# P1#y1

Algorithm impact assessment report

Algorithm Charter for Aotearoa New Zealand

December 2023 

Contents

[Algorithm Impact Assessment 1](#_Toc153546465)

[Report Template 1](#_Toc153546466)

[Algorithm Charter for Aotearoa New Zealand 1](#_Toc153546467)

[1. Purpose and summary 3](#_Toc153546468)

[Purpose 3](#_Toc153546469)

[Summary of key points 3](#_Toc153546470)

[About this AIA Report template 3](#_Toc153546471)

[2. Overview of Key Findings 4](#_Toc153546472)

[2.1 Overall risk profile 4](#_Toc153546473)

[2.2 High-level project summary 4](#_Toc153546474)

[2.3 Key opportunities 5](#_Toc153546475)

[2.4 Summary of potential harms 5](#_Toc153546476)

[2.5 Summary of potential risks to agency 6](#_Toc153546477)

[3. Assessment and approvals 7](#_Toc153546478)

[Author sign-off: 7](#_Toc153546479)

[Reviewer / decision maker: 7](#_Toc153546480)

[Executive Sponsor: 7](#_Toc153546481)

[Appendix 1: Detailed Findings 8](#_Toc153546482)

[Appendix 2: Impact Classification Guide 9](#_Toc153546483)

[Appendix 3: Harm Assessment Guidelines 11](#_Toc153546484)

[Rating Risk 11](#_Toc153546485)

[Appendix 4: Related Impact Assessment Information 12](#_Toc153546486)

[Appendix 5: Glossary 13](#_Toc153546487)

# Purpose and summary

## Purpose

The purpose of this Algorithmic Impact Assessment (**AIA**) Report template is to provide agencies with a template for developing an overall risk rating and reporting the key opportunities, harms, risks and controls identified across the AIA process for a given algorithm project.

This is designed to facilitate informed decision-making on the use of algorithms by government agencies, supporting their adherence to the commitments in the Algorithm Charter.

## Summary of key points

Where an algorithm has been identified as higher risk by the [Algorithm Threshold Assessment](https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-threshold-assessment-questionnaire.docx), the [AIA Questionnaire](https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-impact-assessment-questionnaire.docx) should be completed with reference to and support from the [AIA User Guide](https://data.govt.nz/docs/algorithm-impact-assessment-user-guide).

You can then use this template to prepare an AIA Report that summarises the key findings from the answers in the AIA Questionnaire and clearly articulates the identified harms and risks. That should include:

* A summary of key findings, including the overall risk profile, a high-level project summary, the key opportunities and summaries of the potential harms and risks to the agency concerned
* Assessment and approvals from key decision makers
* A table with more detailed findings that articulate the risks identified through completion of the AIA Questionnaire (with help from the AIA User Guide)
* Key details of any related impact assessment information (e.g. for privacy and/or security).

To help with the risk assessment process, this template also includes guidelines on how harms should be assessed (see Appendix 3) and a Glossary of key terms used (Appendix 5).

## About this AIA Report template

This template is designed to indicate the critical information that should be presented to decision makers and those accountable for an algorithm project. The aim is to give those individuals a clear understanding of the potential opportunities, harms and risks of the project, as well as visibility of the relevant controls.

The report template should be amended as necessary to fit your agency’s needs. In particular, you should amend the risk classifications in Appendix 3 in line with your agency’s overall risk appetite and profile.

Please note that an AIA report should be prepared by someone with a good understanding of risk and the ability to clearly articulate risks and controls. That may be someone in a privacy, legal or other risk function.

You should complex the “Detailed Findings” table in Appendix 1 before inserting the summary in section 2.4 below and finalising the “Overall risk profile” in section 2.1 below.

Please see the Glossary in Appendix 5 for an explanation of the terms used in this document. Terms used more widely across the AIA documentation are defined in the [Glossary in the AIA User Guide](https://data.govt.nz/docs/algorithm-impact-assessment-user-guide/#glossary) .

In the case of a large or complex project, this AIA Report may need to be reviewed a number of times throughout the project lifecycle to ensure its findings continue to be relevant. If applicable, outline any dates or milestones that will be used as a checkpoint to review whether anything significant has changed since the AIA Report was completed.

# Overview of Key Findings

This report summarises the key findings from the AIA process for the [INSERT NAME] project (**Project**) and is based on the responses provided in the [Algorithmic Impact Assessment Questionnaire](https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-impact-assessment-questionnaire.docx).

