Australian Government
Digital Transformation Agency

dta

Assessment against the Pilot Australian Government AI Assurance Framework

Use of Artificial Intelligence in Digital Marketplace Panel 2 tender evaluations

Version 1.0

Digital Transformation Agency



© Commonwealth of Australia (Digital Transformation Agency) 2019

With the exception of the Commonwealth Coat of Arms and where otherwise noted, this product is provided under a Creative Commons Attribution 4.0 International Licence. (http://creativecommons.org/licenses/by/4.0/legalcode)

The Digital Transformation Agency has tried to make the information in this product as accurate as possible. However, it does not guarantee that the information is totally accurate or complete. Therefore, you should not solely rely on this information when making a commercial decision.

Digital Transformation Agency is committed to providing web accessible content wherever possible. If you are having difficulties with accessing this document, please email communications@dta.gov.au.

Version: 1801

Contents

Introduction **3**
Executive Summary 4
1 Basic information **5**
1.1 AI use case profile 6
1.2 Lifecycle stage 7
1.3 Review date 7
1.4 Assessment review history 8
2 Purpose and expected benefits **9**
2.1 Problem definition 10
2.2 AI use case purpose 10
2.3 Non-AI alternatives 11
2.4 Identified stakeholders 11
2.5 Expected benefits 12
3 Threshold assessment **13**
Assessment contact officer recommendation *19*
Executive sponsor endorsement *20*
4 Fairness **20**
4.1 Defining fairness 21
4.2 Measuring fairness 22
5 Reliability and Safety **23**
5.1 Data suitability 23
5.2 Indigenous data 24
5.3 Suitability of procured AI model 24
5.4 Testing 25
5.5 Pilot 26
5.6 Monitoring 27
5.7 Preparedness to intervene or disengage 27
6 Privacy protection and security **28**
6.1 Minimise and protect personal information 29
6.2 Privacy assessment 29
6.3 Authority to operate 30
7 Transparency and explainability **30**
7.1 Consultation 31
7.2 Public visibility 31
7.3 Maintain appropriate documentation and records 32
7.4 Disclosing AI interactions and outputs 33
7.5 Offer appropriate explanations 33

8 Contestability 34
8.1 Notification of AI affecting rights 35
8.2 Challenging administrative actions influenced by AI 35
9 Accountability 36
9.1 Establishing responsibilities 37
9.2 Training of AI system operators 37
10 Human-centred values 38
10.1 Incorporating diversity 39
10.2 Human rights obligations 39
11 Internal review and next steps 40
11.1 Legal review of AI use case 41
11.2 Risk summary table 41
11.3 Internal review of AI use case 42
11.4 External review of AI use case 43

Introduction

This document presents the formal assessment of the Use of AI in Digital Marketplace Panel 2 (DMP2), conducted through the Pilot Australian Government AI Assurance Framework. Developed by the Digital Transformation Agency (DTA), this assessment evaluates the design, deployment, and operation of the DMP2 AI tool, ensuring alignment with Australia's AI Ethics Principles and emerging national best practice for responsible AI in government. The Framework provides a structured, risk-based methodology to guide agencies in the ethical, lawful, and effective use of AI, with a focus on transparency, human oversight, and continuous improvement.

Executive Summary

Purpose of Assessment:

This document details the comprehensive evaluation of the DMP2 AI tool, as required by the Pilot Australian Government AI Assurance Framework. The assessment covers the tool's objectives, operational context, risk profile, and governance arrangements, ensuring that its use in government tender evaluations meets the highest standards of fairness, accountability, and public trust.

Project Overview:

The DMP2 AI tool is designed to augment the evaluation of vendor applications for digital services, automating repetitive tasks and supporting human evaluators with AI-generated recommendations. The tool operates within a secure, government-approved environment, with human oversight ensuring that all final decisions remain with evaluators.

Key Findings:

- The AI tool improves efficiency, consistency, and quality in tender evaluations, enabling more frequent panel refreshes and broader market access.

- Human-in-the-loop processes, privacy controls, and external expert reviews (e.g., CSIRO Data61) are embedded to manage risks and uphold ethical standards.
- Medium risks (e.g., security, reputational impact) are accepted with ongoing monitoring and periodic review; all other risks are assessed as low.
- Stakeholder consultation, transparency mechanisms, and contestability processes are in place to ensure accountability and build trust.

Governance and Continuous Improvement:

The assessment confirms robust governance, clear roles and responsibilities, and ongoing training for operators. The DMP2 AI tool and its supporting framework are subject to regular review and refinement, informed by stakeholder feedback and lessons learned from deployment.

In Summary:

This document is the official assessment of the AI DMP2 tool project, conducted under the Pilot Australian Government AI Assurance Framework. It demonstrates the DTA's commitment to responsible, transparent, and effective AI adoption in government procurement, setting a benchmark for ethical innovation and public sector leadership.

1 Basic information

1.1 AI use case profile

Complete the information below.

Name of AI use case	Digital Marketplace Panel 2 AI in tender evaluations, Digital Sourcing branch
Reference number	DTA240719
Lead agency	DTA
Assessment contact officer	**Name:** Section 22 **Email:** Section 22
Executive sponsors	**Name:** **Chris Fechner** **Chief Executive Officer** **Digital Transformation Agency** **Email: Chris.Fechner@dta.gov.au**

In plain language, briefly explain how you are using or intend to use AI. 200 words or less.

AI use case description
The DTA aims to drive the Government's digital transformation by setting up and managing whole-of-government sourcing arrangements, including the Digital Marketplace Panel 2 (DMP2). The DTA regularly invites the market to tender digital services to join the DMP2, whereby the DTA evaluates applicants to determine whether they are value for money. This process is structured in several phases to ensure effective, efficient, economical and ethical evaluations. The key phases are: • screening of applications to confirm compliance with conditions of participation, minimum content and format requirements, the Panel Agreement, and to clarify unintentional errors • evaluation of a company's financial viability, understanding and application of government procurement connected policies, technical capability and capacity, and pricing of goods and services offered • risk assessment and value for money recommendation Applications are assessed by an evaluation committee (EC), which has identified challenges in the evaluation process that AI could help address, such as: • the scale and frequency of the tender cycle • the repetitive and labour-intensive nature of the process • limited human resources, and • the need for consistency, fairness, and transparency.

The EC proposes using AI to help assess the technical capability and capacity evaluation phase. The AI will recommend ratings based on the evaluation of case studies using the evaluation methodology and criteria set out in the relevant Tender Evaluation Plan (TEP).
Independently, a human will also evaluate the applicant's technical capability and capacity using the same methodology and criteria. If the AI's rating differs from the human evaluator's rating outside an agreed margin, the application will be reviewed by an additional and independent human to reach a consensus score and commentary.
The EC will still be responsible for the final evaluation outcomes. AI will not make any decision.
This use case will use data from DMP2 tenders, with permission from applicants, to test AI functionality and applicability for evaluations. The DTA will apply lessons learnt in this AI use case to other use cases in the DTA.

Briefly explain what type of AI technology you are using or intend to use. 100 words or less.

Type of AI technology

The Digital Marketplace AI Augmentation platform uses advanced artificial intelligence to help automate and improve the assessment of vendor applications. At its core, the platform uses Microsoft's Azure OpenAI service, which provides secure access to powerful large language models (LLMs) similar to those behind ChatGPT.
These models can analyse, summarise, and compare large volumes of text, supporting faster and more consistent decision-making. The system is designed to be secure, scalable, and compliant with government standards, and can be adapted to use other AI models in the future if needed.

