

Artificial intelligence assurance framework

As described by the NSW Government AI Strategy, AI (Artificial Intelligence) is intelligent technology, programs and the use of advanced computing algorithms that can augment decision making by identifying meaningful patterns in data.

The Framework is intended to be used for custom AI systems, customisable AI systems, and for projects developed using generic AI platforms.

Apply the framework before you use or deploy your AI system. All AI systems should be piloted before being scaled.



Contents

1. Minister’s message	3	5. Overall assessment	47
2. About the AI Assurance Framework	5	Governance requirements	51
3. Engaging with benefits and risks	10	7. Glossary	55
4. Self assessment	14	6. Resources	57
General benefits assessment	16		
General risk factor assessment	17		
Community benefit	18		
Fairness	25		
Privacy and security	32		
Transparency	38		
Accountability	43		
Procurement	46		

Minister's message



Message from the Minister

Introducing the AI Strategy



The Hon. Victor Dominello MP
Minister for Customer Service,
Minister for Digital

“Artificial Intelligence will change everything.”



When it comes to technology, some prefer to hit the brake on progress. Others prefer to press the accelerator on oversight. I am in communion with the latter. Time does not change everything. It is merely the canvas upon which change can be seen. The greatest determinant of change is our dreams, our thoughts, our words, our actions.

Artificial Intelligence will change everything. If we want that change to be positive, then we need a strategy to create that pathway. I am proud to say that the work we have done over the past 12 months under this strategy has laid the foundational paving stones on this pathway.

By 2020, we were promised a world of flying cars and robot assistants. On the surface, our ancestors may be underwhelmed by our progress, but dive beneath and you will find society has made huge leaps of technological advancement that go further and faster than Hollywood’s greatest technology tropes.

Nowhere is this truer than AI. We each may not have our own C-3PO, but we live in a world where AI and machine learning are commonplace, whether in voice assistants like Siri, Alexa, and Google, in recommendations for the next show on Netflix, or in the safety systems of our cars. AI is making our lives easier, safer, and more enjoyable.

The same is true in government. In NSW, we are already using AI to help maintain our trains, protect our endangered species, keep patients safe from sepsis in our hospitals, and to bolster our cyber defences. As we go beyond digital and digitise more government services for our communities, we have a huge opportunity for AI to make these services even simpler, personalised, and secure.

As we embrace the use of AI in government, we must remember this: behind every algorithm is the human who created it. AI is not infallible, and far from it. When unchecked, AI can magnify the biases we deal with every day and cause unintended harm at scale. Government cannot shy away from these issues.

This strategy – the first AI strategy for the NSW Government – has recognised these challenges and charted a course for AI to be used safely across government with the right safeguards in place. This includes thorough consideration of the ethics of any AI use, a recognition of the challenges in buying third-party AI products, and the need to build up our own expert AI skills inside government.

We have created the NSW AI Review Committee to guide and oversight the use of AI in government. The first of its kind in Australia, the Committee has been pivotal in building community trust in the work that we do and has been instrumental in developing the AI Assurance Framework, which will enable us to assure our AI projects against the NSW AI Ethics Framework.

Our strategy was the result of an extensive period of consultation with the community, academic leaders, ethics experts, industry partners, and more than 1000 members of the public. Furthermore, the accomplishments we have made under the strategy have been a continuing product of innovative partnerships between government agencies as well as with representatives from industry and academia.

However, this is just the start of the conversation. AI technology is advancing at such a rapid rate that we must not believe that the pathway set by this strategy is complete. It is imperative that we continue our open dialogue with stakeholders and the public to ensure that the NSW Government continues to recognise and respond to the challenges of AI now and in the future.

By getting this right, we will turbocharge a new wave of government services that will make lives easier and safer for the people of NSW.

I am grateful to everyone who gave their invaluable insights to the development and implementation of the strategy. I hope you will keep the conversation going as we continue into our AI-informed future.

The Hon. Victor Dominello MP
Minister for Customer Service.
Minister for Digital

About the Artificial Intelligence Assurance Framework



About the AI Assurance Framework

What is it?

The AI Assurance Framework will help you design, build and use AI technology appropriately. The framework contains questions that you will need to answer at every stage of your project and while you are operating an AI system. If you cannot answer the questions, the framework will let you know how to get help.

The aim of the framework is to support the NSW Government to innovate with AI technology, while making sure we use it safely and secure, with clear accountability for the design and use of our AI Systems.

Who should use it?

The framework is intended to be used by:

- project teams who are using AI systems in their solutions
- operational teams who are managing AI systems
- Senior Officers who are accountable for the design and use of AI systems
- internal assessors conducting agency self-assessments
- the AI review body (TBC)

When should I use it?

All AI systems and projects must be assessed against the Assurance Framework.

You must use the framework:

- during all stages of an AI project from inception to handover
- periodically to review services that use AI systems.



Is applying this framework everything I need to do?

The framework is not a complete list of all requirements for AI projects. Project teams should comply with their agency-specific AI processes, policy requirements and governance mechanisms as well.



Before you start

In addition to the AI Assurance Framework, there are two additional components of the NSW government's overarching approach to AI:

- NSW Government [AI Ethics Policy Framework](#) (mandatory)
- NSW Government [AI Strategy](#).



When you do not need to apply this framework

You do not need to assess your product or service if:

- you are using an AI system that is a widely available commercial application, and
- you are not customising this AI system in any way or using it in a way other than as intended.

Examples: personal digital assistant, smart phones, smart watches, laptops, QR code reader, satnav system, smart card reader, smoke detector, digital thermometer.

AI systems developed on commercially available software platforms are not exempt.

Commitment to Human Rights

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

Questions to ask of any AI project

- Is 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?

There are laws in NSW that protect the human rights of all people.

Examples include:

[Disability Discrimination Act 1992](#) (Cth)

[Anti-Discrimination Act 1977](#) (NSW)

[International Covenant on Civil and Political Rights 1976](#) (OHCHR, UN)

Publicly available resources:

Australian Human Rights Commission <https://humanrights.gov.au/>

Public Sector Guidance Sheets <https://www.ag.gov.au/rights-and-protections/human-rights-and-anti-discrimination/human-rights-scrutiny/public-sector-guidance-sheets>



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

An initial high level risk assessment should be made on all AI projects to indicate whether a more detailed HRIA would be required.

The parameters for an initial assessment should include the:

- Understanding the goals of the AI project
- Potential harms to people arising from use of the AI system
- Scale of any impact or potential harms
- Degree of transparency of the project or system
- Ethical risk severity (for example: financial, physical, mental)
- Quality of data to be used in the project

How to conduct an AI assurance assessment

1. Assess risk factors

Consider and determine the risk factors for your AI project using the risk matrices in this framework

2. Answer questions & document reasons

Consider and capture your responses to the questions in this framework

Make a decision about whether your project should:

- continue as-is
- continue with additional treatments
- Stop

Consider that any information you capture may be subject to GIPA Act or public disclosure

3. Self assess or submit to the AI review body (TBC)

See next slide for when to submit to the AI review body

Responsible Officers:

- use of the AI insights / decisions:
- the outcomes from the project:
- the technical performance of the AI system:
- data governance:

Comments:



Responsible officers to complete this framework:

This assessment is to be completed by (or the result confirmed with) the Responsible Officers. 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.

These four roles have independent responsibilities and must not be held by the same person. The Responsible Officers should be appropriately senior, skilled and qualified for the role.

When to submit your project to the AI review body*

Larger projects or DRF funded projects

If your project or service:

- uses an AI system, and
- is funded from the [Digital Restart Fund](#), or
- exceeds an estimated total cost of \$5 million.

Your self-assessment must be reviewed by the AI review body (TBC)

The committee will review your assessment, and may make recommendations to help you mitigate risks.

Other projects

If your project or service:

- uses an AI system, and
- Identifies residual risks (after mitigations) which are midrange or higher

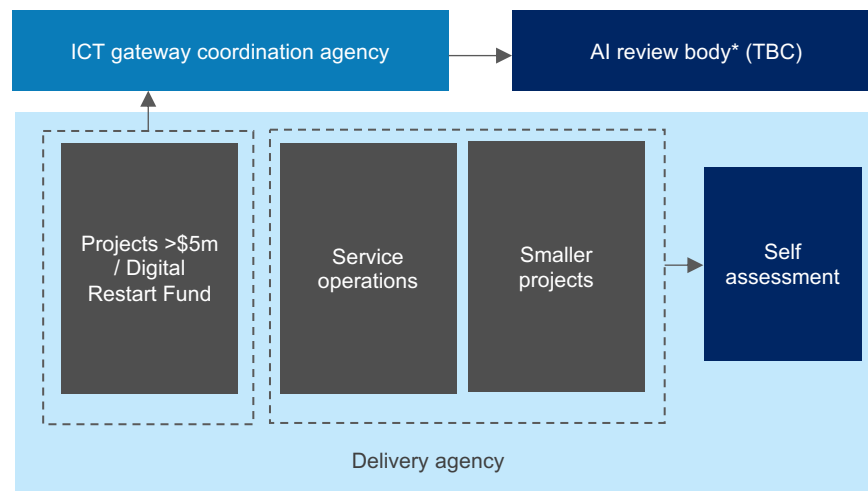
Your self-assessment must be reviewed by the NSW AI review body*

Completing the assessment

In all cases, the project assessment is to be completed by (or the result confirmed with) the Responsible Officers.