## Overall risk profile

|  |
| --- |
| Overall, we identified ***[INSERT NUMBER]*** potential harms in the Project. The following Residual Risk ratings are based on the risk assessment criteria in Appendix 2:   * ***[INSERT NUMBER]* critical** * ***[INSERT NUMBER]* high** * ***[INSERT NUMBER]* medium** * ***[INSERT NUMBER]* low** * ***[INSERT NUMBER]* very low *[DELETE THOSE THAT DON’T APPLY]***   Those potential harms are summarised below in section 1.4, with more “Detailed Findings” set out in Appendix 1.  The following controls are **currently** **in place** to manage the risk of the identified harms occurring:   * ***[INSERT]***   The following controls are **planned** to manage the risk of the identified harms occurring:   * ***[INSERT]*** |

## High-level project summary

|  |  |
| --- | --- |
| *Summarise the Project, including its objectives; who will be impacted; key users; status quo comparison; nature and source of training data, production data and the algorithm; algorithm performance to date; ongoing algorithm monitoring and remediation processes.* | AIA Qs 2, 6, 9 |
|  | |

## Key opportunities

|  |  |
| --- | --- |
| *Summarise the Project value proposition, key opportunities, who is likely to benefit and possible best-case scenario(s) arising from use of the algorithm.* | AIA Qs 2.4, 3.1 |
|  | |

## Summary of potential harms

|  |  |  |  |
| --- | --- | --- | --- |
| ***FIRST*** *complete the table in Appendix 1 and* ***then*** *summarise the overall assessment, controls and Residual Risk rating below.* | | | |
| **Domain** | **Potential harms identified** | **Existing and proposed controls** | **Residual Risk rating** |
| **Governance & human oversight** (AIA Q 4) |  |  |  |
| **Partnership with Māori** (AIA Q 5) |  |  |  |
| **Data** (AIA Q 6) |  |  |  |
| **Privacy** (AIA Q 7) |  |  |  |
| **Unfair outcomes** (AIA Q 8) |  |  |  |
| **Algorithm development, procurement & monitoring** (AIA Q 9) |  |  |  |
| **Safety, security & reliability** (AIA Q 10) |  |  |  |
| **Community engagement** (AIA Q 11) |  |  |  |
| **Transparency & explainability** (AIA Q 12) |  |  |  |

## Summary of potential risks to agency

|  |  |
| --- | --- |
| *Select the relevant risks to the agency arising from the Project and/or the identified harms.* | AIA Qs 3.2, 3.3, 4.2, 4.7, 10.2-10.4 |
| *[DROP DOWN MENU FEATURING THE FOLLOWING – MUST ENABLE MORE THAN ONE OPTION TO BE SELECTED:]*   * Negative impact on service delivery or operational performance * Significant negative media coverage * Compliance breaches * Regulatory involvement or intervention, including corrective actions or penalties for non-compliance * Legal costs * Reputation damage – agency * Reputation damage – Government * Loss of public confidence and social licence * Possible independent investigation commissioned * Major recovery action required to restore credibility * Financial repercussions | |

# Assessment and approvals

## Author sign-off:

|  |  |  |  |
| --- | --- | --- | --- |
| 1. *I confirm this AIA Report has been prepared as a result of input from an appropriate multi-disciplinary group involved in the AIA process for this Project and that it accurately represents the identified potential harms and risks and the existing and/or proposed controls considered sufficient to mitigate them.* | | | |
| Name and Position: |  | | |
| Signature: |  | 1. Date: |  |

## Reviewer / decision maker:

|  |  |  |  |
| --- | --- | --- | --- |
| 1. *I confirm that I understand the potential harms and risks of the Project and the extent to which the proposed controls will be sufficient to mitigate them.* | | | |
| 1. Assessment outcome: | 1. *[APPROVE] OR [UNABLE TO APPROVE] OR [APPROVE SUBJECT TO CONDITIONS – please detail]* | | |
| 1. Name and Position: |  | | |
| 1. Signature |  | 1. Date: |  |

## Executive Sponsor:

|  |  |  |  |
| --- | --- | --- | --- |
| 1. *As the person accountable for this Project, I confirm I understand the potential harms and risks and the sufficiency of the proposed controls. I [do] OR [do not] support the Project proceeding as proposed in this document.* | | | |
| Name and Position: |  | | |
| Comments: |  | | |
| Signature |  | 1. Date: |  |

# Appendix 1: Detailed Findings

This section provides further detail of the potential harms and controls identified during the AIA process for the Project and allocates a risk rating for each harm before the application of controls (the “Raw Risk”) and again following consideration of the application of the proposed control (the “Residual Risk”).