1.2 Lifecycle stage

These stages can take place in an iterative manner and are not necessarily sequential.
They are adapted from the OECD's definition of the AI system lifecycle. Refer to Guidance for further information. Select only one.

Which of the following lifecycle stages best describes the current stage of your AI use case?

Early experimentation	☐
Note: Assessment not required	
Design, data and models	☐
Verification & validation	☐
Pre-Deployment / Deployment	☒
Operation and monitoring	☐
Retirement	☐

1.3 Review date

Assessments must be reviewed when use cases either move to a different stage of their lifecycle or significant changes occur to the scope, function or operational context of the use case. Consult the Guidance and, if in doubt, consult the DTA.

Next date/milestone that will trigger the next review of the AI use case

Review after deployment in real-word evaluations.

1.4 Assessment review history

Summary of changes arising from review	Review date
DTA conducted early experiments on the potential to use AI to support human evaluations of tenders. Synthetic data is used to replicate previous tender responses to simulate a common approach to technical aspects of a DTA tender evaluation. Early indicators are encouraging for AI's potential use but also point to shortfalls in the technology and data available to demonstrate its capability to augment human evaluators.	June 2024
After the early experimentation, the CTO team enhanced the prompt control for more consistent and accurate AI outputs. The team refines building blocks of earlier trials to achieve outputs that are closer to human evaluation outcomes.	11 July 2024
In preparation for designs, data and models, the team adopted this AI Assurance Framework as guiding artefact for a risk-based approach.	15 August 2024
DTA sought written consent from applicants of the DMP2.0 Mar-24 panel establishment round to access technical capability and capacity responses (i.e. case study data) to support future AI trials for tender evaluations.	26 August 2024
With express agreement from applicants, ingest a relatively small dataset from DMP2 applications for the purpose of further validating AI outputs against known human outcomes.	September 2024
Engage external expertise to support an in-depth analysis of the DTA's approach, technology, risks, controls, governance and ethics of applying AI to support tender evaluation processes.	October 2024
In the approach to market for the DMP2.1 Nov-24 panel refresh round, DTA included the following condition in the Request for Tender (RFT): By submitting an Application, the Applicant consents to: a. DTA's use of the AI Tool in the evaluation of their Application; and b. DTA's use of their Application for training and validation of the AI Tool in this and future Application Rounds. The Applicant may withdraw their Application at any stage of the procurement process if they no longer consent to DTA's use of the AI Tool.	6 February 2025
External review report received from CSIRO's Data61 and incorporated into planning. Overall, report outcome was positive.	April 2025
Supporting activities, documentation and approvals progressed, including: • Governance Framework • Assurance Framework (this document) • Risk Management Framework • Project Plan – Adoption of AI in DTA Tender Processes • Authority to Operate (ATO) • Privacy Impact Assessment (PIA) and external probity review • Relevant procurement documentation (per tender): • Procurement Plan	June-September 2025

• Probity Plan • Tender Evaluation Plan (TEP) • Request for Tender (RFT) and associated BuyICT application workflow • Evaluation Standard Operating Procedures (SOPs)	
Prompt engineering, A/B testing and analysis of AI candidates	June-September 2025
AI Candidate E selected as the most suitable	September 2025

2 Purpose and expected benefits

Under *Australia's AI Ethics Principles*, the use of AI should have a clearly defined and beneficial purpose that is consistent with human, societal and environmental wellbeing.

2.1 Problem definition

Use 100 words or less.

Clearly and concisely identify the problem you are trying to solve

To improve the efficiency, effectiveness, and consistency of the tender evaluation process, whilst minimising the high resourcing demands incurred.
The highly repetitive, standardised, and labour-intensive nature of evaluations requires significant time and human effort, leading to increased costs and the potential for evaluator fatigue. This fatigue, combined with the scale and frequency of refresh cycles, may result in cognitive and communication biases, inconsistencies, and errors which could result in evaluator 'burnout' and compromise the fairness and accuracy of the process.

2.2 AI use case purpose

Use 200 words or less.

Clearly and concisely describe the purpose of your use of AI, focusing on how it will address the problem you have identified

The purpose is to deliver three benefits to future DMP2 tender and evaluation processes.
In no particular order, these are:

1. Improve the quality of evaluation, by
    2. providing consistent evaluations of applicant case studies based on formal evaluation criteria.
    3. alleviate reading fatigue / assessment fatigue experienced by evaluators.
    4. improve the consistency of evaluation, by use of a single LLM with the same knowledge base, the same prompts, and the same insights across the entire set of applications.
5. Improve the efficiency of evaluation, reducing the time and resourcing requirements needed to complete evaluations across multiple current, and future modules and categories, by
    - automating repetitive and time-consuming tasks, allowing evaluators to focus on more nuanced and complex aspects of the process.
    - analysing large volumes of data, identifying patterns, and generating preliminary recommendations to reduce impact to human resourcing.

6. Potentially increasing the frequency of tender refresh cycles to provide the Commonwealth with greater access to the market, including the addition of new sellers and innovative technology, whilst maintaining healthy competition.

2.3 Non-AI alternatives

Non-AI alternatives that could address this problem
The non-AI alternative is the status quo. Two evaluators will individually evaluate each application, without any AI assistance, before coming together to reach a consensus score for each application. It may suffer from the problems outlined above (inconsistency between pairs of evaluators; assessment fatigue; delays to completion; lack of sufficient technical knowledge), but without any additional technological solution, these problems would be considered inherent to the process.

2.4 Identified stakeholders

Stakeholder group	Briefly describe how they may be affected
Applicants	• positive: efficient evaluations that result in frequent Tender refreshes • positive: input from the AI system may increase consistency in the evaluation removing incorrect assessment outcomes • negative: perception that if unsuccessful, this is due to AI regardless that the DTA will discard AI results if the rating deems the applicant 'unsuitable'
Buyers	• positive: benefit from improved timeframes for completion of evaluation • positive: benefit from improved quality and consistency of sellers on DMP2 • positive: more efficient procurement processes could lead to a reduction in cost recovered fees for accessing DTA procurement arrangements • positive: increase frequency in tender refresh resulting greater access to the market • negative: a possible reduction in trust of DTA vetting processes leading to buyers looking elsewhere to procure digital goods and services, reducing whole of government efficiencies
Evaluators	• positive: automation of repetitive and manual time-consuming tasks

	• positive: capability uplift with real-world exposure to AI • positive: allow evaluators to focus on more complex tasks and priorities
The broader APS	• positive: visibility about the outcome of this use case, and associated application of policy and framework, successful or otherwise • negative: poor implementation or checks and balances could lead to an aversion to implement AI in other areas of procurement or government generally and cause reputational damage to the agency and APS.

2.5 Expected benefits

What are the expected benefits of the AI use case?

The expected benefits are to:

- Improve the quality of evaluation
- Improve the efficiency of evaluation, reducing the time and resourcing requirements needed to complete evaluations across multiple current, and future modules and categories
- Increase the frequency of tender refresh cycles to provide the Commonwealth with greater access to the market, including the addition of new sellers and innovative technology, whilst maintaining healthy competition.

The expected benefits for each stakeholder group specifically are:

Applicants:

- Efficient completion of evaluations that result in increased opportunity to engage with government through frequent refresh cycles.
- input from the AI system may improve consistency in the evaluation by removing incorrect assessment outcomes.

Buyers:

- Improved timeframes for completion of evaluation
- Improved quality and consistency of sellers on DMP2
- Increase frequency in tender refresh resulting greater access to the market.
- More efficient procurement processes could lead to a reduction in cost recovered fees for accessing DTA procurement arrangements

Evaluators:

- Capability uplift through real-world experience in using AI
- Automation of repetitive, manual, and time-consuming tasks
- Allow evaluators to focus on more complex tasks and priorities.

