Australian Government

Services Australia

Draft AI Assurance Plan

EARef 112 Finalised No Debt

September 2024

Automation and Service Optimisation Branch

Contents

Purpose

The agency needs to engage responsibly, ethically, and safely with AI to innovate and modernise government services so people can get on with their lives.

In 2024, the agency's AI strategy was approved by the Executive Committee. The strategy states that all initiatives using AI will develop an AI Assurance Plan to ensure they are appropriately managed.

This document is to be used as a template for an assurance plan. This document is to be used by initiatives from discovery, to implementation, to on-going support and assurance. If a section does not yet apply to an initiative (e.g. it is only in discovery), **do not delete it** – provide justification as to why it doesn't yet apply.

Background

This document brings key considerations from the agencies AI strategy and the Digital Transformation Agencies Draft Commonwealth AI Assurance Framework into one tool that is to be used to ensure that AI is engaged with in accordance to Whole of Government guidelines and that the agency is engaging responsibly, ethically, and safely with AI.

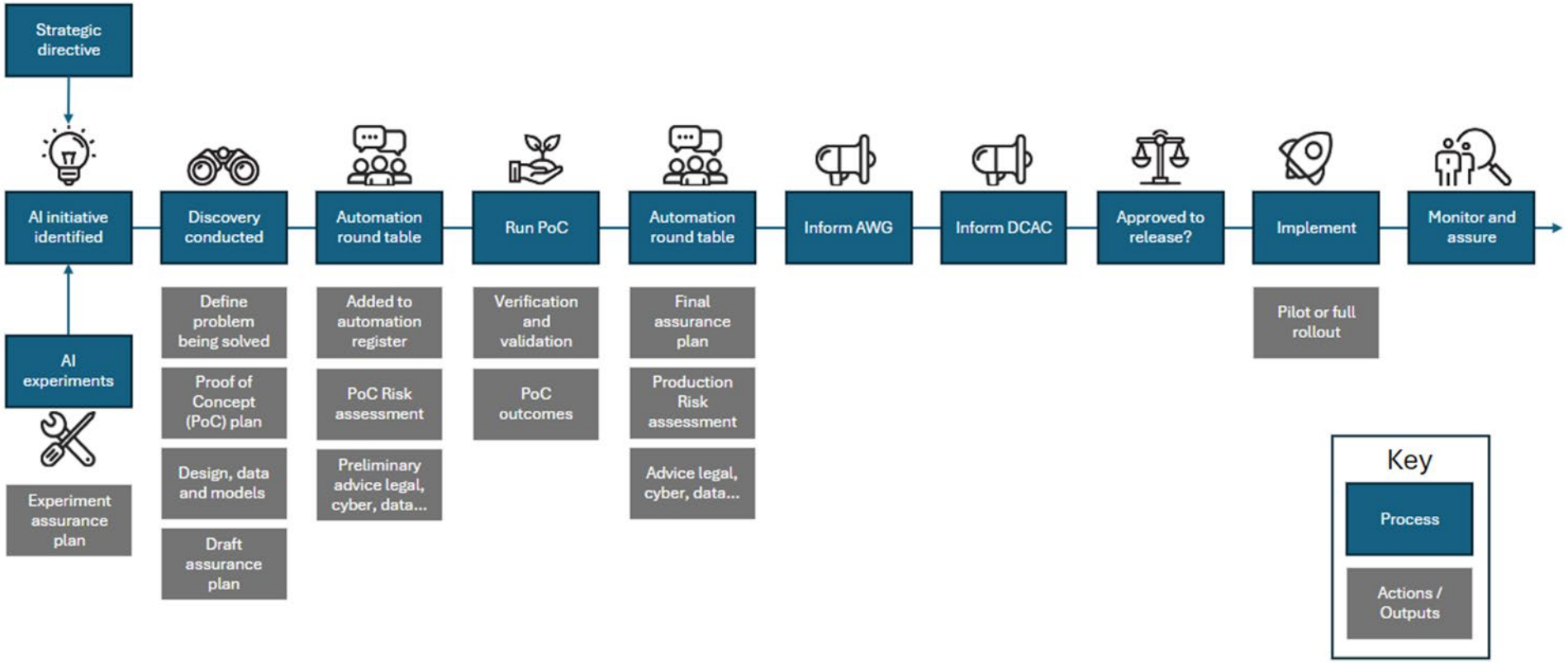
At a time when public trust in the Australian Government is decreasing, the ethical use and protection of personal data when it comes to AI will be pivotal.