* The AI review body is still a work in progress.



Recommendations from the AI review body*

Recommendations from the AI review body* are not binding, but any decision to not meet them should be documented with accompanying reasons.

The Responsible Officers remain responsible for the impact and outcomes of the project.

Engaging with benefits and risks



Evaluating AI benefits and risks

Benefits and risks

NSW Government has a strong commitment to the responsible use of technology.

This means you need to evaluate the potential risks of harms from deployment and operation of AI, as well as its 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 life lifecycle of the technology.

Evaluating and engaging with risk

This AI Assurance Framework is structured in sections that align to the [AI Ethics Principles](#). These principles are mandatory for all NSW Government AI projects.

Each section starts with a page that prompts you to consider the types of risk that your project may bear, and helps shape your response to questions in that section with risk in mind.

At the end of the self-assessment, you will assign a risk rating (highest risk and total number of risks ranked medium or higher) to the different Ethics Principles your AI project. This rating will determine if your project should:

- proceed as is
- proceed, with additional risk mitigations
- stop.



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 proceeds, treat these unanswered questions as representing Midrange risk, commence with a pilot phase and closely monitor for harms and establish controls.



Understanding the balance of benefits and risks

All significant NSW Government ICT projects (including those with AI) are governed by the [ICT Assurance Framework](#).

Some projects carry real risk (for example within Health), but are undertaken to improve existing processes, or because of a clear benefit to community.

Identifying and managing of these risks during the life of the project is an essential requirement, as is clarifying the benefits of the project.

Operational vs non-operational AI

Operational AI

Operational AI systems are those 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 AI systems often work in real time (or near real time) using a live environment for their source data.

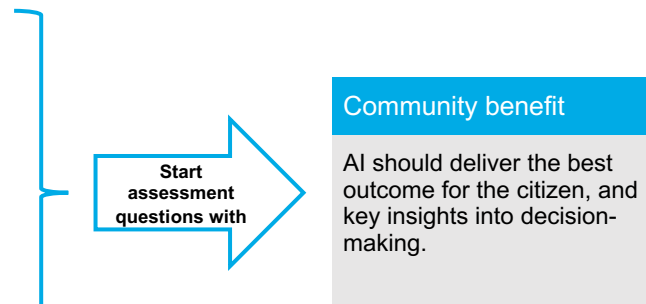
Not all operational AI systems are high risk. An example of lower risk operational AI is the digital information boards that show the time of arrival of the next bus.

Operational AI that uses real-time data to recommend or make a decision that adversely impacts a human will likely be considered High or Very high risk.

Non-operational AI

Non-operational AI systems do not use a live environment for their source data. Most frequently, they produce analysis and insight from historical data.

Non-operational AI typically represents a lower level of risk. However, the risk level needs to be carefully and consciously determined, especially where there is a possibility that AI insights and outputs may be used to influence important future policy positions.

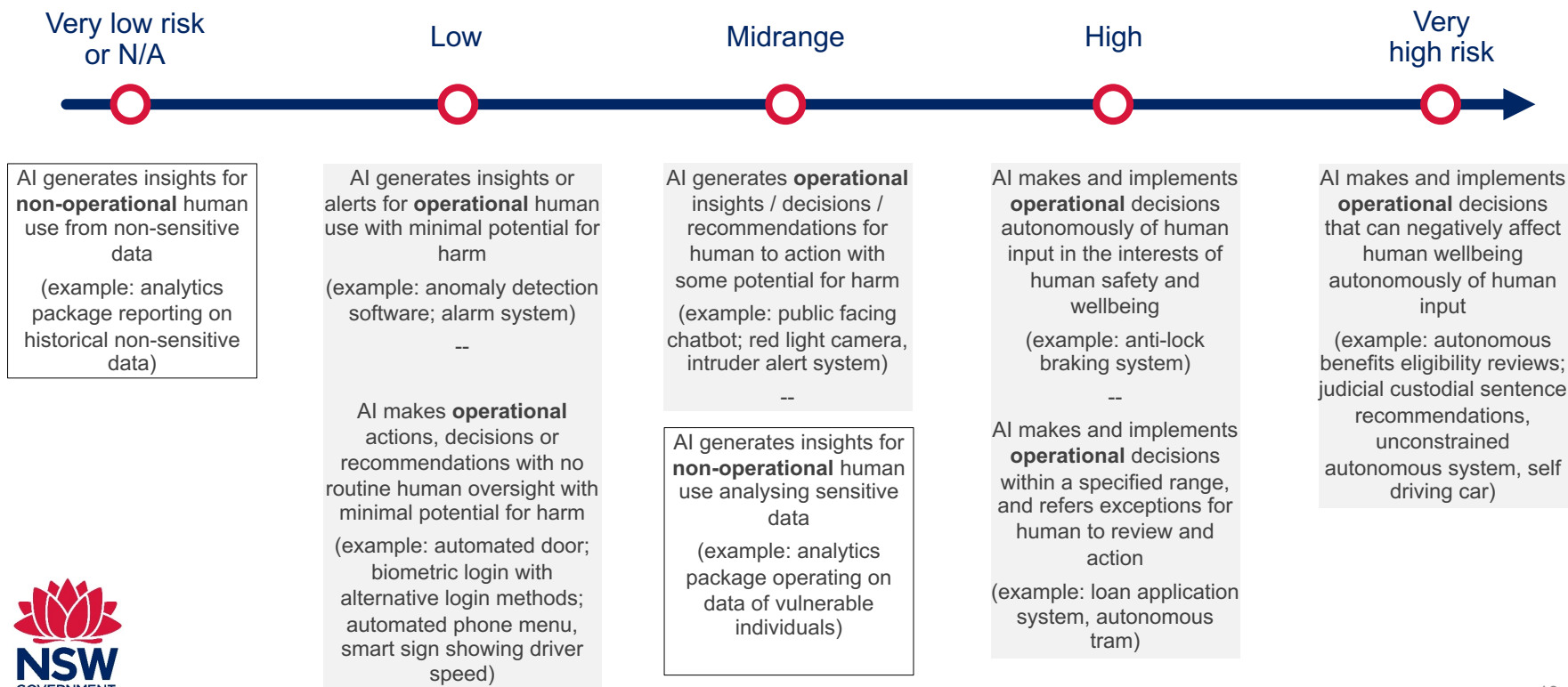


Benefits identification

For all AI systems, the benefits of the AI project should be captured in a [Benefits Realisation Management Plan](#) before commencement.

AI risk factors exist on a spectrum

The key factor that determines risk is how the AI system is used, including whether it is operational or non-operational.



Self assessment



Ethics framework

Mandatory principles

There are five principles that you must apply when using AI. These are mandated through the NSW Government AI Ethics Policy.

Community benefit

AI should deliver the best outcome for the citizen, and key insights into decision-making.

Fairness

Use of AI will include safeguards to manage data bias or data quality risks, following best practice and Australian Standards

Privacy and security

AI will include the highest levels of assurance. Ensure projects adhere to [PPIP Act](#)

Transparency

Review mechanisms will ensure citizens can question and challenge AI-based outcomes. Ensure projects adhere to [GIPA Act](#)

Accountability

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



More information

The Ethics Principles are mandatory. You must consider and apply them when designing, implementing or running an AI System.

You can find out more about the mandatory [Ethical Principles](#) online.

General benefits assessment

Consider the benefits associated with the AI project ...	Very low or N/A	Low	Midrange	High	Very high
Delivering a better quality <i>existing</i> service or outcome (e.g. accuracy or client satisfaction)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reducing processing or delivery times	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generating financial efficiencies or savings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Providing an AI capability that could be used or adapted by other agencies	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Delivering a <i>new</i> service or outcome (particularly if it cannot be done without using AI)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Enabling <i>future</i> innovations to existing services, or new services or outcomes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Benefits realisation management is essential for AI projects

Think about the *potential* benefits of your AI project and the likelihood of these benefits being *realised* in practice; as well the strength of available *evidence* supporting your assessment.

Indicate the overall level of *confidence* in your assessment (e.g. low, midrange, high, very high) and any major variation in the level of confidence between different types of benefit.

Comments:

General risk factor assessment

Consider the risks associated with ...	Very low risk or N/A	Low	Midrange	High	Very high risk
Whether this AI system is delivering a new or existing service	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The potential to cause discrimination from unintended bias	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Whether the AI system is a single point of failure for your service or policy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
If there is sufficient <i>experienced</i> human oversight of the AI system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Over-reliance on the AI system or ignoring the system due to high rates of false alert	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Whether the linkage between operating the AI system and the policy outcome is clear	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Is a new service or policy automatically high risk?