The DTA & wider APS:

- Benefit from learning and insights about the outcome of this use case, and associated application of policy and framework, successful or otherwise

3 Threshold assessment

Risk Matrix					
Likelihood/Consequence	Insignificant	Minor	Moderate	Major	Severe
Almost certain	Medium	Medium	High	High	High
Likely	Medium	Medium	Medium	High	High
Possible	Low	Medium	Medium	High	High
Unlikely	Low	Low	Medium	Medium	High
Rare	Low	Low	Low	Medium	Medium

What is the risk of the use of AI	Risk severity
7. Negatively affecting public accessibility or inclusivity of government services	Low ☒ Med ☐ High ☐

The public may perceive that the use of AI could negatively affect their accessibility, or inclusivity in a tender, due to:

- Bias where the AI perpetuates or amplifies an existing bias in the data that affects a specific cohort, such as SME's or indigenous businesses.

This could have reputational risks to the DTA, via a lack of confidence in the tender process and how the panels and marketplaces have been established. It may disenfranchise certain market participants, and result in a lower number of sellers and diminished choice and competition in marketplaces.

The likelihood/consequence is rare/moderate. Existing risk controls in place are:

- DTA's Contract Limits and Reviews Policy increases market access for businesses of all sizes and enables flexibility to respond to changing strategic direction and emerging technologies;
- DMP2 Tender Evaluation Plan (TEP) ensures that the public has an equal and fair opportunity to apply via BuyICT.

8. Unfairly discriminating against individuals, communities or groups	Low ☒ Med ☐High ☐

The public may perceive that they might be unfairly discriminated against using AI in tender evaluations, resulting from:

- Transparency in the AI decision process.

- Perception that human oversight may be perceived as a formality.

The likelihood/consequence is unlikely/moderate. The existing risk controls are:

- DTA has established complaints and dispute resolution processes which requires the DTA to respond to applicant concerns in a timely, fair, and transparent manner.
- The DTA collaborates with the Commonwealth Ombudsman and the Australian Small Business and Family Enterprise Ombudsman to handle discrimination.
- The TEP ensures a fair, equitable and non-discriminatory treatment of Applicants across RFT rounds.
- Any variations to the TEP will be considered by the Chair in consultation with the Legal, Procurement, and/or Probity Adviser.

9. Perpetuating stereotyping or demeaning representations of individuals, communities or groups — Low ☒ Med ☐High ☐

AI use in a tender could perpetuate stereotyping or demeaning representations of individuals, communities, or groups because of:

- Algorithmic decision making relying on data attributes that could lead to a bias in decision making if the AI prioritises criteria indirectly/directly linked to certain elements of a tender, for instance SMEs or indigenous businesses.

The likelihood/consequence is rare/insignificant.

The TEP and proposed use case have safeguards and controls in place, including:

- At minimum, a human evaluator will form part of the evaluation pair, to reduce outcome bias.
- Where a 'dissenting' view occurs and there is no consensus in scoring between the human evaluator and AI, a second and independent human evaluator will review to form consensus.
- Chair and the Delegate will have oversight of the tender process to ensure compliance, including the final decision on Applicant suitability for panel inclusion.

10. Harming individuals, communities, groups, organisations or the environment — Low ☒ Med ☐High ☐

The public may perceive that the use of AI in tender evaluation may harm individuals, communities, groups, organisations or the environment. The likelihood/consequence is rare/insignificant.

- There is a risk of undue harm to businesses, where the use of AI influences evaluators in a way that results in a negative overall assessment. This may occur because of flawed,

but convincing reasoning presented by the AI. Where it occurs at a certain low level, and if the AI's reasoning is incorrect yet convincing, moderate harm may result, including business facing operational challenges or financial losses, affecting their competitiveness.

This is unlikely as the system is designed to improve the quality of the assessors work by providing additional perspectives, and because assessors will be encouraged to keep a critical eye on the outputs of AI analysis only after undertaking their own assessment prior to accessing AI evaluation data.

11. Raising privacy concerns due to the sensitivity, amount or source of the data being used by an AI system	Low ☒ Med ☐ High ☐

The application data may contain Personally Identifiable Information (PII), that can be used to uncover the applicant's identity. The likelihood/consequence is rare/moderate. Existing risk controls in place are:

- Evaluators to scan for PII while reading and provide risk rating for further actions.
- Redact PII before Seller's BuyICT profile is published.

Planned risk controls include:

- Conduct a privacy threshold assessment;
- Conduct a privacy impact assessment, dependant on the outcome of the privacy threshold assessment.

Assumptions affecting risk assessment:

- a broad range of information, or an opinion, that could identify an individual in accordance with Privacy Act 1988.

12. Raising security concerns due to the sensitivity or classification of the data being used by an AI system	Low ☐ Med ☒ High ☐

The application data may contain confidential and/or sensitive information, that is critical to business operations. The likelihood/consequence is unlikely/major. Existing risk controls in place are:

- The transmission, storage and processing of the application data is contained within the DTA's protected network.
- The classification and protection of the application data is rated to Official Sensitive at the highest level in accordance with DTA's security requirements.
- During onboarding, evaluators to sign confidentiality deed and to fulfill probity requirements in accordance with Probity Plan.
- The DTA does not seek classified information from applicants in tender processes. Applicants are also notified at time of tender to carefully consider the information they share with the DTA, as the information provided may be shared publicly via BuyICT catalogues.

Planned risk controls include:

- Conduct a security risk assessment.
- (depending on the outcome), conduct a security impact assessment.

Assumptions affecting the risk assessment:

- DTA does not collect PROTECTED information in a tender process.
- PROTECTED information must not be sent outside the DTA's PROTECTED environment. If it happens, the person will receive an official breach notification from the Security team.
- The DTA has adequate processes in place to manage the protection of data in accordance with the PSPF. Depending on the circumstances, a breach would trigger events to notify relevant government agencies and affected applicants.

13. Raising security concerns due to the implementation, sourcing or characteristics of the AI system	Low ☐ Med ☒ High ☐

The security concerns that may arise during implementation, sourcing or characteristics of the AI system are about the reliability, accuracy, and fairness of the AI tool's output. The likelihood/consequence is possible/moderate. Existing risk controls include:

- The BuyICT platform allows a clear and accountable audit log by enhanced logging detail, application lists and activity logging.
- The Evaluation SharePoint has clear access control right and is managed by the Project Team.

Planned risk controls include:

- Rigorously test the AI tool via a pilot with a large enough dataset
- Finalise this Commonwealth AI Assurance framework
- Implement oversight on the evaluation processes

14. Influencing decision-making that affects individuals, communities, groups, organisations or the environment	Low ☐ Med ☒ High ☐

The concerns that the AI tool's output could influence tender evaluation decision-making are likely/moderate. These concerns include:

- Inability to fully explain human decisions augmented by AI tool's output.
- Inability to contest AI-assisted decisions.

Planned risk control is to keep record of how the AI tool has ultimately affected the final procurement decision. Applicants also have considerable rights under Commonwealth procurement frameworks to review the outcome of a DTA procurement process. The use and implementation of AI to support the evaluation process would fall within purview of any such review. The TEP outlines the measures to ensure the evaluation is fair, consistent, transparent, and in accordance with relevant Commonwealth policies and legislation.

