Improving algorithm governance in the New Zealand public sector

Johniel Bocacao

April 2024

# Introduction

As the use of algorithms to improve and optimise the operations of businesses increases and becomes increasingly effective, so too do government agencies. Use cases within the New Zealand Government range from simple business rules that triage applications, such as passport and visa applications. Other agencies employ more complex mathematical algorithms that influence how individual cases are handled, such as managing criminals more at risk of offending or targeting health support to improve patient outcomes. Algorithms may also inform policy and funding decisions with even more complex models that integrate data of individuals from different government agencies, such as evaluating the wider social outcomes of certain policy programmes. (StatsNZ, 2018) These algorithms help agencies make evidence-based decisions from the insights unlocked by modern data analytics, potentially reducing the inevitable subjectivity and risk of manual decisions made by humans that can be more prone to error and more difficult to systematically interrogate to understand what (such as biases) affects a decision they reach.

Algorithms are never perfect; there is an inevitable risk of algorithms making incorrect decisions to the detriment of individuals that, for example, miss health interventions after they were needed or are victims of criminals already known to the justice system and were incorrectly managed. The importance of making accurate decisions is compounded for government agencies who have a monopoly on these services, from entry to the country to the legal use of force. Furthermore, a liberal democracy like New Zealand have processes that hold agencies accountable for decisions they make, from hard levers like legislation that facilitate the right to information held by an agency and enforce expectations on how individual information can be used to protect their privacy. There are more indirect soft levers like the risk of losing social license for agencies that consistently make suboptimal or incorrect decisions, to the point where certain communities could accuse an agency of bias against them, or make decisions that disproportionately restrict the rights of individuals. The general public may also hold concerns about the cold automatic nature of machine-based decisions, devoid of the intuition and empathy that comes from a human decision-maker. This risk of losing social license materialises, for example, in media reports of civil society questioning the use of algorithms (Gabel, 2017).

To maintain public confidence in how agencies operate and use individual data, the Government Chief Data Steward (role held by StatsNZ) and the Government Chief Digital Officer (role held by the Department of Internal Affairs) initiated work in 2018 to streamline the New Zealand Government's approach to "ensur[ing] algorithms are used appropriately, are used ethically, and free from bias" (Milicich & Occleshaw, 2018). A survey of government agencies determined how and why they used algorithms at the time to inform a set of principles the assure internal and external stakeholders that algorithms are fit for purpose and meet legal and ethical standards. These principles were embedded in the Algorithm Charter for Aotearoa New Zealand, a non-binding commitment by signatory agencies to comply with said principles, inculcated in each agency through enterprise policies and procedures. The full text is replicated in section 2.3 below.

After 12 months, the two agencies commissioned an independent review to evaluate if its intended aim of improving agency transparency was achieved, ensuring that they weren't stifled in terms of innovation or added compliance burdens. Aspects of the charter excluded from the review included reviewing the text and the voluntary nature of the charter. Findings from interviews and questionnaires were synthesised into common themes and actions, which have been picked up by StatsNZ to implement. This research aims to pick up where StatsNZ left off and seize opportunities to solve already identified gaps in the operationalisation of the Algorithm Charter. This research may also involve identifying opportunities to streamline algorithm development and governance in the New Zealand public sector in general, and critically evaluate how new technological and regulatory developments since the Charter's publication may affect how algorithm governance here should be approached.

## Motivation

At the lowest level of the algorithm development process, the Algorithm Charter requires that "data is fit for purpose by understanding its limitations [and] identifying and managing bias". While this consideration will already be part of the processes of agencies with mature algorithm development capability, and may vary considerably depending on the nature of the business problem, there is a case for providing further guidance in operationalising this principle:

* A standard set of fitness measures, both for the data and the algorithm where applicable, provides a consistent benchmark to readily assure stakeholders (from internal governance advisors to the general public) that an algorithm is fit for purpose.
* A gold standard for measuring fitness can help agencies early in their algorithm maturity to embed best practices during model development and governance.
* Clarifying the importance for monitoring algorithms on an ongoing basis can ensure the data and algorithm continues to be fit for purpose and bias is managed.
* Quantifying bias and fairness is complex multi-dimensional issue. Different definitions of fairness need to be considered and balanced to reach a holistic idea of fairness, such as empirical impartiality, equality and consistent application of an algorithm. Methodologies and tools to help agencies navigate this requirement and make measurement as easy as possible will help agencies grapple one of the most important technical requirements the Charter asks of its signatories.