There are always risks associated with a new service or policy simply because it has not been implemented before.

To address the risks of a new service, think ahead about the potential harms, their likelihood and how readily they can be reversed. Also think about the role of human oversight of the new service.

It is important to document your responses to identified risks and provide evidence of controls enacted to mitigate risks.



Comments:

Community benefit

1. Will the AI system improve on existing approaches to deliver the outcomes described in:

- the NSW Premier's priorities
- the Human Services Outcomes Framework
- the Smart Places Outcomes Framework
- NSW Treasury Budget Outcomes
- Your Agency strategic plans or
- another relevant NSW Outcomes Framework?

Response:

- yes _____ document your reasons, then go to next question
- partially _____ after your pilot, you must conduct a formal benefits review before scaling the project. Document your reasons and go to the next question
- not sure _____ pause the project and prepare a [Benefits Realisation Management Plan](#)
- no _____ do not proceed any further. Discuss this project with the policy or service owner



Benefits

All AI projects should have a benefits register that is kept up to date throughout the project.

The benefits register should be handed over to the service owner at the end of the project.

Community benefit

2. Were other, non-AI systems considered?

Response:

- yes _____ document your reasons, then go to next question
- informally _____ after your pilot, you must conduct a formal benefits review before scaling the project. Document your reasons and go to the next question
- no _____ do not proceed any further. Discuss this project with the policy or service owner



Alternatives

For an AI project 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.

Alignment with legal frameworks

3. Does this project 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)

Response:

- yes _____ document your reasons, then go to next question
- unclear _____ pause the project. Seek advice from an appropriate NSW legal source or the NSW Privacy Commissioner. You may need to redesign your project
- no _____ do not proceed any further unless you receive clear legal advice that allows the project to proceed. Consider redesigning your project.



More information

You must comply with privacy and information access laws at all times, including when you are developing and using AI Systems.

AI Projects: Risk factors for individuals or communities

Consider the risks of...	None, negligible, N/A	Reversible with negligible consequences	Reversible with moderate consequences	Reversible with significant consequences	Significant or irreversible
Physical harms	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Psychological harms	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Environmental harms or harms to the broader community	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unauthorised use of health or sensitive personal information (SIP)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Impact on right, privilege or entitlement	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unintended identification or misidentification of an individual	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Misapplication of a fine or penalty	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other financial or commercial impact	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Incorrect advice or guidance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Inconvenience or delay	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other harms	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Very low risk or N/A

Low

Midrange

High

Very high risk



Comments: *these responses should be considered as residual risks after mitigations are place*

Possible harms

4. Considering planned mitigations, could the AI system cause significant or irreversible harms?

If there is a residual risk of significant or irreversible harms and the project proceeds, you must pilot the project first, then conduct a formal benefits review before scaling the project.

For more information on when a Human Rights Impact Assessment is required see <https://humanrights.gov.au/>

Response:

- no _____ document your reasons, then go to next question
- yes, but it's better than existing systems _____ you must seek approval from an ethics committee. You must have clear legal advice that allows this project to proceed. Consult with all relevant stakeholders. Consider a Human Rights Impact Assessment.
- yes _____ do not proceed any further unless you receive clear legal advice that allows the project to proceed. If you have legal approval: discuss the project with all relevant stakeholders, seek approval from an ethics committee, consider a Human Rights Impact Assessment.
- unclear _____ pause the project and prepare a [Benefits Realisation Management Plan](#)



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 do real 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 harms

5. Considering planned mitigations, could the AI system cause reversible harms?

If there is a residual risk of mid-range (or higher) harms and the project proceeds, you must pilot the project first, then conduct a formal benefits review before scaling the project.

Response:

- no _____ document your reasons, then go to next question
- yes, but it's better than existing systems _____ you may need to seek advice from an ethics committee. You should clearly demonstrate that you have consulted with all relevant stakeholders before proceeding to pilot phase. Consider a Human Rights Impact Assessment.
- yes _____ If the risk of harms identified are mid-range or higher, do not proceed any further unless you receive clear legal advice that allows the project to proceed. If you have legal approval: discuss the project with all relevant stakeholders, you may need ethics approval, consider a Human Rights Impact Assessment.
- yes _____ If the risk of harms identified are low or very low, document your reasons, then go to next question
- unclear _____ pause the project and prepare a [Benefits Realisation Management Plan](#)



Irreversible harms vs reversible harms

An irreversible harm occurs when it is impossible to change back to a previous condition. For example, if an AI system makes an incorrect decision to deny somebody a pension without an option to have that overturned.

You should consider how outcomes can be overturned in the event there is harm caused or the AI system leads to an incorrect decision.

Possible secondary or cumulative harms

6. Considering planned mitigations, could the AI System result in secondary harms, or result in a cumulative harm from repeated application of the AI System?

If there is a residual risk of mid-range (or higher) harms and the project proceeds, you must pilot the project first, then conduct a formal benefits review before scaling the project.

Response:

- no _____ document your reasons, then go to next question
- yes, but it's better than existing systems _____ you may need to seek advice from an ethics committee. You should clearly demonstrate that you have consulted with all relevant stakeholders before proceeding to pilot phase. Consider a Human Rights Impact Assessment.
- yes _____ If the risk of harms identified are mid-range or higher, do not proceed any further unless you receive clear legal advice that allows the project to proceed. If you have legal approval: discuss the project with all relevant stakeholders, you may need ethics approval, consider a Human Rights Impact Assessment.
- yes _____ If the risk of harms identified are low or very low, document your reasons, then go to next question
- unclear _____ pause the project and prepare a [Benefits Realisation Management Plan](#)



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. 24

Fairness: risk factors for AI projects

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

Fairness

7. Can you explain why you selected this data for your project and not others?

Response:

- yes _____ document your reasons, then go to next question
- unclear _____ consult with relevant stakeholders to identify alternative data sources or implement a data improvement strategy or redesign the project
- it's better than existing systems _____ document your reasons. You should clearly demonstrate that you have consulted with all relevant stakeholders before proceeding to pilot phase.
- no _____ pause the project and consider how absent data or poor quality data will impact your system.



Data relevance and permission

Your AI system may draw in multiple datasets from different sources to find new patterns and insights.

You need to determine 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.

Fairness

8. Is the data that you need for this project available and of appropriate quality given the potential harms identified?

If your AI project is a data creation or data cleansing application, answer according to the availability of any existing data that is needed for the project to succeed, for example, training datasets.

Response:

- yes _____ document your reasons, then go to next question
- unclear _____ consult with relevant stakeholders to identify alternative data sources or implement a data improvement strategy or redesign the project
- it's better than existing systems _____ document your reasons. You should clearly demonstrate that you have consulted with all relevant stakeholders before proceeding to pilot phase.
- no _____ pause the project and consider how absent data or poor quality data will impact your system.



Data quality

Data quality is often described in terms of minimum requirements for accuracy, timeliness, completeness, and consistency.

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.

Fairness

9. Does your data reflect the population that will be impacted by your project or service?

- yes _____ document your reasons, then go to next question
- it's better than existing systems _____ you may need to seek advice from an ethics committee. You should clearly demonstrate that you have consulted with all relevant stakeholders before proceeding to pilot phase. Consider a Human Rights Impact Assessment
- no or unclear _____ pause the project and address the gaps in your solution design
- N/A _____ document your reasons as to why this does not apply, then go to next question

Response:

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 solution might impact different individuals in minority groups when developing this AI system?

Minority groups may include:

- those with a disability
- LBGQIT+ and gender fluid communities
- people from CALD backgrounds
- Aboriginal and Torres Strait Islanders
- children and young people

- yes _____ document your reasons, then go to next question
- it's better than existing systems _____ you may need to seek advice from an ethics committee. You should clearly demonstrate that you have consulted with all relevant stakeholders before proceeding to pilot phase. Consider a Human Rights Impact Assessment
- no or unclear _____ pause the project and address the gaps in your solution design
- N/A _____ document your reasons as to why this does not apply, then go to next question



Diversity and inclusion, and the impact on minorities

Services or decisions can impact different members of the relevant community in different ways.

Whether due to cultural sensitivities, or underrepresentation in training data sets. It is important to think deeply about everyone who might be impacted by AI Systems.

Fairness

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?

Response:

- yes _____ document your reasons, then go to next question
- no or unclear _____ for operational AI systems, pause the project until you have established performance measures and targets. for non-operational systems, results should be treated as indicative and not relied on.
- N/A _____ document your reasons as to why this does not apply, then go to next question



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

13. Do you have a way to monitor and calibrate the performance (including fairness) of your AI system?

Operational AI systems which are continuously updated / trained can quickly move outside of performance thresholds. Supervisory systems can monitor system performance and alert when calibration is needed.

Response:

- yes _____ document your reasons, then go to next question
- no or unclear _____ for operational AI systems, pause the project until you have established performance measures and targets. for non-operational systems, results should be treated as indicative and not relied on.
- N/A _____ document your reasons as to why this does not apply, then go to next question



Measuring AI system performance

Operational AI systems should have clear performance monitoring and calibration schedules.