15. Posing a reputational risk or undermining public confidence in the government	Low ☐ Med ☒ High ☐

The likelihood/consequence of reputational risks is possible/minor. Public confidence in, and reputational risks to the DTA and whole-of-Government may include:

- Lack understanding and transparency about the use of Applicant data in training the AI and quality of the outcome achieved.
- Negative media coverage due to misinformation or a misunderstanding of how the AI has been implemented.
- Criticism by groups that represent key stakeholders such as the Business Council of Australia.

Planned risk control is to provide clear information to the market, prior to the deployment of any AI tool, outlining how the application data will be used by AI tools in a tender evaluation.

The DTA will seek external expertise to validate its approach to technology, risk, governance, ethics and controls in the use of AI to support tender evaluations.

Assumptions affecting the risk assessment:

- The use case complies with the Commonwealth AI Assurance Framework.
- Risk associates with the use of AI in a controlled environment, with specific use criteria are offset by the efficiency and quality benefits to the tender evaluation process.
- The use of AI in tenders will offer a genuine benefit to the broader Commonwealth and industry.
- That the risks identified can be sufficiently mitigated through the engagement of an external provider, with the expertise to provide governance oversight and validation throughout the development of the AI tool.

Assessment contact officer recommendation

If the assessment contact officer is satisfied that all risks in the threshold assessment are low, then they may recommend that a full assessment is not needed, and that the agency accept the low risk.

If one or more risks are medium or above, then a full assessment must be completed, unless you amend the AI use scope, function or risk controls such that the assessment contact officer is satisfied that all risks in the threshold assessment are low.

You may decide not to accept the risk and not proceed with the AI use case.

Recommendation	A full assessment is necessary for this use case.
Comments (optional)	
Name & position	Section 22 Principal Technical Advisor
Date	27 October 2025

Executive sponsor endorsement

Endorsement	I have reviewed the recommendation, am satisfied by the supporting analysis and agree that a full assessment is necessary for this use case.
Comments (optional)	
Name & position	 **Chris Fechner** **Chief Executive Officer** **Digital Transformation Agency**
Date	11 December 2025

4 Fairness

Under *Australia's AI Ethics Principles*, AI systems should throughout their lifecycle be inclusive and accessible and should not involve or result in unfair discrimination against individuals, communities or groups.

4.1 Defining fairness

Where appropriate, you should consult relevant domain experts, affected parties and stakeholders to determine how to contextualise fairness for your use of AI. Consider inclusion and accessibility. Consult the Guidance document for prompts and resources to assist you.

Do you have a clear definition of what constitutes a fair outcome in the context of your use of AI?

Yes ☒	N/A ☐	No ☐

A fair outcome, in the context of the Digital Marketplace Panel 2 (DMP2) AI evaluation assistant, is one where all applicants are assessed consistently, impartially, and transparently against the published evaluation criteria, regardless of their background, size, or other attributes. Fairness means that:

- **All applicants have equal opportunity:** The AI system must not introduce or perpetuate bias or discrimination based on irrelevant attributes (e.g., business size, ownership, or other protected characteristics).
- **Decisions are based strictly on merit:** Evaluations are conducted using the same methodology and criteria for all applicants, as set out in the Tender Evaluation Plan (TEP).
- **Human oversight is maintained:** Any significant discrepancies between AI and human ratings trigger further independent review by humans to reach consensus.
- **Inclusion and accessibility are considered:** The evaluation process is designed to be accessible to all eligible applicants, including SMEs and Indigenous businesses, and to avoid barriers that could exclude or disadvantage any group.
- **Transparency and contestability:** Applicants are informed about the use of AI in the evaluation process, and there are clear mechanisms for contesting outcomes.

This definition has been developed in consultation with relevant domain experts, affected parties, and stakeholders, and is informed by the principles outlined in Australia's AI Ethics Principles and the Guidance document. The approach is further supported by external review (e.g., CSIRO Data61 report) and ongoing engagement with human evaluators ("human in the loop") to ensure fairness is contextualised and upheld throughout the lifecycle of the AI use case.

4.2 Measuring fairness

Measuring fairness is an important step in identifying and mitigating fairness risks. A wide range of metrics are available to address various concepts of fairness. Consult the Guidance document for resources to assist you.

Do you have a way of measuring (quantitatively or qualitatively) the fairness of system outcomes?		
Yes ☒	**N/A ☐**	**No ☐**

The fairness of system outcomes in the DMP2 AI evaluation assistant is measured using a combination of quantitative and qualitative approaches:

Quantitative Measures:

- **Comparison of AI and Human Evaluator Scores:** For each case study, the AI's scores are compared to the independent human evaluator's rating. A difference between the two ratings of more than an agreed value trigger further human review, ensuring that outlier decisions are investigated and consensus is reached.
- **Distribution Analysis:** Statistical analysis can be performed on the distribution of scores across applicants to identify any patterns of bias or unfairness.
- **Consistency Metrics:** The variance in scores for similar applications is measurable to confirm the AI system is applying criteria consistently.

Qualitative Measures:

- **Stakeholder Feedback:** Feedback avenues are available for applicants, evaluators, tender resources, executives and other stakeholders regarding their experience of the process and perceived fairness.
- **Case Study Review:** A sample of AI-generated comments are reviewed by tender specialists to assess whether the outcomes align with the published evaluation criteria.
- **External Review:** Independent experts (e.g., CSIRO Data61) are engaged to assess the appropriateness of the AI system and its outputs.

Controls and Oversight:

- **Human-in-the-loop:** All AI recommendations are subject to human oversight, and any significant differences between AI and human ratings are resolved through independent human review.
- **Transparency:** Applicants are informed about the use of AI and can contest outcomes if they believe the process was unfair.

These measures are designed to ensure that fairness risks are identified and mitigated, and that the evaluation process remains consistent, impartial, and transparent for all applicants.

5 Reliability and Safety

Under *Australia's AI Ethics Principles*, AI systems should throughout their lifecycle reliably operate in accordance with their intended purpose.

5.1 Data suitability

Consider data quality and factors such as accuracy, timeliness, completeness, consistency, lineage, provenance and volume.

If your AI system requires the input of data to operate, or you are training or evaluating an AI model, can you explain why the chosen data is suitable for your use case?

Yes ☒	N/A ☐	No ☐

The data chosen for the DMP2 AI evaluation assistant is suitable for the intended use case for the following reasons:

- **Accuracy:** The dataset consists of actual case study responses submitted by applicants during DMP2 tender rounds. These responses are directly relevant to the evaluation criteria and reflect real-world scenarios.
- **Timeliness:** Data is sourced from the most recent panel establishment and refresh rounds, ensuring it reflects current market conditions and applicant capabilities.
- **Completeness:** The dataset includes all required fields and supporting information as specified in the Tender Evaluation Plan (TEP), enabling comprehensive assessment of technical capability and capacity.
- **Consistency:** Data is collected and formatted according to standardised requirements set out in the DMP2 application process, reducing variability and supporting consistent evaluation.
- **Lineage and Provenance:** All data is obtained with the consent of applicants, and its origin is documented and traceable to specific tender rounds. This ensures transparency and accountability in data use.
- **Volume:** The dataset is sufficiently large to enable robust validation of AI outputs against known human evaluation outcomes, supporting reliable model performance.

Additional controls are in place to ensure data quality, including:

- Redaction of personally identifiable information (PII) prior to publication.
- Alignment with Commonwealth privacy and security standards.

These measures ensure that the data used is fit for purpose, supports reliable AI operation, and upholds the principles of fairness, transparency, and accountability.

5.2 Indigenous data

Consider whether your use of Indigenous data and AI outputs is consistent with the expectations of Indigenous people, and the Framework for Governance of Indigenous Data (GID). See definition of Indigenous data in Guidance material.