Our responsibility is to make the most of government investments in our capability in a way that maximises the benefit to our customers through proactive, informed and carefully considered action. For this reason, Services Australia has an imperative to engage with AI now.
This assurance plan is part of the agency's Automation governance and assurance framework, which includes the consideration of AI uses cases. It should be used by projects/initiatives that are using AI as part of their solution, and should be used to assist and complement the existing agency governance and assurance processes.

The use of this assurance plan does not remove the need for a project/initiative owner to engage with corporate governance requirements, project management processes, risk management responsibilities processes, change management processes, and the Chief Information and Digital Officer's architectural, design, cyber security, procurement, and release management processes. It should not be considered, or used in isolation of these other project/initiative owner responsibilities.

The Commonwealth AI Assurance Framework is the bases for the questions that are required to be answered in this assurance plan. The purpose of Commonwealth AI Assurance Framework is to guide Australian government agencies through impact assessment of AI use cases against Australia's AI Ethics Principles. It is intended to complement and strengthen – not duplicate – existing frameworks, legislation and practices that touch on government's use of AI. It should be read and applied alongside the Policy for the responsible use of AI in government and existing frameworks and laws to ensure agencies are meeting all their current obligations.

High-level AI Governance process



1. Basic information

1.1 AI use case profile

Complete the information below.

Name of AI use case	Finalised No Debt project
Reference number	EA Ref 112
Lead agency	Services Australia
Assessment contact officer	Name: Stacey Miller-O'Neill Email: Section 22
Executive sponsor	Name: Jason Kallus, National Manager, Integrity Transformation Branch Email: Section 22 Name: Mark Morrison, National Manager, Data Analytics Innovation Branch Email: Section 22

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

AI use case description

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

Type of AI technology
Supervised machine learning model on Tabular dataset was used to determine the confidence score of a potential overpayment resulting in an FND outcome

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	☐
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.

What is the next date/milestone that will trigger the next review of the AI use case?
The project has been endorsed to move to deployment stage, which for this model, is a one-off deployment. Post deployment and in the monitoring, stage would be the next review timeframe - January 2025.

1.4 Assessment review history

Track the review history for this assessment in the table below. Include brief summaries of changes arising from reviews (50 words or less).

Summary of changes arising from review	Review date
Initial Assessment September 2024	January 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

Assessing, calculating, and finalising a potential overpayment requires 3 systems, 9 tools, and up to 328 steps.

The agency currently has a potential overpayment backlog of approximately 1.8 million debts, with 10,000-15,000 new potential overpayments added to the backlog each week as of August 2024, with no simple method of identifying which ones may result in an FND outcome. This means that Payment and Assurance Operations (PAO) staff spend time assessing each potential overpayment, with approximately 1 in 10 resulting in an FND outcome.

FND outcomes are unproductive as the assessment is a highly manual process that results in a nil outcome for both the customer and the agency.

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

Using Data Science techniques, develop a solution that assigns a confidence score to potential overpayments that are likely to result in FND. This assists with segmenting potential overpayments, to enable allocation of productive work to PAO staff during the MSA period.

2.3 Non-AI alternatives

Use 100 words or less.

Briefly outline non-AI alternatives that could address this problem

The current process for FND is highly manual with multiple touchpoints. On average, staff require the use of 3 systems, 9 tools and up to 328 to make the final determination.

There is currently no easy way of identifying a potential overpayment that may result in an FND outcome, meaning that staff must spend time to assess each overpayment, with approximately 1 in 10 resulting in an FND outcome. This is an unproductive use of staff time as it takes them away from other higher priority work.

Alternative strategies to reduce the potential overpayment backlog have not yielded expected results and the backlog continues to grow.

2.4 Identifying stakeholders

Identify stakeholder groups that may be affected by the AI use case and briefly describe how they may be affected, whether positively or negatively. This will guide your consideration of expected benefits and potential risks in this assessment.
Consider holding a brainstorm or workshop to help identify affected stakeholders and how they may be affected. A discussion prompt is provided in the guidance document.