For operational AI systems 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.

Sensitive data considerations for AI projects

Do you use 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
Children	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Religious individuals	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Racially or ethnically diverse individuals	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Individuals with political opinions or associations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Individuals with trade union memberships or associations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gender and/or sexually diverse individuals	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Individuals with a criminal record	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Specific health or genetic information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Personal biometric information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other sensitive person-centred data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Very low risk or N/A	Low	Midrange	High	Very high risk



Comments: *these responses should be considered as residual risks after mitigations are place*

Privacy and security

14. Have you applied the “Privacy by Design” and “Security by Design” principles in your project?

Response:

- yes — document your reasons, then go to next question
- partially — pause the project, consult with your stakeholders and determine how you will improve your data or practices
- no or unclear — pause the project until you have received appropriate advice including from the Information and Privacy Commission. You may need to redesign your project.



Privacy by design, security by design

Even small AI projects may have privacy or security vulnerabilities. For example, an analytics project 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.

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.

Privacy and security

15. Have you completed a privacy impact assessment (either third party or self assessed)?

Response:

- yes _____ document your reasons, then go to next question
- no _____ pause the project until you have completed a privacy impact assessment.



Privacy impact assessment

Even projects 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

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 also the NSW Privacy Commissioner's [fact sheet](#) on Reasonably Ascertainable Identity

Response:

- yes — document your reasons, then go to next question
- 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 this project to proceed. The project should be carefully monitored for harms during the pilot phase.
- partially — pause the project until you have consent, or redesign your project
- no — pause the project until you have either consent or clear legal advice authorising use of this information
- N/A — document your reasons as to why this does not apply, then go to next question



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 Page 50: Governance Requirements.

Privacy and security

17. Does your AI System adhere to the mandatory requirements in the NSW Cyber Security Policy?

Have you considered end-to-end Security Principles for your project?

Response:

- yes _____ document your reasons, then go to next question
- no or _____ pause the project until these requirements can be met partially
- N/A _____ document your reasons as to why this does not apply, then go to next question



Cyber security

As with any emerging technology, AI can pose new cyber security risks and so it is important to 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 including malware and ransomware attacks.

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 (see slide 32)?

Response:

- no _____ document your reasons, then go to next question
- yes _____ seek explicit approval from the Responsible Senior Officer to proceed with this risk. Consider seeking approval from an ethics committee.
- unclear _____ pause the project and clarify the nature of the data, address any inadvertent use of sensitive data in you system



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 Page 50: Governance Requirements.

Transparency: risk factors for AI projects

Consider the risks associated with...	Very low risk or N/A	Low	Midrange	High	Very high risk
Incomplete documentation of AI system design, or implementation, or operation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
No or limited access to model's internal workings or source code ("Black Box")	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Being unable to explain the output of a complex model	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A member of the public being unaware that they are interacting with an AI system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
No or low ability to incorporate user feedback into an AI system or model	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Is a 'black box' AI system automatically high risk?

The inner workings of commercial AI systems are not always accessible and even if they are, they can be very complex to interpret.

To address the risks this poses, think proactively about the role of human judgement in use of an "unexplainable" insight or decision. If you cannot explain the ways in which insights are outputted from an AI system, what are the potential harms that may arise? What's the likelihood of these harms and how readily they can be reversed? It is important that these considerations are documented. This is particularly important if midrange or higher risks are identified.

Comments: *these responses should be considered as residual risks after mitigations are place*

Transparency

19. Have you consulted with the relevant community that will benefit from (or be impacted by) the AI system?

Response:

- yes _____ document your reasons, then go to next question
- 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 this project to proceed. The project should be carefully monitored for harms during the pilot phase.
- it's better than existing systems _____ you may need to seek advice from an ethics committee. Document your reasons. You should clearly demonstrate that you have consulted with all relevant stakeholders before proceeding to pilot phase.
- 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



Consultation

You must consult with the relevant community when you design your AI system. This is particularly important for operational AI systems.

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 systems intended to operate under legislation which allows use without community consultation, the public benefits must be clear before proceeding to pilot phase.

Transparency

20. Are the scope and goals of the project publicly available?

Response:

- yes _____ document your reasons, then go to next question
- no _____ make sure you communicate the scope and goals of the project to relevant stakeholders and the relevant community who are impacted before proceeding beyond pilot
- N/A _____ document your reasons as to why this does not apply, then go to next question



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

21. Is there an easy and cost effective way for people to appeal a decision that has been informed by your AI system?

Response:

- yes _____ document your reasons, then go to next question
- 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



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, 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 decision.

Transparency

22. Does the system using the AI allow for transparent explanation of the factors leading to the AI decision or insight?

Response:

- yes _____ document your reasons, then go to next question
- no, but a _____ 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 during pilot phase..
person makes the final decision
- 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



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 (e.g. a person overriding an automated barrier).

Accountability: risk factors for AI projects

Consider the risks associated with ...	Very low risk or N/A	Low	Midrange	High	Very high risk
Insufficient training of AI system operators	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Insufficient awareness of system limitations of Responsible Officers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
No or low documentation of performance targets or "Fairness" principles trade-offs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
No or limited mechanisms to record insight / AI System decision history	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The inability of third parties to accurately audit AI system insights / decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



The skill and training of the operators of the AI system are the most important elements

With all automated systems, there is always the risk of over-reliance on result. It is important that the operators of the system, including any person who exercises judgement over the use of insights, or responses to alerts, is appropriately trained on the use of the AI system. Training must include the ability to critically query insights generated, and to understand the limitations of the AI system.

For operational AI systems, the users must be confident they can readily reverse any harms resulting from the use of an AI generated insight or decision, or ensure a Responsible Officer is empowered to make a decision on the use of an AI generated insight. For non-operational AI systems, the users must be skilled in the interpretation and critiquing of AI generated insights if the insight is to be relied upon.

Comments: *these responses should be considered as residual risks after mitigations are place*

Accountability

23. Have you established who is responsible for:

- use of the AI insights and decisions
- policy/outcomes associated with the AI system
- monitoring the performance of the AI system
- data governance?

Response:

- yes _____ document your reasons, then go to next question
- no or unclear ____ pause the project while you identify who is responsible and make sure they are aware and capable of undertaking their responsibilities
- N/A _____ document your reasons as to why this does not apply, then go to next question



Responsible officers:

This assessment is to be completed by or (the result confirmed with) the Responsible Officers. 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.

These four roles must not be held by the same person. The Responsible Officer should be appropriately senior, skilled and qualified for the role.

Accountability

24. Have you established a clear processes to:

- intervene if a relevant stakeholder finds concerns with insights or decisions?
- ensure you do not get overconfident or over reliant on the AI system?

Response:

- yes _____ document your reasons, then go to next question
- 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



Human intervention and accountability

For operational AI systems, you must make sure that humans are accountable and can intervene. This may also be relevant for non-operational AI systems

This will help you to build public confidence and Control in your AI system.

Procurement

25. If you are procuring all or part of an AI system, have you satisfied the requirements for:

- transparency?
- privacy and security ?
- fairness?
- accountability?

Response:

- yes _____ document your reasons
- no _____ pause your project. Make sure you can meet the requirements before you continue.



Engaging with NSW procurement

The procurement process may be the best place to ensure the mandatory policy requirements for AI systems are considered early on. Mechanisms for ensuring performance, ongoing monitoring and calibration may be negotiated and built into contractual agreements with vendors.

Please make sure you seek help from procurement experts and communicate regularly with the Responsible Officer.

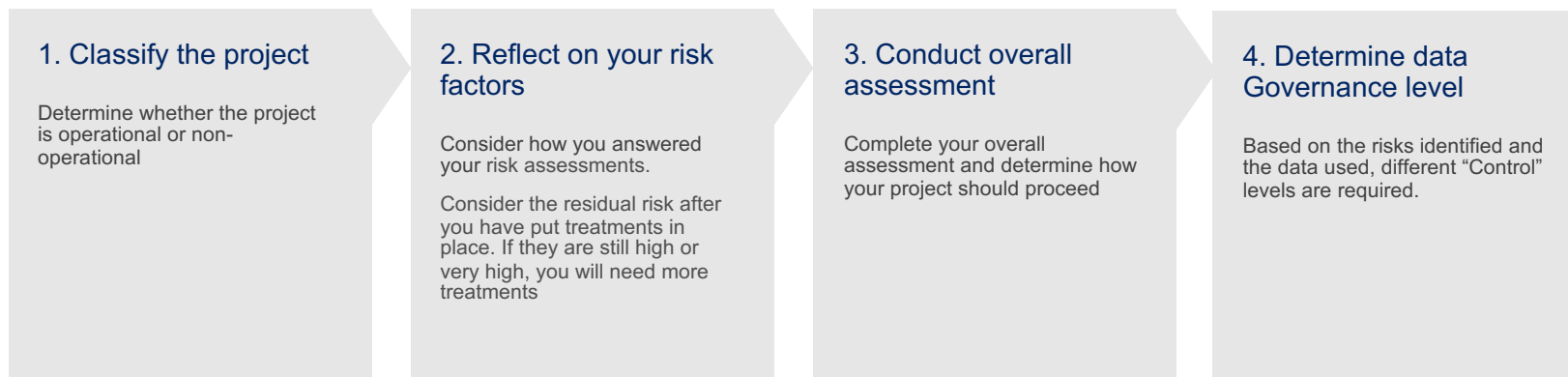
Overall assessment



Completing the overall assessment

How to complete this section

You will need to make an overall assessment that will determine whether the project should continue as-is, continue with additional mitigations, or stop altogether.