If your AI system uses Indigenous data, including where any outputs relate to Indigenous people, have you ensured that your AI use case is consistent with the Framework for Governance of Indigenous Data?

Yes ☐	N/A ☒	No ☐

The DMP2 AI evaluation assistant does not specifically use Indigenous data, nor do its outputs relate directly to Indigenous people or communities. All data used in the evaluation process is sourced from tender applications submitted to the DMP2, with the consent of applicants, and is handled in accordance with Commonwealth privacy, security, and ethical standards.

Should future use cases involve Indigenous data or outputs relating to Indigenous people, the DTA will ensure full alignment with the Framework for Governance of Indigenous Data (GID), including consultation with Indigenous stakeholders and adherence to expectations regarding data sovereignty, governance, and benefit sharing.

5.3 Suitability of procured AI model

May include multiple models or a class of models. Includes using open-source models, application programming interfaces (APIs) or otherwise sourcing or adapting models. Factors to consider are outlined in Guidance.

If you are procuring an AI model, can you explain its suitability for your use case?

Yes ☒	N/A ☐	No ☐

The AI model selected for the DMP2 evaluation assistant is suitable for the intended use case for the following reasons:

- **Model Type:** The solution uses a large language model (LLM), such as GPT-4, deployed via Microsoft Azure OpenAI Service. This model is well-suited to natural language understanding and generation tasks, which are central to evaluating written case study responses.
- **Alignment with Evaluation Criteria:** The model is configured with static prompts that reflect the evaluation methodology and criteria set out in the Tender Evaluation Plan (TEP). This ensures that AI-generated recommendations are relevant and consistent with human evaluation standards.
- **Scalability and Reliability:** The chosen model is capable of processing large volumes of text efficiently, supporting the scale and frequency of DMP2 tender cycles without compromising reliability.
- **Security and Compliance:** The model is hosted within a secure, government-approved cloud environment (Microsoft Azure), ensuring compliance with Commonwealth security, privacy, and data protection requirements.
- **Human Oversight:** AI score outputs are compared with human evaluators, and any significant discrepancies between AI and human ratings are subject to further independent human review, maintaining accountability and reducing risk of overreliance.
- **External Validation:** The approach and technology have been subject to external review (e.g., CSIRO Data61), with recommendations incorporated to strengthen governance, ethics, and risk controls.

These factors collectively ensure that the procured AI model is fit for purpose, supports the goals of the DMP2 evaluation process, and aligns with ethical, legal, and operational requirements.

5.4 Testing

Outline any areas of concern in results from testing. If testing is yet to occur, outline elements to be considered in testing plan (for example, the model's accuracy).

Has the AI system been tested sufficiently and are you satisfied with its reliability and safety for the context of your use case?		
Yes ☒	**N/A ☐**	**No ☐**

The DMP2 AI evaluation assistant has undergone multiple rounds of testing to assess its reliability and safety for the intended use case:

- **Early Experimentation:** Initial trials were conducted using synthetic data to simulate previous tender responses. These experiments helped identify both the strengths and limitations of the AI model in replicating human evaluation outcomes.
- **Prompt Refinement:** The technical and business teams enhanced prompt control to improve the consistency and accuracy of AI outputs, iteratively refining the approach based on test results.
- **Validation Against Human Outcomes:** With the consent of applicants, a dataset from DMP2 applications was ingested to validate AI outputs against known human evaluation outcomes. Discrepancies between AI and human ratings were analysed and used to further calibrate the system.
- **External Review:** Independent experts (e.g., CSIRO Data61) were engaged to review the approach, technology, risks, controls, governance, and ethics, with recommendations incorporated into the testing and deployment plan.
- **Human Oversight:** AI-generated scores are subject to review by human evaluators, and any significant differences trigger further independent human review to reach consensus.

Areas of concern identified and addressed:

- **Model accuracy and consistency:** Ongoing monitoring and prompt refinement have addressed initial concerns about variability in AI outputs.
- **Bias and fairness:** Distribution analysis and stakeholder feedback are used to identify and mitigate any patterns of bias.
- **Data privacy and security:** Data used in testing is handled in accordance with Commonwealth privacy and security requirements.

Planned testing:

- Validation with new datasets from future tender rounds.
- Review and adjustment of prompts, as needed, to ensure alignment with the evaluation criteria.
- Engagement with external experts for independent assessment, where needed.

These measures provide confidence that the AI system is reliable and safe for use in the DMP2 evaluation process.

5.5 Pilot

If answering 'yes', explain what you have learned or hope to learn in relation to reliability and safety and, if applicable, outline how you adjusted the use of AI.

Have you conducted, or will you conduct, a pilot of your use case before deploying?		
Yes ☒	**N/A ☐**	**No ☐**

A pilot has been conducted prior to full deployment of the DMP2 AI evaluation assistant. The pilot involved:

- **Testing with Synthetic and Real Data:** Early experimentation used synthetic data to simulate tender responses. This was followed by validation using case study data from DMP2 applications (with applicant consent).
- **Comparison of AI and Human Evaluations:** The pilot compared AI-generated ratings with those of human evaluators to identify discrepancies, assess consistency, and calibrate the AI system.
- **Stakeholder Engagement:** Feedback was gathered from the business team and tender specialists to assess the perceived reliability, safety, and fairness of the AI-assisted process.
- **External Review:** Independent experts (e.g., CSIRO Data61) reviewed the pilot approach, technology, and governance, providing recommendations that were incorporated into the final design.

Key learnings and adjustments:

- **Reliability:** The pilot confirmed that the AI system could reliably replicate human evaluation outcomes.
- **Safety:** Human-in-the-loop oversight is essential to mitigate risks of overreliance or erroneous recommendations.
- **Bias and Fairness:** The pilot identified areas where prompt refinement and additional controls were needed to minimise bias and ensure fair treatment of applicants.
- **Process Improvements:** The pilot informed improvements to data handling and the process for resolving discrepancies.

The pilot has provided valuable insights into the reliability and safety of the AI system, supporting its readiness for broader deployment in the DMP2 evaluation process.

5.6 Monitoring

If answering 'yes', explain how you will monitor and evaluate performance.

Have you established a plan to monitor and evaluate the performance of your AI system?		
Yes ☒	**N/A ☐**	**No ☐**

Monitoring has been established to evaluate the performance of the DMP2 AI evaluation assistant:

- **Ongoing Performance Tracking:** The system's outputs are compared against human evaluator ratings to identify discrepancies, trends, and potential areas for improvement.
- **Bias and Fairness Analysis:** Statistical analysis is conducted on evaluation outcomes to detect and address patterns of bias.
- **Stakeholder Feedback:** Feedback is sought from evaluators, tender resources, executives and other stakeholders to assess the perceived reliability, fairness, and transparency of the AI-assisted process.
- **Audit Trails:** All AI-generated recommendations and subsequent human decisions are logged and auditable, supporting transparency and accountability.

- **External Review:** The DTA may conduct external reviews (e.g., by CSIRO Data61 or other independent experts) to validate the system's performance, governance, and ethical compliance, as required.
- **Continuous Improvement:** Insights from monitoring activities are used to refine prompts and improve training for evaluators to ensure alignment with evaluation criteria.

This monitoring plan ensures that the AI system remains reliable, safe, and aligned with its intended purpose throughout its lifecycle.

5.7 Preparedness to intervene or disengage

See Guidance document for resources to assist you in establishing appropriate processes.

Have you established clear processes for human intervention or safely disengaging the AI system where necessary (for example, if stakeholders raise valid concerns with insights or decisions or an unresolvable issue is identified)?