Stakeholder group	Briefly describe how they may be affected
Project Team - Integrity Transformation	Adopting this model would align to the Integrity Transformation Branch's priorities to: • Identify potential overpayments in the backlog that would likely result in an FND outcome. • Provide treatment options for identified FND debts, such as keyword and quarantine or priority allocation. • Provide a measure of the volume of FND cases in the potential overpayment backlog. • Support the prioritisation of higher priority work. • Support advanced functionality and alternative treatments.
End User - Payment Assurance Operations Staff	Pending a treatment pathway, enable PAO staff to focus on more productive work over the MSA period.
AI model or AI system engineers – Machine Learning Operations	Monitoring and Maintenance of the model once deployed.
Operations Management	Workload prioritisation and allocation

2.5 Expected benefits

Considering the stakeholders identified in the previous question, identify the expected benefits of the AI use case. This should be supported by quantitative and/or qualitative analysis.
Qualitative analysis should consider whether there is an expected positive outcome and whether AI is a good fit to accomplish the relevant task, particularly compared to non-AI alternatives identified. Benefits may include gaining new insights or data.
Consult the guidance document for resources to assist you. Aim for 300 words or less.

What are the expected benefits of the AI use case?
Service Delivery Integrity - Demand Management • Improves demand and supply management of potential overpayments in the agency

- Improve the agency's capacity planning and resource allocation by being able to identify FND potential overpayments, which do not require customer contact, will allow the agency to finalise these debts during the MSA period. This will assist in reducing the existing potential overpayment backlog.

Enhanced Opportunities - Foundational Capability

- Creates a reusable, scalable, upgradable solution for identification of debt features that can assist with efficient resource allocation
- As the predictive model is refined through use cases, it can be utilised to identify additional work types for targeted allocation of the 'Right Work, to the Right Person at the Right Time'.

Cost to Government - Improved Productivity

- Improves productivity by enabling allocation of ready to action work items during the MSA period.
- Improved productivity by enabling the processing of potential overpayments during the MSA period and reducing reallocations of work as staff will be able to finalise FND potential overpayments.

3. Threshold assessment

Using the risk matrix, determine the severity of each of the risks in the table below, accounting for any risk mitigations and treatments. Provide a rationale and an explanation of relevant risk controls that are planned or in place. The Guidance document contains consequence and likelihood descriptors and other information to support the risk assessment.

The risk assessment should reflect the intended scope, function and risk controls of the AI use case. Keep the rationale for each risk rating clear and concise, aiming for no more than 200 words per risk.

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
1. Negatively affecting public accessibility or inclusivity of government services	**Low ☐** Med ☐ High ☐
Rationale for risk rating and explanation of relevant risk controls: Accessibility to government services is not impacted by this model or ongoing BAU debt collection process.	
2. Unfairly discriminating against individuals, communities or groups	**Low ☐** Med ☐ High ☐

Rationale for risk rating and explanation of relevant risk controls: Principles of justice and equity have been considered. No additional bias has been built into the model and customer demographic data has not been used to determine the confidence score.	
3. Perpetuating stereotyping or demeaning representations of individuals, communities or groups	**Low ☐** Med ☐ High ☐
Rationale for risk rating and explanation of relevant risk controls: No groups of customers are under or overrepresented with the use of the model. All customers impacted already appear in the potential overpayment backlog, the model uses existing debt attributes to provide a confidence score of FND or no FND to each debt shell.	
4. Harming individuals, communities, groups, organisations or the environment	**Low ☐** Med ☐ High ☐
Rationale for risk rating and explanation of relevant risk controls: The outcome of this initiative does not change outcome of following regular BAU processes. As part of this project, staff will still determine the actual outcome under BAU processes.	
5. Compromising privacy due to the sensitivity, amount or source of the data being used by an AI system	**Low ☐** Med ☐ High ☐
Rationale for risk rating and explanation of relevant risk controls: Section 42 It is also our understanding that the FND - MSA strategy will not involve any automated debt decision making. The current intention is for potential overpayments with a higher confidence rating be allocated to staff who will continue to manually investigate and determine if an FND outcome is appropriate.	
6. Raising security concerns due to the sensitivity or classification of the data being used by an AI system	**Low ☐** Med ☐ High ☐
Rationale for risk rating and explanation of relevant risk controls: Personal information has been collected but will not be used in the modelling process. Information being used in the modelling process cannot be used to identify the customer. Data Extraction team extract from the Enterprise Data Warehouse (EDW), encrypt the identifiable data and then provide to Data Science team in SAS. Data Science team place it into the data lake/hdfs. Data Science team do not see the raw or unencrypted data.	
7. Raising security concerns due to the implementation, sourcing or characteristics of the AI system	**Low ☐** Med ☐ High ☐
Rationale for risk rating and explanation of relevant risk controls: Data does not leave Services Australia's core systems, greatly reducing any risks. Personal information extracted from the EDW before being encrypted in SAS and moved to the Data Science SAS location. The Data Science team then move the data to the datalake.	