People and partnership. Segue from talking about biases. Difficult to engage communities as they often don't have the capability to discuss. Information asymmetry, without guidance may lead to mistaken concerns, wrong questions being asked, leading to a worse position, making decisions that lead to worse models. What is the right level of engagement with communities, particularly priority communities like Māori who have Treaty rights to have a say in how decisions made about them are made.

Algorithm specific advice on PHRAE.

Investigate the human oversight principle. Are operators falling to automation bias? How is this verified when there aren't hard controls in place, i.e. trust that the human is reviewing decisions. How the role of humans in automated decisions to mitigate any impact

Linking all motivations is how agencies make these considerations transparent. Different users will need different information: developers as models are refined and monitored, general public, 'gatekeepers' like internal governance bodies, risk managers, priority communities.

Reported difficulty in understanding what’s in scope of 'algorithm'. Where does GenAI fit?

## Research Objectives

Critical analysis looking at the algorithm lifecycle more generally across agencies - critique each step of the way from all perspectives: how developers operationalise Charter considerations (both pre and post prod: data fit for purpose, human oversight, PHRAE), how developers seek approval from gatekeepers, how organisations engage affected communities, how organisations monitor and manage risk on an ongoing basis. [First principles review of the Charter? Do we need to go wider? Maybe a different Generative Algorithm Charter to define expectations when agencies e.g. refine a foundational model for specific purposes.]

Second stage will develop solution(s) that fill identified gaps, solve these problems. [Identify must/should/could measures for monitoring algorithms]. [Both quantitative and qualitative for managing soft risk, both conventional data sci measures and ones called by Charter (human oversight, people, partnership)]. [Provide guidance on complex issues like quantifying bias].

[As seamless as possible, integrate into existing data science pipelines]. [Consistent pattern general enough to apply to all agencies, modular enough that measures that may not apply to a problem can be explained away]. [Omission of this regular, seamlessly produced artefact can easily call into question an agency's commitment to transparency]. [So easy that inability to produce the artefact will call into question an agency's ability to explain a model, let alone observe or monitor it].

[Flexible enough that it can serve mutliple audiences]. [Data scientists to compare model options, performance/compliance]. [Communities to guide their discussions around what an algorithm means for them, how it performs for them vs others/all]. [Gatekeepers to provide a framework around discussions when evaluating compliance]. [General public to easily understand that government uses their data, runs algorithms appropriately through a publicly accessible, central register].

# Background

## Algorithms in the Public Sector

* Define what’s in scope of AC
* Examples of algorithms, what’s not an algorithm
* Generative AI - pending question
* Interim Generative AI guidelines for NZ Government but only for services - not for GAI dev itself.

## Measuring Algorithmic Fitness

* Specific guidance will help agencies identify and mitigate the risk of data drift: when the underlying structure of fresh input data changes as the algorithm remains constant; and concept drift: when the relationship between the input data and the decision changes.
* Fairness and bias.
  + Modify data before trianing
  + Modify algorithm trained
  + Modify predictions of model
  + Root cause analysis, e.g.
* At the most basic level, one aspect of fairness may be impartiality in ensuring that a model is trained empirically on the available evidence base. Another aspect is the definition of equality, be that equality of assessment (same threshold across all subpopulations), opportunity (same true positive rate), outcome (same accuracy) or odds (same true and false positive rate). Even when bias is minimised from the underlying data and algorithm, the way that algorithm is employed can still introduce unfairness, requiring broader process design considerations.