Yes ☒	N/A ☐	No ☐

Clear processes have been established to ensure human intervention and safe disengagement of the AI system when necessary:

- **Human-in-the-loop Oversight:** AI-generated scores are reviewed by human evaluators. If there is a discrepancy between the AI and human ratings of more than an agreed margin, an independent human review is triggered to reach consensus.
- **Escalation Pathways:** If stakeholders raise valid concerns about AI-generated scores or comments, these concerns are escalated to the evaluation committee and, if required, to senior management for resolution.
- **Disengagement Protocols:** The AI system can be paused or disengaged from the evaluation process if unresolvable issues are identified, such as systemic bias, technical failures, or ethical concerns. In such cases, the evaluation reverts to fully manual human assessment until the issue is resolved.
- **Documentation and Audit Trails:** Interventions, escalations, and disengagements are documented to ensure transparency and accountability.
- **Monitoring:** Monitoring of system outputs and stakeholder feedback enables early detection of issues that may require intervention.

These processes ensure that the AI system operates safely, ethically, and in accordance with its intended purpose, with robust safeguards for human oversight and intervention.

6 Privacy protection and security

Under *Australia's AI Ethics Principles* AI systems should throughout their lifecycle respect and uphold privacy rights and data protection, and ensure data security.

6.1 Minimise and protect personal information

See Guidance on data minimisation and privacy enhancing technologies.

Are you satisfied that any collection, use or disclosure of personal information is necessary, reasonable and proportionate for your AI use case?

Yes ☒	N/A ☐	No ☐

The project is satisfied that the collection, use, and disclosure of personal information in the DMP2 AI evaluation assistant is necessary, reasonable, and proportionate for the following reasons:

- **Data Minimisation:** Only the minimum amount of personal information required to support the evaluation process is collected and used. Unnecessary personal data is excluded from the dataset used in the AI solution.
- **Consent:** Applicants provide consent for their data to be used in AI trials and evaluations and are informed of their rights to withdraw consent at any stage of the procurement process.
- **Privacy Controls:** Personally identifiable information (PII) is redacted before any data is published. Evaluators are trained to identify and manage PII appropriately.
- **Compliance:** Data handling practices comply with the Privacy Act 1988 and relevant Commonwealth privacy and data protection policies.
- **Security:** Data is stored and processed within secure, government-approved environments, with access controls and audit trails in place to prevent unauthorised access or disclosure.

These measures ensure that personal information is handled responsibly, with respect for privacy rights and in accordance with legal and ethical obligations.

6.2 Privacy assessment

Has the AI use case undergone a Privacy Threshold Assessment or Privacy Impact Assessment?

Yes ☒	N/A ☐	No ☐

The DMP2 AI evaluation assistant has undergone a Privacy Threshold Assessment to determine that a full Privacy Impact Assessment (PIA) is required. The assessment considered the nature and sensitivity of the data used, the potential for identification of individuals, and the risks associated with data handling and processing.

A full Privacy Impact Assessment is being conducted in accordance with Commonwealth privacy requirements. This process ensured that all personal information is handled lawfully, ethically, and in accordance with the Privacy Act 1988, with appropriate controls in place to mitigate privacy risks.

The outcomes of these assessments will inform the design and operation of the AI system, including data minimisation, consent processes, redaction of personally identifiable information, and secure data storage and processing.

6.3 Authority to operate

Engage with your agency's IT Security Adviser and consider the latest security guidance and strategies for AI use (such as Engaging with AI from the Australian Signals Directorate).

Has the AI system been authorised or does it fall within an existing authority to operate in your environment, in accordance with Protective Security Policy Framework (PSPF) Policy 11: Robust ICT systems?

Yes ☒	N/A ☐	No ☐

The DMP2 AI evaluation assistant operates within the DTA's secure ICT environment and complies with the Protective Security Policy Framework (PSPF) Policy 11: Robust ICT systems. The AI model is hosted on Microsoft Azure, a government-approved cloud platform, and all data transmission, storage, and processing are contained within the DTA's protected network.

The system is in the process of being reviewed and authorised by the agency's IT Security Adviser, ensuring alignment with the latest security guidance and strategies for AI use, including recommendations from the Australian Signals Directorate's "Engaging with AI" guidance. Access controls, audit trails, and regular security assessments are in place to maintain compliance and protect sensitive information.

These measures ensure the AI system is authorised to operate and meets all relevant Commonwealth security requirements.

Once this Authority to Operate has been completed, the link will be included in this section.

7 Transparency and explainability

Under *Australia's AI Ethics Principles*, there should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI and can find out when an AI system is engaging with them.

7.1 Consultation

Refer to the list of stakeholders identified in section 2. Seek out community representatives with the appropriate skills, knowledge or experience to engage with AI ethics issues. Consult the Guidance document for prompts and resources to assist you.

Have you consulted stakeholders representing all relevant communities or groups that may be significantly affected throughout the lifecycle of the AI use case?

Yes ☒	N/A ☐	No ☐

Stakeholders representing relevant groups have been consulted. This includes:

- **Applicants:** Direct engagement with applicants during tender rounds, including obtaining consent for the use of their data in AI trials and providing opportunities for feedback. This cohort includes SMEs and Indigenous businesses.
- **Tender Team:** Regular workshops and feedback sessions with tender specialists to ensure the AI system supports their needs and aligns with established evaluation practices.
- **External Experts:** Engagement with independent experts (e.g., CSIRO Data61) to review the approach, technology, risks, governance, and ethics, and to provide recommendations for improvement.
- **Advisors and Policy Stakeholders:** Consultation with relevant specialists, including legal, procurement, and probity advisers, to ensure compliance with Commonwealth policies and frameworks.

These consultations have informed the design, implementation, and ongoing monitoring of the AI system.

7.2 Public visibility

See Guidance document for advice on appropriate transparency mechanisms, information to include and factors to consider in deciding to publish or not publish AI use information.

Will appropriate information (such as the scope and goals) about the use of AI be made publicly available?		
Yes ☒	**N/A ☐**	**No ☐**

Appropriate information about the use of AI in the DMP2 evaluation process will be made publicly available. This includes:

- **Scope and Goals:** The purpose, scope, and intended benefits of the AI evaluation will be published, including how AI is used to support human evaluators and improve the efficiency and consistency of tender assessments.
- **Process Transparency:** Information about the evaluation methodology, criteria, and the role of AI in the process will be accessible to applicants and other stakeholders.
- **Privacy and Security:** Details regarding data handling, privacy protections, and security measures are available to ensure public confidence in the responsible use of AI.
- **Stakeholder Engagement:** Opportunities for feedback will be provided, and applicants will be informed about their rights and the mechanisms available to challenge procurement outcomes.
- **Compliance and Governance:** The DTA will publish documents in line with Commonwealth requirements.

These transparency measures are designed to build trust, support accountability, and ensure that all affected parties understand how AI is being used in the DMP2 evaluation process.

7.3 Maintain appropriate documentation and records

Ensure you comply with requirements for maintaining reliable records of decisions, testing and the information and data assets used in an AI system. This is important to enable internal and external scrutiny, continuity of knowledge and accountability.

Have you ensured that appropriate documentation and records will be maintained throughout the lifecycle of the AI use case?		
Yes ☒	**N/A ☐**	**No ☐**

Appropriate documentation and records will be maintained for use of the DMP2 AI evaluation assistant. This includes:

- **Decision Records:** Key decisions relating to the design, deployment, and operation of the AI system are documented.
- **Testing and Validation Logs:** Records of testing activities, validation results, and issues identified are maintained.
- **Data Asset Available:** Data assets are retained.
- **Audit Trails:** AI-generated content, human interventions, and final evaluation outcomes are kept ensuring accountability and enabling review.
- **Risk and Compliance Documentation:** Risk assessments, privacy and security reviews, and compliance with relevant frameworks and legislation are documented.