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

Rationale for risk rating and explanation of relevant risk controls:

No automation decision making, a human in the loop process has been designed and built into the model that replicates existing BAU process.

A risk raised at DTEC on 30 August, that staff may see the keyword which could influence decision making. The proposed mitigation strategies for these risks, are that:

1. Staff will receive a mixture of different work types for action including work outside of the Youth Allowance stream.
2. Keyword applied to enable workload allocation/reporting, shall not be identifiable as a potential FND to remove the risk of influence/bias.

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

Rationale for risk rating and explanation of relevant risk controls:

There is always a perceived risk with AI and debt. All steps have been taken to protect the customers data, store and move the data within approved architectural patterns individual branch control plans. Use an unbiased data set, ensure the process simulates BAU process as close as possible and maintain the human in the loop with decision making.

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 **has been undertaken** for this project
Comments (optional)	As part of the Data Science Section's commitment to good AI governance processes, we have opted to complete the full assessment although we have rated the project Threshold Assessment as an overall low risk, with the exception of *8. Influencing decision-making that affects individuals, communities, groups, organisations or the environment.* This Threshold was rated as medium as staff may identify keywords applied to debt shells mean a higher likelihood of the outcome being an FND determination. Relevant mitigations have been put in place to address this risk. Collaboration between Data Science Section and Integrity Transformation Branch commenced in November 2022 and the output is planned as a once off static model deployment during the MSA period in December 2024. Information on previous model validation trials can be found at *5.Reliability and Safety section* to ensure adequate testing and validating of the model has occurred. I am confident that throughout the lifecycle of this project, appropriate engagement with program areas, AI Assurance and Governance processes and Enterprise Governance Committee Supporting Forums has occurred with advice sought, Section 42 and Data Trust and Ethics Committee throughout this project.
Name & position	Section 22 Assessment Contact Officer, Data Science Lead Project Manager
Date	13 September 2024

Executive sponsor endorsement

Endorsement	I have reviewed the recommendation, am satisfied by the analysis and agree that a **full assessment** will support AI governance assurance processes
Comments (optional)	Section 22
Name & position	Jason Kallus, National Manager Integrity Transformation
Date	13 September 2024

Executive sponsor endorsement

Endorsement	I have reviewed the recommendation, am satisfied by the analysis and agree that a **full assessment** will support AI governance assurance processes
Comments (optional)	
Name & position	Mark Morrison, National Manager Data Analytics Innovation
Date	

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 ☐

The model uses characteristics in existing potential overpayments to predict whether potential overpayments in the backlog will likely result in an FND outcome or not. Once a prioritised list is formed this will be allocated to staff for action during the MSA period. All potential overpayment work items allocated and processed as part of the treatment strategy will retain human oversight:

- the delegate will be unaware of the predictive score,
- the predictive score will not change/impact the delegate's decision,
- the decision under the social security law is a decision made by a human delegate,
- processing of debt shells is fully in line with existing Business-as-Usual processes.
- staff will continue to undertake all relevant checks of a customer's record, engage with a customer (as necessary) and make independent decisions based on relevant Legislation and agency Policies.
- ensuring a fair outcome for all.

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 ☐

Explain your answer:

The AI does not unfairly discriminate certain individuals based on their background or ethnicity. This model will be used primarily to assist with workload allocation and will not inherently impact fairness outcomes for individuals as an FND results in a **nil** outcome for a customer. All allocated potential overpayments will be processed by staff.