Please complete with reference to the Impact Classification Guide in Appendix 2 below, particularly for identifying the “Likelihood”, “Impact” and “Risk Rating”. A sample entry is provided by way of example only.

Table – Harm Details

| **Harm Description** | **Key Harm Drivers** | **Consequence** | **Raw Risk** | | | **Proposed Controls** | **Residual Risk** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Likelihood** | **Impact** | **Risk Rating** | **Likelihood** | **Impact** | **Risk Rating** |
| *E.G. An algorithm to be used during the recruitment process to screen for preferred candidates was trained on data about the agency’s current and previous employees, creating a risk that it could learn to prioritise similar types of people, excluding those of a different ethnicity, gender or physical ability.* | * *The use of training data from current and previous employees only, which risks reproducing historic / societal biases in the context of employment offers.* * *Lack of clear governance and human oversight of the algorithm, including little to no bias or “fairness” testing and insufficient monitoring of bias over the algorithm’s lifecycle.* | * *Discrimination* * *Exclusion from employment* * *Loss of opportunity* | *Likely* | *Significant* | *High* | * *Re-train the model using appropriate, sufficiently representative training data* * *Implement appropriate testing and review mechanisms across the algorithm lifecycle to check for potential biased results.* * *Adopt processes outside the algorithm to encourage a broad cross-section of job seekers, such as making sure job descriptions are appropriately worded to reduce unconscious bias.* | *Unlikely* | *Minimal* | *Very Low* |
|  |  |  |  |  |  |  |  |  |  |

# Appendix 2: Impact Classification Guide

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |  |
| Table 1 – Harm assessment matrix | | | **Impact** | | | | | | | |  |
|  | |  | **Minimal** | | | **Minor** | **Moderate** | | **Significant** | **Severe** |  |
| **Likelihood** | | **Almost Certain** | Low | | | Medium | High | | Critical | Critical |  |
| **Likely** | Low | | | Medium | Medium | | High | Critical |  |
| **Possible** | Very Low | | | Medium | Medium | | Medium | High |  |
| **Unlikely** | Very Low | | | Low | Medium | | Medium | Medium |  |
| **Rare** | Very Low | | | Very Low | Very Low | | Low | Low |  |
| Table 2 – Likelihood rating & criteria | | | | **Table 3 – Impact rating & criteria** | | | |
| **Likelihood of occurrence** | **Criteria:** probability in 12 month period | |  | **Impact of harm** | **Impact criteria *(non-exhaustive & by way of illustration only)*** | | | | | | | |
|  |
| Almost Certain | > 80% | |  | Severe | **Very significant or irreversible consequences:** E.g. excluded from accessing services; significant breach of human rights; significant loss of rights or freedoms; significant negative impacts on specific groups; financial distress; significant loss of agency or autonomy; serious or long-term psychological or physical damage, death (including suicide); significant public health implications | | | | | | | |
| Likely | 60 - 80% | |  | Significant | **Significant consequences and considerable damage:** E.g. inability to access services; serious breach of human rights; loss of rights or freedoms; material negative impacts on specific groups; financial loss; loss of employment; loss of agency or autonomy; negative impacts on mental and/or physical health; public health implications | | | | | | | |
| Possible | 40 - 60% | |  | Moderate | **Significant inconvenience:** E.g. difficulty accessing services; breach of human rights; negative impacts on rights or freedoms; disadvantages to specific groups; extra costs; employment changes; anxiety, confusion, stress; minor physical ailments | | | | | | | |
| Unlikely | 20 - 40% | |  | Minor | **Inconvenience:** E.g. harder to access services; difficulty asserting rights or freedoms; significant inconvenience, frustration or irritation; | | | | | | | |
| Rare | < 20% | |  | Minimal | **Minor inconveniences or no impact** | | | | | | | |