Risks Identified

Community benefit	Fairness	Privacy and security	Transparency	Accountability
AI should deliver the best outcome for the citizen, and key insights into decision-making.	Use of AI will include safeguards to manage data bias or data quality risks, following best practice and Australian Standards	AI will include the highest levels of assurance. Ensure projects adhere to PPIP Act	Review mechanisms will ensure citizens can question and challenge AI-based outcomes. Ensure projects adhere to GIPA Act	Decision-making remains the responsibility of organisations and Responsible Officers.
Highest risk:	Highest risk:	Highest risk:	Highest risk:	Highest risk:
No. of Risks:	No. of Risks:	No. of Risks:	No. of Risks:	No. of Risks:



Monitoring ongoing risks

Operational AI projects which progress with high and very high risks must plan for regular external risk audits to cover

- 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



Tallying risks

Highest risk refers to the most significant risk identified in each of the five principle areas (e.g., “Community Benefit” or “Fairness”).

No. of Risks refers to the count of medium, high and very high risks
All medium, high and very high risks must have effective mitigations.

Projects which progress with medium, high and very high risks must have project-specific legal advice.

Overall assessment

Is this an operational AI system?

It is operational AI if the system uses real-time or near real-time data to:

- make recommendations for humans to act on in real-time or near real-time
- or
- take actions itself in real-time or near real-time.

- | | |
|--|---|
| ● yes, and the decisions it makes or informs include <u>high</u> or <u>very high</u> risk factors _____ | do not proceed without project-specific legal advice. If the project proceeds, pilot first with ongoing controls and monitoring. A formal review should be conducted after pilot phase. Use of an external review committee is recommended. |
| ● yes, and the decisions it makes or informs include <u>medium</u> risk factors _____ | do not proceed without project-specific legal advice. Pilot first with ongoing controls and monitoring required once pilot commences. |
| ● yes, and the decisions it makes or informs include low or negligible risk factors _____ | your project can proceed with appropriate ongoing controls and monitoring. Pilot the project first. |
| ● no, it relies on historical data, however its outputs may be used to inform policy and other important decisions _____ | your project can proceed, but you need to review your risk treatments and make sure there are sufficient controls in place |
| ● no, it only uses historical data for reporting or informing purposes only _____ | your project can proceed with appropriate ongoing controls and monitoring |



Monitoring ongoing performance

For operational AI systems, ongoing performance monitoring is essential. Even low risk systems such as an automated barrier, could rapidly change to operate outside of normal parameters. Mechanisms to monitor calibrate system performance should be identified before scaling beyond pilot phase.



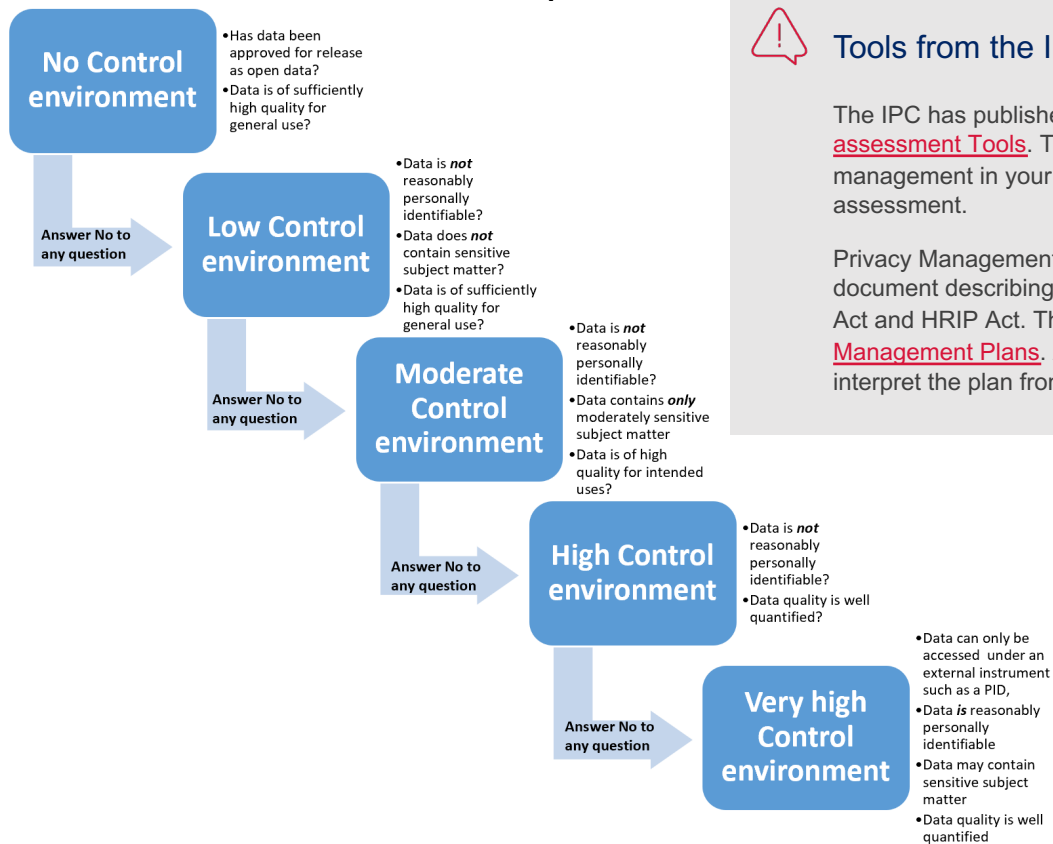
The importance of documenting your assessment

You must make sure your answers, explanations and risk mitigating controls are recorded in your document management system.

For operational AI systems which include medium risks or higher, the public benefits must be clear and documented before proceeding to pilot phase. Project specific legal advice is required. Projects should be actively monitored for potential harms and remedies identified.

Determining Data Governance Requirements

What the Level of Control is Required?



Tools from the IPC

The IPC has published Information Governance Agency [Self-assessment Tools](#). These tools may be useful to self assess privacy management in your organisation. The IPC recommends regular self-assessment.

Privacy Management Plan – agencies must have a strategic planning document describing how the organisation will comply with the PPIP Act and HRIP Act. The IPC has a Guide on making [Privacy Management Plans](#). Agencies can seek support to access and interpret the plan from the agency’s Privacy Officer.

Data Governance Environment Required

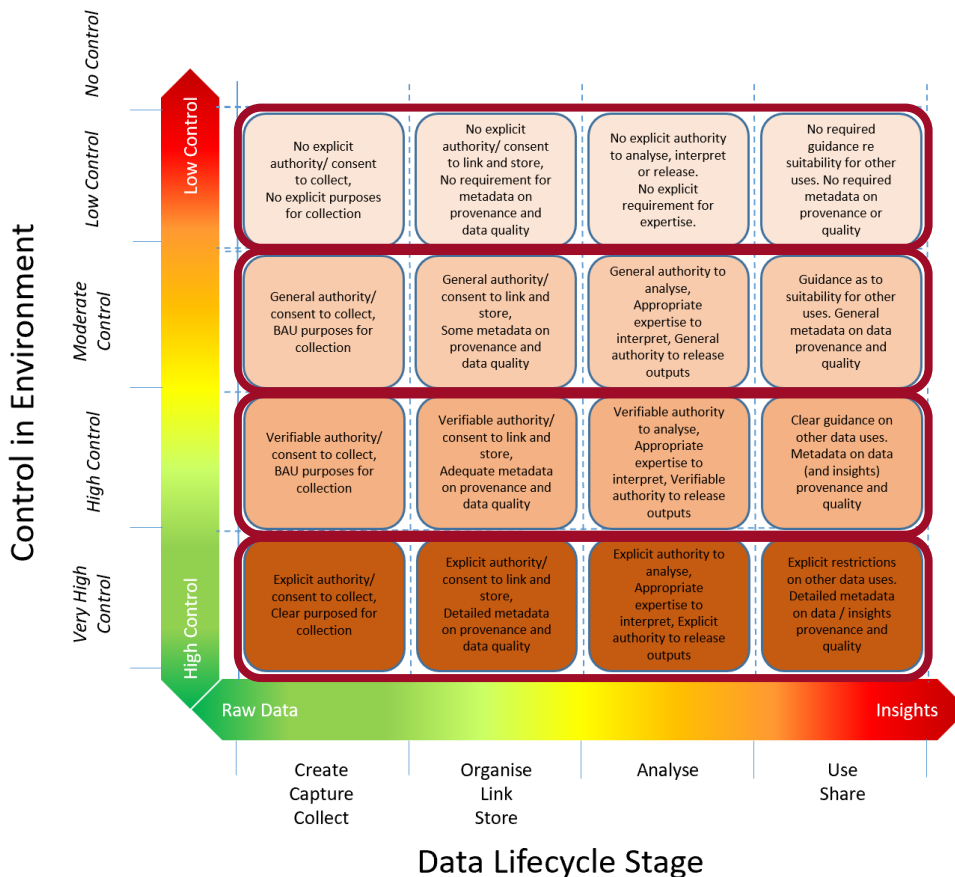
Control (Capability + Governance + Purpose) at Each Stage

Low control. May have assumed authority to collect, use, and reuse data. May have metadata on data provenance and quality. Data - low level of personal information.