## Algorithm Charter of Aotearoa New Zealand

* History and development, who did Stats interview
* Find suggestions from submissions made
* Year 1 review, parameters, limitations.
  + Interviews from agencies with high-risk algorithms
  + Interviews with civil society: academics, peak bodies, MDS/Tiriti experts
  + Questionnaires to analytics Tier 2s, non-signatory agencies.
  + Agencies were excluded on the basis of minimal use of algorithms.
  + Out of scope: amendments, private sector, \*\*outcomes\*\*.
* Outstanding and actioned recommendations from the Year 1 review
* **Transparency** - Maintain transparency by clearly explaining how decisions are informed by algorithms. This may include:
  + Plain English documentation of the algorithm
  + Making information about the data and processes available (unless a lawful restriction prevents this)
  + Publishing information about how data are collected, secured and stored.
* **Partnership** - Deliver clear public benefit through Treaty commitments by:
  + Embedding a Te Ao Māori perspective in the development and use of algorithms consistent with the principles of the Treaty of Waitangi.
* **People** - Focus on people by:
  + Identifying and actively engaging with people, communities and groups who have an interest in algorithms, and consulting with those impacted by their use.
* **Data** - Make sure data is fit for purpose by:
  + Understanding its limitations,
  + Identifying and managing bias.
* **Privacy, human rights and ethics** - Ensure that privacy, ethics and human rights are safeguarded by:
  + Regularly peer reviewing algorithms to assess for unintended consequences and act on this information.
* **Human oversight** - Retain human oversight by:
  + Nominating a point of contact for public inquiries about algorithms,
  + Providing a channel for challenging or appealing of decisions informed by algorithms,
  + Clearly explaining the role of humans in decisions informed by algorithms.

## Related Regulation

Regulation for the purposes of this research comprises of hard levers of setting the rules. Legislation

### Official Information Act

Freedom of information around government decision making. Is it

### Privacy Act

Comply with information privacy principles. Make available any information an agency holds on a requesting individual.

## Related Frameworks

Frameworks are softer levers for setting rules, usually within an agency or a community of agencies that

### Algorithm Impact Assessment

Drafted as a result of Year 1 review for assessing algorithm risk at the pre-deployment phase, including whether it is high risk enough to fall under the Charter.

### Internal Agency Frameworks

* PHRAE - MSD impact-side evaluation prior to deployment
* MDL - MSD framework for model development and maintenance
* DPUP - SWA
* StatsNZ IDI - Nga Tikanga Paihere. Stats framework for applying tikanga Māori to data use

### Māori Algorithmic Sovereignty

As per (Brown et al., 2023), how they were transformed from MDS principles already circulated to agencies.

## Related Efforts

### Centre for Data Ethics and Innovation

### Algorithm Transparency in Open Government Partnership

Probably where we talk about public participation, IAP2 Spectrum.

## Related International Efforts

* Canadian directives on responsible AI
* Apply EU AI Act ideas in a non-commercial context
* Biden administration EO, guidance to federal agencies on AI use

# Timeline

|  | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Analysis | Lit rev | X | X | Analysis | X |  |  |  |  |  |  |
| Interview dev |  | X |  |  |  |  | X |  |  |  |  |
| Primary research |  |  | Discov | X |  |  |  | Testing | X |  |  |
| Ideation |  |  |  | X | X |  |  |  | X | X |  |
| Solution devt |  |  |  |  | X | X | X |  |  | X | X |

Table 1 - Proposed Gantt chart (column represents 2 months)

## Critical Analysis (10 months)

Critical analysis of the state of play across the breadth of the Charter? Frame as not just a review but a contribution in its own right

### Interview development (2 months)

Write up questions and script, devise sample

### Primary research (4 months)

Undertake interviews

### Ideation and Synthesis (overlapping 4 months)

Integrate learnings from research

## Solution development (overlapping 14 months)

Similar waterfall structure: develop tool, devise and perform user tests, integrate feedback, finalise solution. This may be around quantification of bias, developing a model register API/framework around what measures/stats needs to be published.

# Resources

## Interview Participants

Who is in and out of scope for interviewing, how I’ll build a representative sample. Ethical considerations.

## Trial Data for Solution Development

What data I’ll be using to test solutions

# Complimentary Material

## Past Experience

# Bibliography

StatsNZ, D. O. I. A.. (2018). Algorithm assessment report 2018 - data.govt.nz. https://data.govt.nz/toolkit/data-ethics/government-algorithm-transparency-and-accountability/algorithm-assessment-report/. 2024-04-08

Gabel, J.. (2017). ACC’s computer-aided decision-making questioned by Otago experts. https://itbrief.co.nz/story/accs-computer-aided-decision-making-questioned-otago-experts. 2024-04-08

Milicich, R. & Occleshaw, T.. (2018). Report: Review of Government Algorithms. https://data.govt.nz/assets/Blog-files/Review-of-Government-Algorithms-Report-14-May-2018-for-release.pdf. 2024-04-08

Brown, P. T., Wilson, D., West, K., Escott, K. R., Basabas, K., Ritchie, B., Lucas, D., Taia, I., Kusabs, N. & Keegan, T. T.. (2023). Māori algorithmic sovereignty: idea, principles, and use. arXiv:2311.15473