Table 4 – Harm severity & implications

|  |  |
| --- | --- |
| **Harm Severity** | **Implications** |
|
| Critical | Highest priority to mitigate; this level of harm cannot be accepted. |
| High | High priority; this level of harm cannot be accepted without appropriate controls and a formal risk acceptance process. |
| Medium | Moderate priority; this level of harm cannot be accepted without appropriate controls and some form of risk acceptance process. |
| Low | Low priority; this level of harm can be accepted. |
| Very Low | Very low priority; this level of harm can be accepted. |

Table 5 – Risk escalation & reporting

This table sets out who must be informed and has authority to accept risk based on its severity.

|  |  |
| --- | --- |
|  | **Risk Escalation and reporting levels for each level of risk** |
| **Zone 4** | **Chief Executive / Deputy Chief Executive** |
| **Zone 3** | **Senior Leadership Team** |
| **Zone 2** | **Business Owner** |
| **Zone 1** | **Service Manager or Project Manager** |

# Appendix 3: Harm Assessment Guidelines

## Rating Risk

The likelihood and impacts of the identified harms should be rated using the risk matrix in Appendix 2 above, including Tables 2 and 3 setting out the “Likelihood” and “Impact” criteria and rating levels (see further discussion below).

When completing the Detailed Findings in Appendix 1, you should initially assess the identified harms with no controls in place. This will give you the “Gross Risk rating”.

You should then assess the effectiveness of the proposed controls to formulate the “Residual Risk rating”.

### Likelihood Assessment

The scale used to assign a “Likelihood” rating is presented in Table 2 of Appendix 2 above.

It’s likely there will not be readily available information about how often the types of harms under consideration have occurred in the past. That does not necessarily mean the likelihood of the harm occurring is low. Rather, it may indicate the absence of appropriate controls to detect the harm or that the agency has not previously been exposed to the particular risk.

### Impact Assessment

The scale used to assign an “Impact” rating is presented in Table 3 of Appendix 2 above.

All impacts need to be considered and assessed in the relevant business context and be informed by the “Purpose” of the Project defined in the Algorithm Impact Assessment Questionnaire.

The effect of a harmful impact occurring should be assessed using your agency’s own approved risk rating scale, where one is available. If an identified harm has multiple potential consequences, then you should rate the risk according to the consequence with the most significant impact. However, where multiple consequences for a single harm are assessed at the same level, you should consider whether the harm should be evaluated at a higher level than for just a single consequence. For example, a harm with two moderate impacts might be judged to have a significant impact when they are combined.

Rating the impact of a harm must include consideration of any possible knock-on or downstream effects of the consequences of the identified harm, including cascade and cumulative effects and impacts on wider sections of the population.

# Appendix 4: Related Impact Assessment Information

### Privacy Impact Assessment (PIA) - *where required by your Privacy team*

|  |  |  |
| --- | --- | --- |
| Date PIA completed: |  | |
| Status: | New PIA | Update. Date of previous version: |
| Link to PIA report: |  | |
| Overall risk rating: |  | |
| Prepared by:  *Include position & email* |  | |

### Security Impact Assessment (SIA) - *where required by your Security team*

|  |  |  |
| --- | --- | --- |
| Date SIA completed: |  | |
| Status: | New SIA | Update. Date of previous version: |
| Link to SIA report: |  | |
| Overall risk rating: |  | |
| Prepared by:  *Include position & email* |  | |

# Appendix 5: Glossary

|  |  |
| --- | --- |
| **Author** | The person responsible for preparing the AIA Report. That may be a different person to the “Author” of the AIA Questionnaire. |
| **Business Owner** | The person with the accountability and authority to manage a risk. |
| **Control** | A risk treatment implemented to reduce the likelihood and/or impact of a risk. |
| **Harm** | Adverse consequences for people of an algorithm’s deployment and operation in the real world. The AIA process aims to anticipate, identify and avoid potential harms. |
| **Impact** | The effect of an identified harm. |
| **Likelihood** | The chance or probability of a harm occurring. |
| **Raw Risk** | The risk of a harm occurring without any controls being applied. |
| **Residual Risk** | The risk remaining after controls have been applied. |
| **Risk** | The effect of uncertainty on business objectives. In this context, it specifically refers to considering whether a harm might occur and the subsequent impact on the agency in question. |
|  |  |

Please see the [Glossary](https://data.govt.nz/docs/algorithm-impact-assessment-user-guide/#glossary) in the AIA User Guide for an explanation of the other terms used in this template report and the other AIA documents.