These practices ensure continuity of knowledge, support transparency and accountability, and enable effective internal and external review of the AI use case.

7.4 Disclosing AI interactions and outputs

Consider members of the public or government officials that may interact with the system or decision-makers that may rely on its outputs.

Will people directly interacting with the AI system or relying on its outputs be made aware of the interaction or that they are relying on AI -generated output? How?

Yes ☒	N/A ☐	No ☐

All individuals who interact with the DMP2 AI evaluation assistant or rely on its outputs are made aware of the involvement of AI in the process. This is achieved through several mechanisms:

- **Applicant Notification:** Applicants are informed in the Request for Tender (RFT) documentation that their submissions may be evaluated using an AI tool, and that their data may be used for training and validation purposes. Consent is sought, and applicants retain the right to withdraw their application.
- **Process Documentation:** Tender Evaluation Plans outline the use of AI in the evaluation process.
- **Transparency in Outcomes:** Where AI-generated outputs support evaluation outcomes, this is disclosed to decision-makers, and records are maintained to ensure transparency and accountability.
- **Contestability:** Mechanisms are in place for applicants to contest decisions, ensuring that the process remains open and transparent.

These measures ensure that parties are aware when AI is used and can understand its role in the evaluation process.

7.5 Offer appropriate explanations

If your AI system will materially influence administrative action or decision making by or about individuals, groups, organisations or communities, will your AI system allow for appropriate explanation of the factors leading to AI generated decisions, recommendations or insights?

Yes ☒	N/A ☐	No ☐

The DMP2 AI evaluation assistant is designed to allow for appropriate explanation of the factors leading to AI-generated recommendations and insights:

- **Transparent Criteria:** The AI system uses static prompts that directly reflect the formal evaluation methodology and criteria set out in the Tender Evaluation Plan (TEP). This ensures that recommendations are traceable to specific, published criteria.
- **Rationale Documentation:** For each evaluation, the AI provides a comment, referencing relevant aspects of the applicant's case study and the evaluation criteria applied.
- **Human Oversight:** Human evaluators may review AI-generated scores and can request access to AI generated comments if needed.
- **Auditability:** Decisions, including the final consensus reached by evaluators, are documented and auditable.

- **Stakeholder Communication:** Applicants can request explanations for evaluation outcomes.

These measures ensure that any administrative action or decision materially influenced by the AI system is accompanied by a clear, understandable explanation of the factors involved, supporting transparency, contestability, and accountability.

8 Contestability

Under *Australia's AI Ethics Principles*, when an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system.

8.1 Notification of AI affecting rights

See Guidance document for help interpreting 'administrative action', 'materially influenced' and 'legal or similarly significant effect' as well as recommendations for notification content.

Will individuals, groups, organisations or communities be notified if an administrative action with a legal or similarly significant effect on them was materially influenced by the AI system?

Yes ☒	N/A ☐	No ☐

Applicants will be notified if an administrative action with a legal or similarly significant effect on them was materially influenced by the AI system. The notification process includes:

- **Tender Documentation:** Applicants are informed in the Request for Tender (RFT) and related documentation that their submissions may be evaluated using an AI tool, and that AI-generated scores and comments may inform the assessment process.
- **Outcome Communication:** Applicants are notified of administrative actions, such as panel inclusion or exclusion. However, it is important to note that the exclusion process will not involve AI. If AI and a human agree that an Applicant is recommended to be found unsuitable, a second, independent human will perform the evaluation and the AI results will not be considered.
- **Right to Contest:** Applicants are provided with information about their rights to contest the outcome, or request a review if they believe the decision was unfair.
- **Transparency:** The DTA maintains clear records of how AI recommendations have contributed to final decisions.

These measures ensure that all affected parties are aware of the involvement of AI in decisions that may impact their rights or interests, supporting transparency, contestability, and accountability.

8.2 Challenging administrative actions influenced by AI

Administrative law is the body of law that regulates government administrative action. Access to review of government administrative action is a key component of access to justice. Consistent with best practice in administrative action, ensure that 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.

Is there a timely and accessible process to challenge the administrative actions discussed at 8.1?

Yes ☒	N/A ☐	No ☐

There is a timely and accessible process in place to challenge administrative actions that have been materially influenced by the AI system:

- **Complaints and Dispute Resolution:** The DTA has established complaints and dispute resolution processes, requiring prompt, fair, and transparent responses to applicant concerns regarding evaluation outcomes.
- **External Review Mechanisms:** Applicants can seek review through the Government Procurement (Judicial Review) Act 2018 (Cth).
- **Transparency and Record-Keeping:** Decisions, including those informed by AI, are documented and auditable, supporting contestability and review.
- **TEP Compliance:** The Tender Evaluation Plan (TEP) outlines measures to ensure fair, equitable, and non-discriminatory treatment of applicants, and any variations are considered by the Chair in consultation with legal, procurement, and probity advisers.

These processes ensure that no person loses a right, privilege, or entitlement without access to a review process or an effective way to challenge an AI-generated or informed decision, consistent with best practice in administrative action.

9 Accountability

Under *Australia's AI Ethics Principles*, those responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.

9.1 Establishing responsibilities

Where feasible, it is recommended that these three roles not all be held by the same person. The responsible officers should be appropriately senior, skilled, and qualified.

Who is responsible for:	
use of AI insights and decisions	Evaluation Committee (EC) Chair and Delegate, supported by the EC members. The EC Chair and Delegate are accountable for how AI-generated insights are used in the evaluation process, ensuring all decisions are consistent with the Tender Evaluation Plan (TEP) and subject to human oversight.
monitoring the performance of the AI system	Principal AI Technical Advisor, with oversight from the CTO and project team. The Principal Technical Advisor is responsible for ongoing monitoring, validation, and reporting on the AI system's performance, including bias, reliability, and safety. Oversight is provided by the CTO and project team to ensure technical and ethical standards are maintained.
data governance	Data Governance Lead, in consultation with Legal, Procurement, and Probity Advisers (where required). The DTA Data Governance Lead ensures data used in the AI system is managed in accordance with the Tender Evaluation Plan.

9.2 Training of AI system operators

With all automated systems, there is always the risk of overreliance on results. It is important that the operators of the system, including any person who exercises judgment over the use of insights, or responses to alerts, are appropriately trained on the use of the AI system. Training should be sufficient to understand how to appropriately use the AI system, and to monitor and critically evaluate outcomes.

Is there a process in place to ensure operators of the AI system are sufficiently skilled and trained?

Yes ☒	N/A ☐	No ☐

There are processes in place to ensure that key operators of AI evaluation assistant are sufficiently skilled and trained:

- **Suitably trained staff**: AI system operators are suitably experienced in the responsible use of AI, including understanding the AI's capabilities, limitations, and the importance of human oversight.
- **Critical Evaluation Skills:** Evaluators will apply critical evaluation skills independently of any use of AI. AI will only be introduced following completing of independent human evaluation. If the difference between evaluation scores from the human evaluator and the AI evaluator exceed a threshold, the AI score will be discarded in favour of a second independent human evaluator.
- **Ethics and Compliance:** AI system operators are appropriately skilled on relevant ethical principles, privacy requirements, and Commonwealth policies governing AI use. Alternatively, external experts may be sought to inform these processes.
- **Documentation and Guidance:** Evaluators are trained and have access to documentation to support them in their role.

These measures ensure that all operators are equipped to use the AI system appropriately, monitor outcomes, and maintain high standards of fairness, reliability, and accountability.

