economic Forum resources. Containing many valuable resources for managing risk, explanation of country frameworks and news.

NIST AI resources. The National institute of standards and technology, U.S Department of commerce. Provides many resources for A.I including risk management, standards and measures.

Human Technology Institute. The University of Technology Sydney provide great research papers relevant to the AI challenges facing Australia.

James Martin Institute for public policy. JMI is an independent, non-partisan policy institute with charitable status, working to ensure that government can more effectively harness expertise and evidence for the public good.

NSW Ombudsman ADM resources. The NSW ombudsman resources regarding automated decision making, referenced in the self-assessment, are highly recommended.

There are too many resources to list, if you come across AI risk management resources, you think would be very useful for NSW public servants to be aware of, please let us know.

Risk factors to potential mitigations matrix example

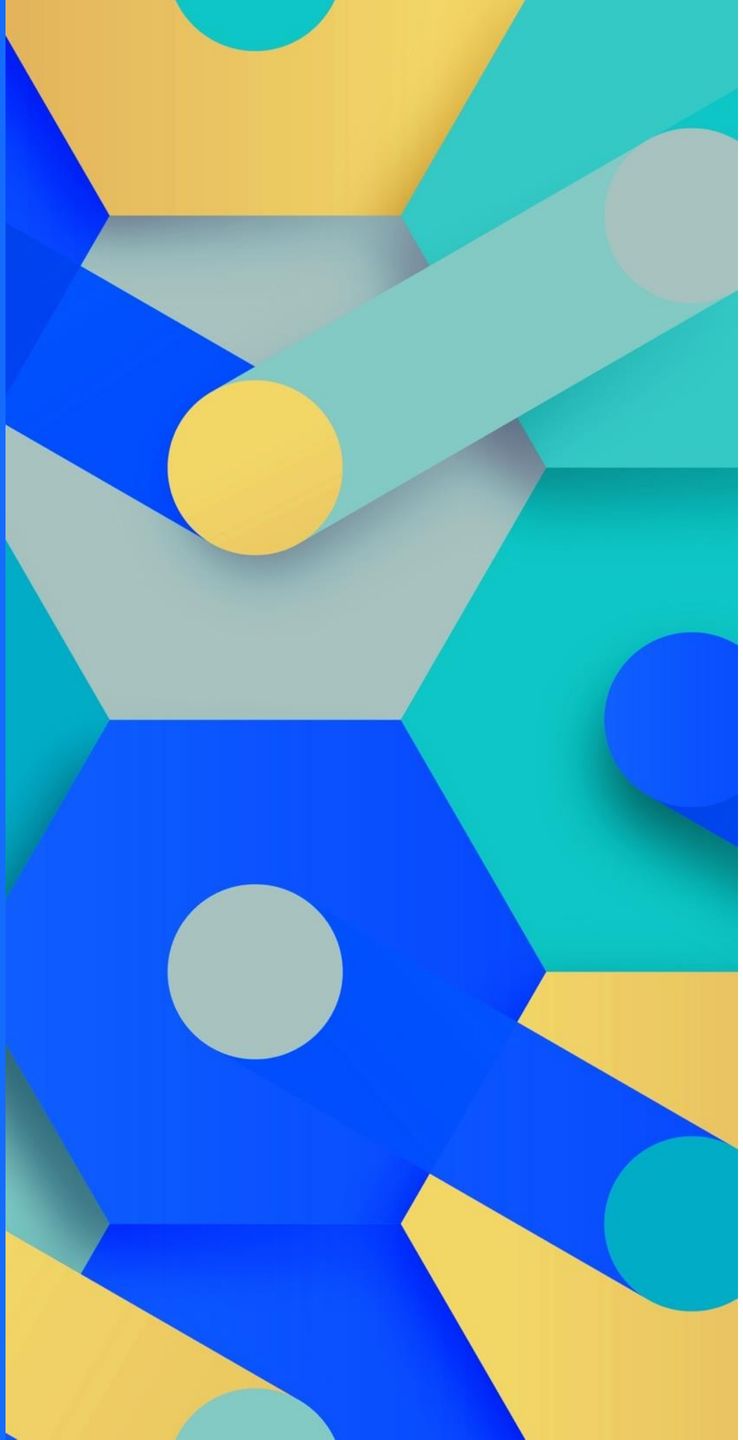
Example template

This is an example of how you could map the risk factors provided in each of the principle areas within the self-assessment to the effectiveness of general mitigations. The example provided is for Fairness.

Assign cells a readiness rating based on how feasibly and effectively each mitigation approach (columns) is likely to handle each risk driver (rows): N/A Low Mid-range High Very High		Mitigation Approaches: potential methods to manage exposure/vulnerability – identify and assess project-specific approaches										
		Domain-specific data / tools	Robust IT infrastructure	Expert and community co-design	Internal testing and research	Data governance / cyber security	AI safety and ethics standards	User risk awareness and training	In-the-loop requirements	Transparency requirements	Accountability mechanisms	Performance monitoring and expert review
Fairness: Drivers of exposure and vulnerability	Using incomplete or inaccurate data											
	Having poorly defined descriptions and indicators of “Fairness”											
	Not ensuring ongoing monitoring of “Fairness indicators”											
	Decisions to exclude outlier data											
	Informal or inconsistent data cleansing and repair processes											
	Informal bias detection methods (e.g., no automated testing)											
	Re-running scenarios could produce different results (reproducibility)											
	Inadvertently creating new associations between data / metadata											
	Differences in the data/methods used for training compared with actual use											

For more information

e: AISecretariat@customerservice.nsw.gov.au



NSW
GOVERNMENT