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 ☐

Explain your answer:

The inputs of the FND model were determined in conjunction with business SME's who were experienced with processing YAL debts. These inputs/features represent information that would be required in the assessment of a YAL PIT debt or are associated with an FND outcome

5.2 Indigenous data

*Consider whether your use of Indigenous data and AI outputs is consistent with the expectations of Indigenous people, and the forthcoming **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 (expected in 2024)?

Yes ☐	N/A ☐	No ☐

Explain your answer:

The model uses all customer data which may include indigenous customer data, but this does not feature in the model.

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 ☐

Explain your answer:

n/a

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 ☐

Explain your answer:

Two separate testing and validation exercises were conducted:

July 2023 the Project Team validated 250 debts with ~95% (74/78) of potential overpayments with a probability score of >90%, resulting in a FND outcome.

March 2024 a Minimum Viable Product (MVP) was conducted with Payment Assurance Operations Division staff to validate 420 records. This was conducted via a blind test model under Business-as-Usual conditions, where staff were unaware of the predictive model outputs or associated probability scores.

Quality Assurance checks of the predictive model outputs and subsequent human delegate decisions were completed by SMEs within the PAO Branch and the Project Team.

The MVP demonstrated significant alignment of the model predictive scores against outcomes determined by human delegates.

- 94% of the records that had a 90%+ probability score resulted in an FND outcome.
- 83% of records that had a 80%+ probability score resulted in an FND outcome.

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 ☐

Explain your answer:

Beta Validation / Pilot was conducted Feb-Apr 2024.

Results report and methodology can be provided.

See 5.4 for some information relating to results.

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 ☐

Explain your answer:

To monitor the model, we will compare the outcomes predicted by the model to the actual outcomes determined by staff, initially on a daily basis for the duration of the MSA period.

This process will involve:

- Extracting the actual outcome of the potential overpayment as determined by the service officers.
- Comparing these actual outcomes to the results predicted by the model.
- If there is a significant deviation between these two, then stop-gap measures may be applied to override the prioritisation of the potential overpayments.
- Noting that if the model did make an incorrect prediction of FND, CSO's follow BAU procedures to place the potential overpayment on hold until after the MSA period.
- If there is a significant deviation between predictive model scores and human delegate outcomes, a stop gap measure will be applied to cease allocation.

5.7 Preparedness to intervene or disengage

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

Have you established clear processes to intervene or safely disengage the AI system if stakeholders raise valid concerns with insights or decisions or an unresolvable issue is identified?

Yes ☐	N/A ☐	No ☐

Explain your answer:

To be further refined as the model deployment plan is established with business stakeholders.

Identifying a significant deviation will depend on the number of potential overpayments processed throughout the period.

As an example, if there is a significant deviation between predictive model scores and human delegate outcomes, a stop gap measure will be applied to cease allocation.

It is worth noting that any deviation of the model will pose minimal risks to customers but may result in rework for staff after the conclusion of the MSA period, as staff are instructed not to contact customers during this period.

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 ☐

Explain your answer:

Personal information was collected for building the test dataset for the model. It was encrypted prior to use and no personally identifiable information was used in any model builds.

- The information reported by the customer to the Agency would reasonably be understood to be used in calculating any potential debt.
- All information used within the project, including the FND confidence level output, will remain within the Agency's system.
- The information extracted by the program is already held within the agency and is the same information staff currently have available to make an FND determination.
- FND confidence level data outputs will be de-identified.

6.2 Privacy assessment

Has a Privacy Threshold Assessment or Privacy Impact Assessment been undertaken?

Yes ☐	N/A ☐	No ☐

Explain your answer:

Section 42

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 to operate in your environment, in accordance with policy requirements in PSPF Policy 11: Robust ICT systems?

Yes ☐	N/A ☐	No ☐

Explain your answer:

The use of AI here as a small-scale. The AI will produce a static list of potential overpayment work items that are highly likely to result in an FND outcome and will be fed into the workload manager allocation tool by the Operations Management Division for the appropriately skilled staff to work on during upcoming MSA period.

It is noted that if work items are endorsed to be processed and allocated as part of Business-as-Usual it will retain human oversight:

- the delegate will be unaware of the predictive score,
- the predictive score will not change/impact the delegate's decision,

the decision under the social security law is a decision made by a human delegate.