10 Human-centred values

Under *Australia's AI Ethics Principles*, AI systems should throughout their lifecycle respect human rights, diversity and the autonomy of individuals.

10.1 Incorporating diversity

Consider how you have incorporated diversity of perspective through the lifecycle of your AI use case – for example, through the choice of data, composition of development and deployment teams and the stakeholder and user groups to choose to consult.

Are you satisfied that you have incorporated diversity and people with appropriately diverse skills, experience and backgrounds throughout the lifecycle of your AI use case?

Yes ☒	N/A ☐	No ☐

Diversity has been incorporated throughout the lifecycle of the DMP2 AI evaluation assistant use case in several key ways:

- **Data Diversity:** The dataset includes case study responses from a wide range of applicants, including SMEs, Indigenous businesses, and organisations of varying sizes and backgrounds, ensuring that the AI system is exposed to diverse perspectives and experiences.
- **Team Composition:** The development, deployment, and evaluation teams include members with varied skills, professional backgrounds, and expertise in technology, procurement, ethics, legal, and policy domains.
- **Stakeholder Consultation:** Input has been sought from a broad spectrum of stakeholders, including, evaluators, and external experts (such as CSIRO Data61).
- **Inclusive Design:** The evaluation process and supporting documentation have been designed to be accessible and inclusive, minimising barriers for all eligible applicants and ensuring fair treatment across cohorts.
- **Continuous Improvement:** Feedback is sought and used to refine the AI system, prompts, and evaluation criteria, supporting ongoing improvement and responsiveness to stakeholder needs.

These measures ensure that the AI use case respects human rights, values diversity, and supports the autonomy of individuals and organisations throughout its lifecycle.

10.2 Human rights obligations

It is recommended you complete this question after completing previous sections of the assessment. This approach will enable a more considered assessment of the human rights implications of your AI use case.

Have you consulted an appropriate source of legal advice or otherwise ensured that your AI use case and the use of data align with human rights obligations?		
Yes ☒	**N/A ☐**	**No ☐**

Appropriate legal advice has been sought, and will continue to be sought where needed, to inform the DMP2 AI evaluation assistant and its use of data align with Commonwealth laws and policy. This includes:

- **Legal Review:** The use case and associated data handling practices have been reviewed by the DTA's legal advisers to confirm compliance with relevant Commonwealth legislation.
- **Policy Alignment:** The AI system is designed and operated in accordance with the Commonwealth policy.
- **Stakeholder Consultation:** Input from legal, procurement, and probity advisers has informed the development and deployment of the AI system, ensuring that human rights, fairness, and non-discrimination are upheld throughout the lifecycle of the use case.

11 Internal review and next steps

11.1 Legal review of AI use case

The final AI solution will be reviewed by the DTA's probity advisor and privacy advisor, with recommendations addressed prior to production deployment of the solution.

11.2 Risk summary table

In the table below, list any risks identified in section 3 (the threshold assessment) or subsequently as having a risk severity of 'medium' or 'high'. Also list any instances where you have answered 'no' in any of the questions in sections 4 to -10.

As you proceed through internal review (section 11.3) and, if applicable, external review (section 11.4), list any agreed risk treatments and assess residual risk using the risk matrix in section 3.

Risk summary table

Risk	Risk treatments	Residual risk
Negatively affecting public accessibility or inclusivity of services	- Application of DTA's Contract Limits and Reviews Policy - TEP ensures equal and fair opportunity - Ongoing bias monitoring and stakeholder feedback	Low
Unfairly discriminating against individuals, communities or groups	- Complaints and dispute resolution processes - Collaboration with Ombudsman - TEP compliance - Human oversight	Low
Perpetuating stereotyping or demeaning representations	- Human-in-the-loop evaluation - Independent review for dissenting scores - Chair/Delegate oversight	Low
Harming individuals, communities, groups, organisations or environment	- Human oversight - Critical review of AI outputs - Encouragement of evaluator independence	Low
Raising privacy concerns due to sensitivity, amount or source of data	- Redaction of PII - Privacy Threshold Assessment and, if required, Privacy Impact Assessment - Evaluator training	Low
Raising security concerns due to data sensitivity/classification	- Secure network and storage - Access controls - Security risk assessment - Probity requirements	Medium
Raising security concerns due to AI system implementation	- Pilot testing - Oversight and logging - Evaluator training - External review	Medium

Influencing decision-making affecting individuals/groups	- Record-keeping of AI influence - Contestability and review rights - Human oversight	Medium
Posing reputational risk or undermining public confidence	- Public transparency - External validation - Clear communication of AI use	Medium
Any ‘no’ answers in sections 4–10	(No ‘no’ answers recorded; all controls in place)	N/A

Note:

- *Residual risk ratings are assessed using the risk matrix in section 3.*
- *Risk treatments are based on controls and processes described in previous sections.*
- *The table should be updated as further internal or external reviews are completed and additional risk treatments are agreed.*

11.3 Internal review of AI use case

An internal agency governance body designated by your agency's Accountable Authority must review the assessment and the risks outlined in the risk summary table.

The governance body may decide to accept any 'medium' risks, to recommend risk treatments, or decide not to accept the risk and recommend not proceeding with the AI use case.

List recommendations of your agency governance body below.

Recommendations of internal agency governance body
• **The agency's Accountable Authority has reviewed the assessment and the risks outlined in the risk summary table.** • **All identified medium risks have been considered, and the proposed risk treatments are deemed appropriate and sufficient to mitigate these risks to an acceptable level.** • **The Accountable Authority recommends accepting the residual medium risks, subject to ongoing monitoring and periodic review.** • **It is recommended that the DMP2 AI evaluation assistant proceed to the next stage, with continued oversight and engagement with stakeholders, and with a commitment to update risk treatments as new information or feedback becomes available.**

11.4 External review of AI use case

If, following internal review (section 11.3), there are any residual risks with a 'high' risk rating, consider whether the AI use case and this assessment would benefit from external review.

If an external review recommends further risk treatments or adjustments to the use case, your agency must consider these recommendations, decide which to implement, and whether to accept any residual risk and proceed with the use case.

If applicable, list any recommendations arising from external review below and record the agency response to these recommendations.

Has your AI use case been subject to external review?	**Yes** ☒ N/A ☐ No ☐
Who conducted the external review?	**CSIRO Data61**
Date external review last completed	**April 2025**

External review recommendation	Agency response
Further refinement of prompt control to improve consistency and accuracy of AI outputs.	Prompt control has been enhanced and regularly reviewed; iterative improvements are made based on ongoing validation and feedback.

Enhanced bias monitoring to detect and mitigate any patterns of unfairness in AI-generated recommendations.	Bias monitoring processes have been implemented, including statistical analysis and stakeholder feedback, with regular reporting.
Strengthened documentation of decision-making processes, including clear audit trails for AI and human evaluations.	Comprehensive documentation and audit trails are maintained for all AI-generated recommendations, human interventions, and outcomes.
Ongoing stakeholder engagement to ensure diverse perspectives and address concerns throughout the lifecycle of the AI system.	Stakeholder engagement is ongoing, with regular consultations, feedback sessions, and transparent communication of AI use and outcomes.
Periodic revalidation of the AI system as new data and evaluation criteria are introduced.	The AI system is subject to periodic revalidation and external review, especially when new data or criteria are incorporated.
Maintain robust human oversight throughout the lifecycle to ensure accountability and contestability.	Human-in-the-loop oversight is a core feature; all AI outputs are reviewed by evaluators, with escalation protocols for discrepancies.

This document has been created in part with the assistance of an AI large language model (LLM).