Moderate control. Must have understanding of data quality and provenance, capable analysts and domain experts, adequate governance / security at each stage. May have broad authority to collect, use, and reuse data. Data - moderately sensitive / moderate level of personal information.

High control. Must have understanding of data quality and provenance, highly skilled analysts and domain experts, strong governance / security at each stage. May have general authority to collect, use, and reuse data. Data - high sensitivity / high level of personal information.

Very high control. Must have explicit purpose and authority, high quality data and metadata, expert analysts and domain experts, strong governance / security at each stage. Explicit restrictions on secondary use of data and insights. Data - very high sensitivity and very high personal information



- Control = (proven) capability * (assessable) governance * (verifiable) purpose
- Capability includes skill in all stages of Data Lifecycle - data analysis, data provenance, governance, security
- High Control = skilled people working in strong governance environment with clearly authorised purpose
- No Control environment = no assessments or no restriction on people accessing or utilising data
- Requires an objective, repeatable, standardised assessment of
 - capability,
 - governance,
 - purpose,
 - data quality and provenance
 - sensitivity of data
 - degree of personal information contained in datasets



Characterising Levels of “Control” Required for Data Governance

No Control environment suitable for:

- Data which is **not** reasonably personally identifiable
- Data which has been approved for release as open data
- Data which is of sufficiently high quality for general use

Low Control environment suitable for:

- Data which is **not** reasonably personally identifiable
- Data does not contain sensitive subject matter
- Data which is of sufficiently high quality for general use

Moderate Control environment suitable for:

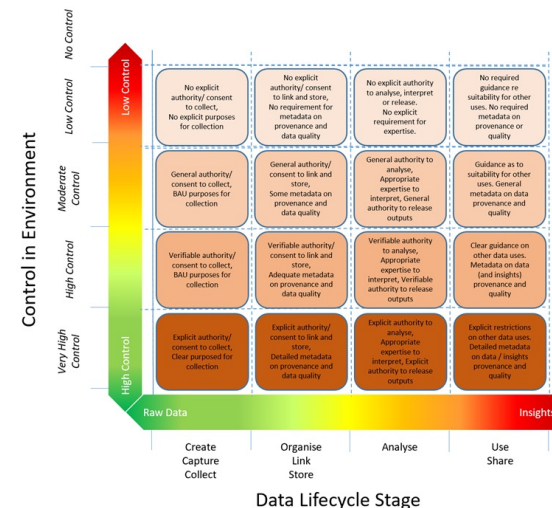
- Data which is **not** reasonably personally identifiable
- Data which contains some sensitive subject matter
- Data which is of high quality for general use
- People with access have met General requirements for a “Safe Person”
- General restrictions have been placed on access to data and use of insights

High Control environment suitable for:

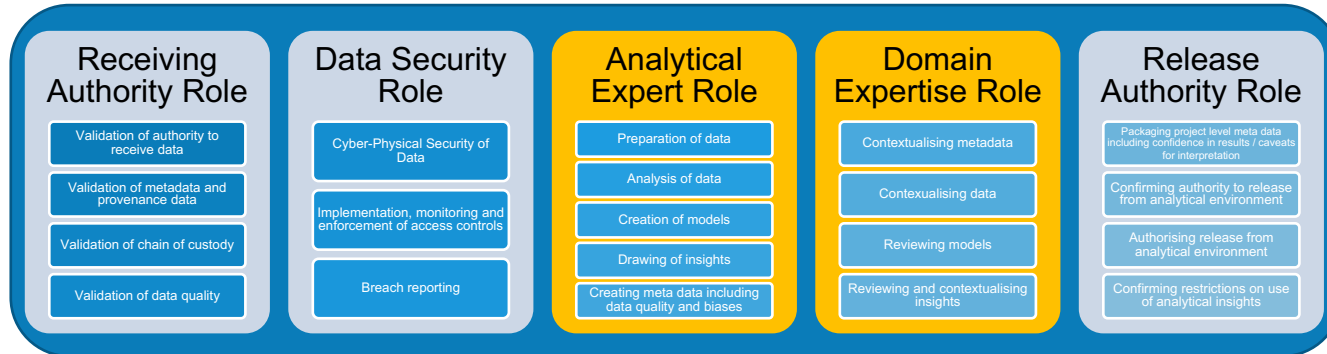
- Data which is **not** reasonably personally identifiable
- Data which contains sensitive subject matter
- Data quality which is well quantified
- People with access have met General requirements for a “Safe Person”
- Specific restrictions have been placed on access to data and use of insights, as well as release mechanism for insights

Very high Control environment required for:

- Data which can only be accessed under an external instrument such as a Public Interest Direction (PID),
- Data which **is** reasonably personally identifiable
- Data which contains sensitive subject matter
- Data quality which is well quantified
- People with access have met General requirements and Project specific requirements for a “Safe Person”
- Specific restrictions have been placed on access to data and use of insights, as well as release mechanism for insights



Requirements for “Safe Person” for High Control and Very High Control Environments



The General requirements for a “Safe Person” to work in a project is someone who is

- verifiably skilled and experienced in their domain’s techniques – e.g. analytical expert, governance expert, cyber expert
- Screened and/or endorsed by independent authorities – endorsed by executive manager, completed police check or working with children check (if necessary)
- Understands and agrees to be bound by relevant legal frameworks – including PIPPA, HRIPA
- Understands and agrees to follow formal governance processes used in the analytical environment
- Understands the roles of others in the analytical chain / governance process, and agrees to respect roles / work with these other roles
- Understands and is able to use the specific tools and processes in the analytical environment

Project specific Requirements

- Is expressly authorised to work with the subject data for an authorised project
- Understands and agrees to be bound by project legal agreements or restrictions – e.g. PID, other project specific restrictions

Individual Privacy Considerations

- Personal Connection to the dataset – understanding the degree of separation between the people represented in the dataset, or the region represented, and the analyst.
- Accountability – the personal consequences for the analyst in the event that reidentification does occur (personally identifiable information (PII) is attained), PII is released, or that PII is used inappropriately by the analyst.



Glossary

AI – means is intelligent technology, programs and the use of advanced computing algorithms that can augment decision making by identifying meaningful patterns in data.

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 – refers to a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods

Data Lifecycle –refers to the entire period of time that data exists in your system. This life cycle encompasses all the stages that your data goes through, from first capture onward.

Data Quality – 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.

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

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.

Glossary

Non-operational AI – systems do not use a live environment for their source data. Most frequently, they produce analysis and insight from historical data.

Operational AI – are those 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 AI systems 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 DCJ – “A child or young person is at risk of significant harm if the circumstances that are causing concern for the safety, welfare or well being of the child or young person are present to a significant extent.