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 ☐

Explain your answer:

Internal stakeholders have been engaged for the duration of the model build lifecycle including:

- Payment Assurance Program and Appeals Division, Integrity Transformation Branch
- Data and Analytics Division, Data Analytics Innovation Branch
- Legal Services Division, Program Advisory Legal Branch
- Data and Analytics Division, Data Strategy and Governance Branch
- Payment Assurance Operations Division
- Payment Assurance Priorities Board
- Debt and Compensation Program Branch
- Student Programs Program
- Data Trust and Ethics Committee (DTEC) endorsement received on 30th August 2024

Fraud Division

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 ☐

Explain your answer:

Existing publicly available data on the collection policy states that we may use information provided for debt collection purposes.

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 ☐

Explain your answer:

Yes, full project and business documentation including SRO signoffs have been received.

- Plan on a Page - approved business case
- Current state project artefacts
- Fortnightly project status reporting
- Discovery report
- DataTEST and Risk Plan
- PTA
- Legal Advice
- Methodology Paper - Alpha
- Beta Validation Trial paper and outputs
- SHAP plots - Explainability
- Governance board papers - PAP Board and DTEC
- Project Artefact Guide
- BIR's
- Quality Assurance and User Guide - Data Extraction

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 ☐

Explain your answer:

It is noted that if work items are endorsed to be allocated and processed as part of Business-as-Usual to retain human oversight:

- the delegate will not know the predictive score,
- the predictive score will not change the delegate's decision,
- the decision under the social security law is a decision made by a human delegate.

7.5 Offer appropriate explanations

If your AI system will materially influence 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 ☐

Explain your answer:

No decision making or automation occurs by the model. The model includes A human in the loop approach to decision making.

It is noted that if work items are endorsed to be processed and allocated as part of the work allocation that occurs during the MSA period 16.12.24 – 1.1.25 (dates to be confirmed by PAO Board)

- the delegate will be unaware of the predictive score,
- the predictive score will not change/impact the delegate's decision,
- the decision under the social security law is a decision made by a human delegate.
- staff will maintain assessment of potential overpayments in line with current Business-as-Usual practices.
- staff will continue to undertake all relevant checks of a customer's record, engage with a customer (as necessary) and make independent decisions based on relevant Legislation and agency Policies.

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 their rights was materially influenced by the AI system?		
Yes ☐	**N/A ☐**	No ☐

Explain your answer:

Existing publicly available data on the collection policy states that we may use information provided for debt collection purposes.

The rights of individuals are not materially influenced by this model. The service officer processing the potential overpayment as part of their work allocation, will not be provided the outcome of the model and will review the case and make an informed decision based on their own review of the situation and details in the system, ensuring a Human in the Loop decision making process at all times.

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 ☐

Explain your answer:

Decision making will be by an existing suitably trained and skilled service officer. Customers have the same rights to complain, challenge or appeal any decisions.

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	NM Jason Kallus, Integrity Transformation
monitoring the performance of the AI system	NM Mark Morrison, Data Analytics Innovation
data governance	NM Heather Ralph, Data Strategy and Governance

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 ☐

Explain your answer:

Machine Learning Engineers will have training on the model, build components and expected use for monitoring and maintenance

Business stakeholders will have training and support on model outputs

BAU Process for staff – no training required

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 ☐

Explain your answer:

Yes, all stakeholders involved in the project from Data Extraction, model build, testing, validation, pilot processing, project team represent diversity in backgrounds, ages, skills knowledge, role etc.

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 ☐

Explain your answer:

- Yes PTA and Section 42 ,
- DataTEST completed iteratively with advice sought from Data Ethicist
- DTEC Endorsement received on 30th Aug 2024

11. Internal review and next steps

11.1 Legal review of AI use case

This section must be completed by a qualified legal adviser. Ensure any supporting legal advice is available for the remaining review steps. Repeat this step if there are significant changes.

I **[am/am not]** satisfied that the AI use case and the use of data meet legal requirements.

Legal review outcome	I **[am / am not]** satisfied that the AI use case and the use of data meet legal requirements.
Comments (optional)	
Name & position of legal adviser	
Date	

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 4-10.

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

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

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.

Recommendations from external review and agency response

servicesaustralia.gov.au