Useful resources



Resource 1A – Policies, Guides and Frameworks

Relevant strategy, policies and guides

- [AI Strategy](#)
- [Digital policy landscape](#)
- [AI Ethics Policy](#)
- [Cybersecurity policy](#)

Project governance

- [ICT Assurance](#)
- [Benefits Realisation Management Framework](#)
- [Digital restart fund](#)

Privacy

- [NSW Information and Privacy Commission](#)
- [Privacy by design](#)

Further Information on the Personal Information Factor (PIF) Tool

The PIF for dataset is driven by both the minimum identifiable cohort size (MICS) and the amount of information which would be revealed if individuals in this cohort were reidentified. PIF Tool Demonstration video available at <https://www.youtube.com/watch?v=wrD6FI2U4Rs>

An open source PIF Tool is available at <https://github.com/PIFtools/piflib>

More detail on the PIF tool
[2019 ACS Report, Privacy Preserving Data Sharing Frameworks](#)



Contact

For more information contact the NSW Data Analytics Centre:

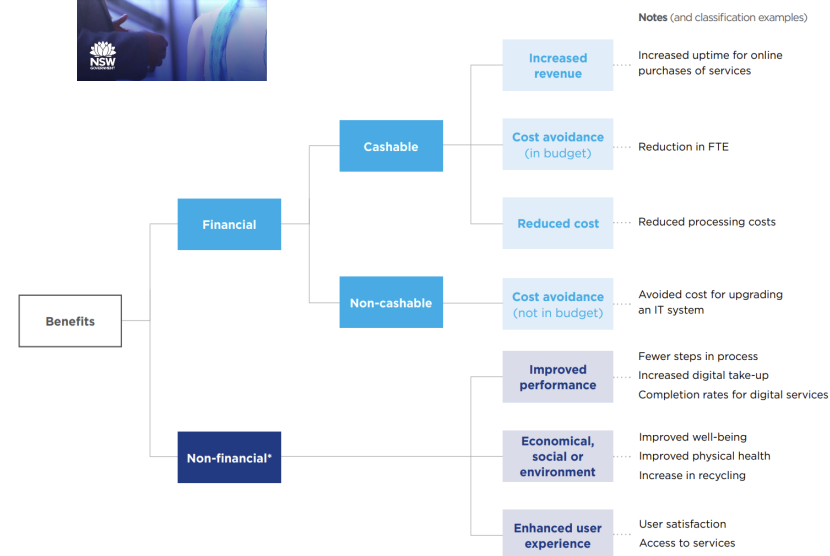
- datansw@customerservice.nsw.gov.au

Resource 1B – Benefits Realisation Framework

Community Benefit from the Use of AI Systems

Governance is key to implementing benefits management, as benefits need to be owned by appropriate sponsors and managers from within the organisation. To support active program sponsorship at the senior leadership and executive level:

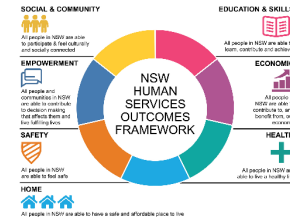
- develop a program vision statement, to be promoted by senior leadership, to assist with the transformational change required to realise the program benefits.
- review the underlining principles of benefits realisation management outlined in *Part 1* of the **NSW Benefits Realisation Management Framework** and how these principles support each phase of benefits management in the governance section of *Part 3: Guidelines*
- use benefits management deliverables to clearly articulate the program outcomes and intended benefits
- when possible, manage, report and approve benefit deliverables within existing governance meetings, noting that the size, complexity, priority and risk of a program and its benefits will affect the level of governance required to control its delivery and benefit realisation
- when possible, integrate benefits management processes with other business processes or NSW Government frameworks used within the organisation
- use the 'RACI' (Responsible, Accountable, Consulted Informed) in *Part 3: Guidelines* to review and agree on responsibilities for managing and realising benefits




*These are examples and should be tailored to the program/project environment

Resource 1C - 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 

Hypothesis	Stakeholders	Desired Outcomes	Benefits
Key Questions	Data Available		
Background/Problem		Current Metrics	Value derived from project

Resource 1D – Co-design Example

Community Benefit from the Use of AI Systems

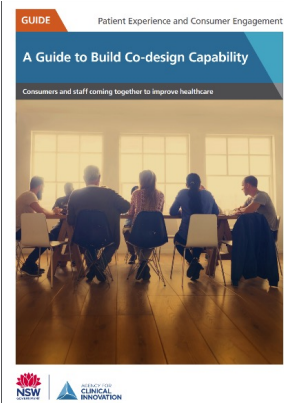
Co-design is a way of bringing major stakeholders together to improve services. It creates an equal and reciprocal relationship between all stakeholders, enabling them to design and deliver services in partnership with each other.

Planning, designing and producing services with people that have experience of the problem or service means the final system is more likely to meet their needs.

This way of working demonstrates a shift from seeking involvement or participation after an agenda has already been set, to seeking consumer leadership from the outset so that consumers are involved in defining the problem and designing the system.

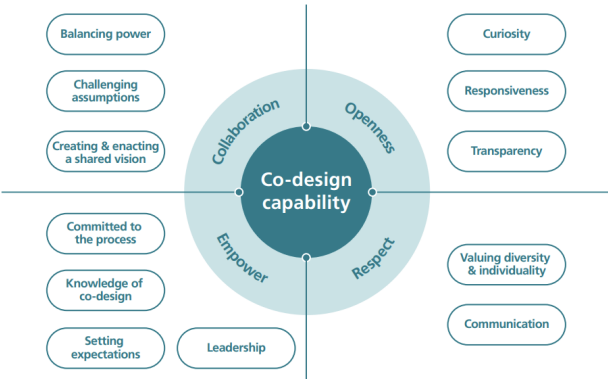
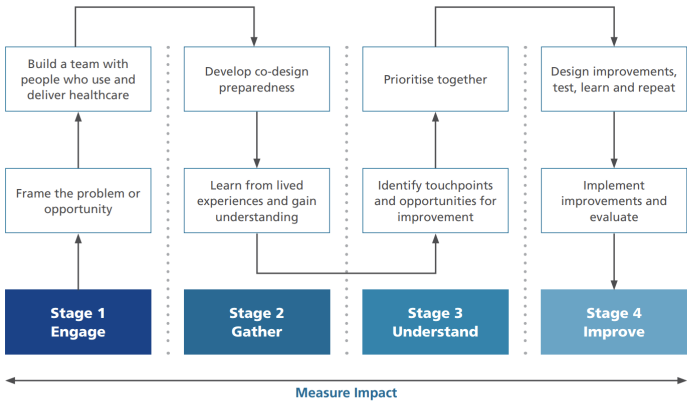
Co-design typically uses a staged process that adopts participatory and narrative methods to understand the experiences of receiving and delivering services, followed by consumers and health professionals co-designing improvements collaboratively.

An example is available from the NSW Agency for Clinical Innovation via the link below.



https://aci.health.nsw.gov.au/data/assets/pdf_file/0013/502240/Guide-Build-Codesign-Capability.pdf

Figure 1. The co-design process



Resource 2 - Recommended Harm Mitigation Approaches

Harm Type	Ethics Expert Review of AI System	Policy Domain Expert Review of AI System	Data Governance / Cyber Security Focus	Analytical Expert Review of AI System	Co-Design of project / actions
Physical	X	X			X
Psychological	X	X			X
Unauthorised Use of Health / Sensitive Personal Information			X	X	X
Unauthorised Use of Personal Information			X	X	X
Impact on Right, Privilege or Entitlement	X	X			X
Misidentification of Individual				X	X
Misapplication of Penalty / Fine		X		X	X
Other Financial Impact		X		X	X
Incorrect guidance / advice				X	X
Inconvenience, Delay				X	X
Other Harms	X	X	X	X	X

Resource 3 - Existing Developing Standards Families

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

Resource 3 - Developing Standards for AI ISO/IEC/JTC1 SC42

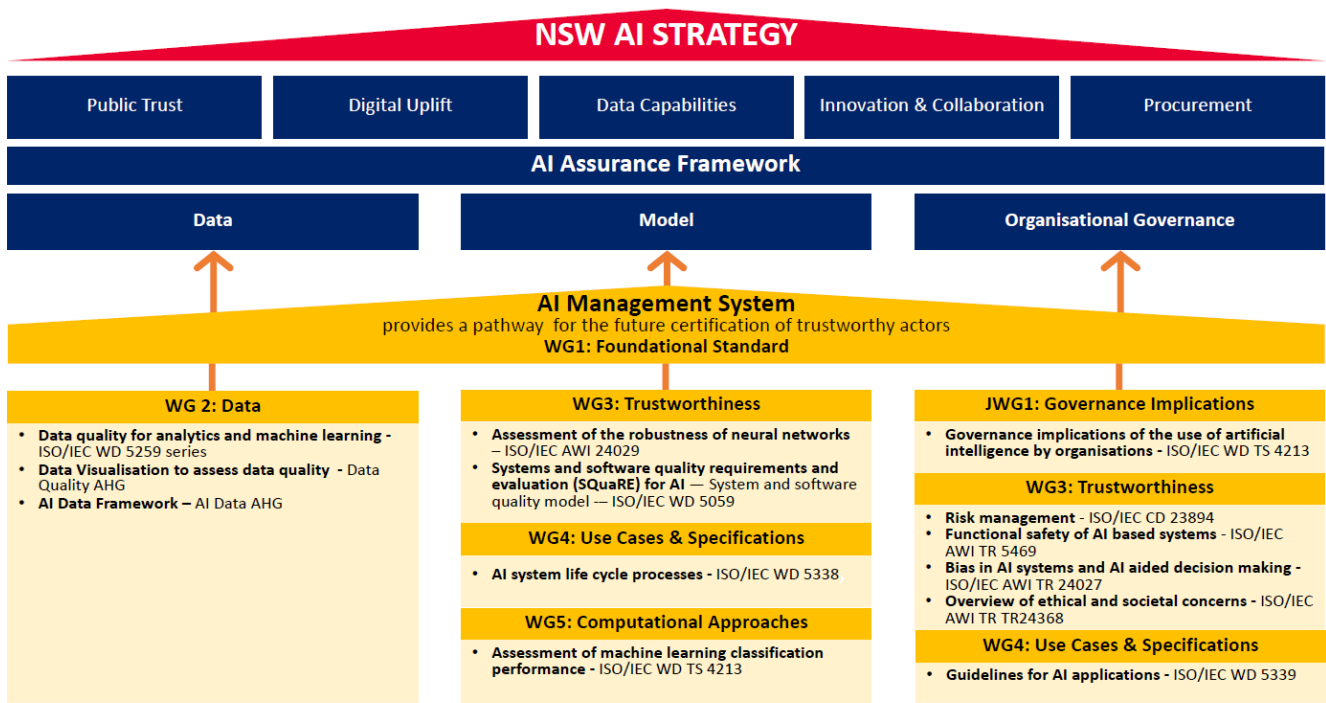


SC 42 is developing an **AI Management System as a pathway to certification**, leveraging the work that has been conducted under all the working groups.

5 standards are now published, and 21 standards and projects are under development.

Including observers, currently 50 countries involved.

Resource 3 - Aligning Standards with NSW AI Strategy



SC 42 work is complemented by
SC32 WG6 data sharing and use
SC32 WG2 metadata
SC 40 Information management



Resource 3 - Developing International Standards for AI

WG1: Foundational Standards

ISO/IEC WD 42001 - AI - Management system (AIMS)

This document enables an organisation to show that it has implemented and continually iterates on the improvement of processes unique to the development or use of an AI system.

For example processes identifying and treating bias of learning data, fairness, inclusiveness, safety, security, privacy, accountability, explainability and transparency, etc.

The risk-based nature of the MSS for AI ensures that the organization carefully considers the impacts and risks and how these can be addressed through the development or procurement of AI systems.

This standard is intended to create a pathway for the future certification of trustworthy AI actors.

WG3: Trustworthiness

ISO/IEC CD 23894 - Risk Management (based on ISO 31000), provides guidelines on managing risk during the development and application of AI techniques and systems. It describes processes for the effective implementation and integration of AI risk management.

ISO/IEC AWI TR 5469 - Functional safety of AI based systems, classifies AI applications and technology, and investigates the extent to which their use can be trusted in safety related applications. This includes the properties and risk-factors, verification and validation techniques and the control and mitigation methods that might be applied.

ISO/IEC AWI 24029- Assessment of the robustness of neural networks

- Part 1: Overview of existing methods to assess the robustness of neural networks(NN), and the different levels of testing that could inform risk assessment for NN
- Part 2: Methodology for the use of formal method, provides a methodology for the use of formal methods to assess robustness properties of NN (how to select, apply and manage formal methods to prove robustness properties).

ISO/IEC AWI TR 24027 - Bias in AI systems and AI aided decision making, provides guidance on the types and forms of bias in AI systems, measurement techniques and methods for assessing and mitigating bias.

ISO/IEC WD 5059 Systems and software quality requirements and evaluation (SQuaRE) for AI - System and software quality models, outlines a quality model for AI systems, and is an AI -specific extension of the SQuaRE series (ISO/IEC 25010 and ISO/IEC 24029). Its purpose is to set out quality characteristics against which stated quality requirements can be compared.

Proposed NWIP - (SQuaRE) for AI – Quality Assurance Process Framework for AI Systems, is based on WD 5059 above. The framework includes requirements for a quality strategy, guidance on how to specify acceptable thresholds, and references materials on how to verify and validate AI systems.

ISO/IEC AWI TR 24368 - Overview of ethical and societal concerns provides information in relation to principles, processes and methods in this area.

Resource 3 - Developing International Standards for AI

WG2: Data

ISO/IEC WD 5259 - Data quality for analytics and machine learning

- Part 1: Overview, terminology, and examples.
- Part 2: Data quality measures, identifies characteristics and measurements of data quality.
- Part 3: Data quality management requirements and guidelines, specify requirements and guidelines for managing and improving data quality;
This standard will enable certification of data quality for analytics and ML.
- Part 4: Data quality process framework, provides guidance on organisational approaches to data quality for training and evaluation in analytics and machine learning.

Data Quality AHG is developing NWIPs on

- Part 5: Data quality assurance and Part 6: Data quality governance for WD 5259 above; and
- Data visualisation to assess data quality by following a data flow model.

AI Data AHG is developing a NWIP: AI Data Framework

AI Data is defined by the data input to, used by, and output from AI systems. This document describes:

- AI Data terminology, types, concepts and how AI Data is organised into datasets
- AI Data functional view including modelling, representation methods, description methods ect,
- AI Data transformation processes and management issues including how to preserve AI Data quality and history.
- AI Data ecosystem including translation languages, formats and analysis

Resource 4 - Relevant Standards for Data Sharing and Use

Publisher	Designation	Title
ISO	19944-2	Cloud and distributed platforms - Cloud services and devices: data flow, data categories and data use - Part 2: Guidance on application and extensibility
ISO	19944-1:2020	Cloud and distributed platforms - Data flow, data categories and data use - Part 1: Fundamentals
ISO	15489-1: 2016	Information And Documentation - Records Management - Part 1: Concepts And Principles
BSI	BS 30301	Information And Documentation. Management Systems For Records. Requirements (British Standard)
ISO	38500:2016	Information technology - Governance of IT for the organisation
ISO	24368	Information technology — Artificial intelligence — Overview of ethical and societal concerns
ISO	24668	Information technology — Artificial intelligence — Process management framework for Big data analytics
ISO	20546	Information Technology - Big data - Overview and vocabulary
ISO/IEC	20547-3:2020	Information technology — Big data reference architecture — Part 3: Reference architecture
ISO	20547-1	Information Technology - Big data reference architecture Part 1: Framework and application process
ISO	20547-4	Information Technology - Big data reference architecture Part 4: Security and privacy
ISO	14662:2010	Information technology - Business Operational view - Part 8 Identification of privacy protection requirements as external constraints on business transactions
ISO	15944-1:2011	Information technology - Business Operational View - Part1: Operational aspects of Open-edi for implementation
ISO	15944-12:2020	Information technology - Business operational View - Part12: Privacy protection requirements (PPR) on information lifecycle management (ILCM) and EDI of personal information (PI)
ISO	22624	Information technology — Cloud computing — Taxonomy based data handling for cloud services
ISO	19583-23	Information Technology - Concepts and usage of metadata - Data element exchange (DEX) for 11179-3
ISO	19583-1	Information Technology - Concepts and usage of metadata - Part 1: Metadata concepts
ISO	38508	Information technology - Governance of IT - Governance of data - Guidelines for data classification
ISO/IEC	38505-1:2017	Information technology - Governance of IT - Governance of data - Part 1: Application of ISO/IEC 38500 to the governance of data
ISO	38505-2:2018	Information technology - Governance of IT - Governance of data - Part 2: Implications of ISO 38505-1 for data management
ISO	11179-3:2013	Information technology - Metadata registries (MDR)- Part 3: Registry metamodel and basic attributes
ISO	14662:2010	Information technology - Open-edi reference model
ISO	27001:2013	Information technology - Security techniques - Information security management systems - Requirements
ISO	27550:2019	Information technology — Security techniques — Privacy engineering for system life cycle processes
ISO	27701:2019	Security techniques - Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management - Requirements and guidelines
BSI	PAS 183:2017	Smart Cities - Guide to establishing a decision making framework for sharing data and information services
ISO/IEC	10032:2003	Information technology — Reference Model of Data Management
ISO/IEC	11179-1:2015	Information technology — Metadata registries (MDR) — Part 1: Framework
ISO/IEC	11179-2:2019	Information technology — Metadata registries (MDR) — Part 2: Classification
ISO/IEC	11179-3:2013	Information technology — Metadata registries (MDR) — Part 3: Registry metamodel and basic attributes

Resource 5 – Considerations (Risk Factors) for Data Use

Sensitivities about data itself:

1. Concerns that data contains high levels of personal information
2. Concerns that data contains uniquely identifiable individuals
3. Concerns that sensitive subjects are captured in data (culturally subjective but often described e.g. religion)
4. Concerns about data quality (accuracy, timeliness, completeness, and consistency)
5. Concerns about fitness-for-purpose of data for analysis

Sensitivities about capability and governance:

6. Concerns that context is not captured with data (metadata, provenance, consent)
7. Concerns about authority to share data for analysis
8. Concerns about poor governance or accidental release of data or insights (outputs)
9. Concerns that expert knowledge / context is required to appropriately interpret data and results of analysis
10. Concerns about authority to release results of analysis

Sensitivities about use of insights:

11. Concerns about the level confidence in outputs (accuracy, precision, consistency, explainability, bias)
12. Concerns about unintended consequences from how outputs (insights / data driven decisions) will be used
13. concerns about whether human judgement will be applied before an insight becomes a decision
14. Concerns possible harms resulting from use of outputs (reversible, reversible with cost, irreversible)
15. Concerns that results from analysis may lead to negative surprises (especially for data not analysed before)
16. Concerns that commercial value may be degraded if insights are shared

L M H



PIF: L M H

Inherent Sensitivity: L M H