

# Towards trusted and trustworthy automated decision making in Aotearoa

What we can learn from the Digital Council's 2020 research

**Professor Anna Brown**

Toi Āria: Design for Public Good  
College of Creative Arts  
Massey University  
Wellington, New Zealand

**Anna Pendergrast**

Co-Lead, Antistatic



**antistatic**

# Established to provide independent advice to the NZ government on how to maximise the societal benefits of digital and data-driven technologies

The Digital Council for Aotearoa New Zealand was a ministerial advisory group established by the Ministers of Government Digital Services and Statistics in early 2020. The Council wrapped up in late 2022.

---

There were three areas of focus in its work programme:

trust (2020)

digital inclusion (2021)

innovation (2022)

# We were team members for the Digital Council's 2020 research project into trusted and trustworthy automated decision-making

A team effort:

**Digital Council research leads:**

Marianne Elliott and Colin Gavaghan

**Participatory Research Lead:**

Toi Āria team, led by Anna Brown

**Report writing and supporting policy analysis:**

Antistatic, Anna and Kelly Pendergrast

**Literature review and supporting advice:**

Brainbox, Tom Barraclough and Curtis Barnes

**Secretariat support from DIA**

**Māori and ADM research paper and wānanga:**

Te Kotahi Research Institute



# Undertake a participatory design engagement focussed on automated decision-making (ADM) — specifically systems where parts of the process are carried out by computer algorithms

**Aim of the research:** to complete a literature review and to engage with specific publics to understand their perceptions and levels of trust in digital and data-driven technologies.

**Why:** to orient government and public applications of such technologies for a digital future centered on the needs and aspirations of people in Aotearoa.

# Understanding trust

## Exploring definitions and terms:

- » The process of defining 'trust' was complex and unresolved. For this project, when we talk about trust, we are talking about people feeling comfortable and confident when they are affected by other people's decisions
- » Te Reo Māori doesn't have equivalent kupu
- » Trust is not sufficient on its own, needs to be trustworthy too



# Harnessing design for positive social change through effective community engagement

[www.toiaria.org](http://www.toiaria.org)

- We aim to do [some] good
- We nurture partnerships with people and communities
- We build relationships

TOI ĀRIA

OUR APPROACH OUR PROJECTS OUR PEOPLE OUR STORY

## Our Projects

Case studies, reports, research papers and more

We help organisations and communities bridge the gap between organisational aspirations and real-life challenges. Through partnerships we work with communities on the things that matter to them. Whether it's housing, community comfort, or data, we want to go where the need is.

Our team has many tools and approaches for facilitating conversations with affected communities. Our aim is to collaborate with organisations and communities to understand and design how best to incorporate the views, insights and experiences of those most affected by the tools, services or policies being developed.

ALL CASE STUDIES PUBLICATIONS

**CASE STUDY**  
**Towards trustworthy and trusted automated decision-making in Aotearoa**  
 Understanding the trust required to harness the full societal benefits of digital and data-driven technologies

**PUBLICATION**  
**Outlook for earthquake early warning for Aotearoa New Zealand**  
 Insights from initiating a community-of-practice

**CASE STUDY**  
**Better Later Life — He Oranga Kaumātua**  
 Improving interactions with government services for New Zealanders over the age of 65

**PUBLICATION**  
**Current Co-design in Aotearoa Literature Review**  
 A snapshot review of scholarship and literature about co-design in Aotearoa New Zealand

**CASE STUDY**  
**Te Mauri Moemoeā — A Whānau Dreaming App**  
 Designed for and with rangatahi and their whānau — that encourages and guides dreaming big

**PUBLICATION**  
**Toward Algorithmic Accountability in Public Services**  
 A Qualitative Study of Community Perspectives on Algorithmic Decision-Making in Child Welfare

**CASE STUDY**  
**Our Data, Our Way**  
 What New Zealand people expect from guidelines for data use and sharing

**PUBLICATION**  
**Digital Contact Tracing for COVID-19**  
 A primer for policymakers on how digital contact tracing works

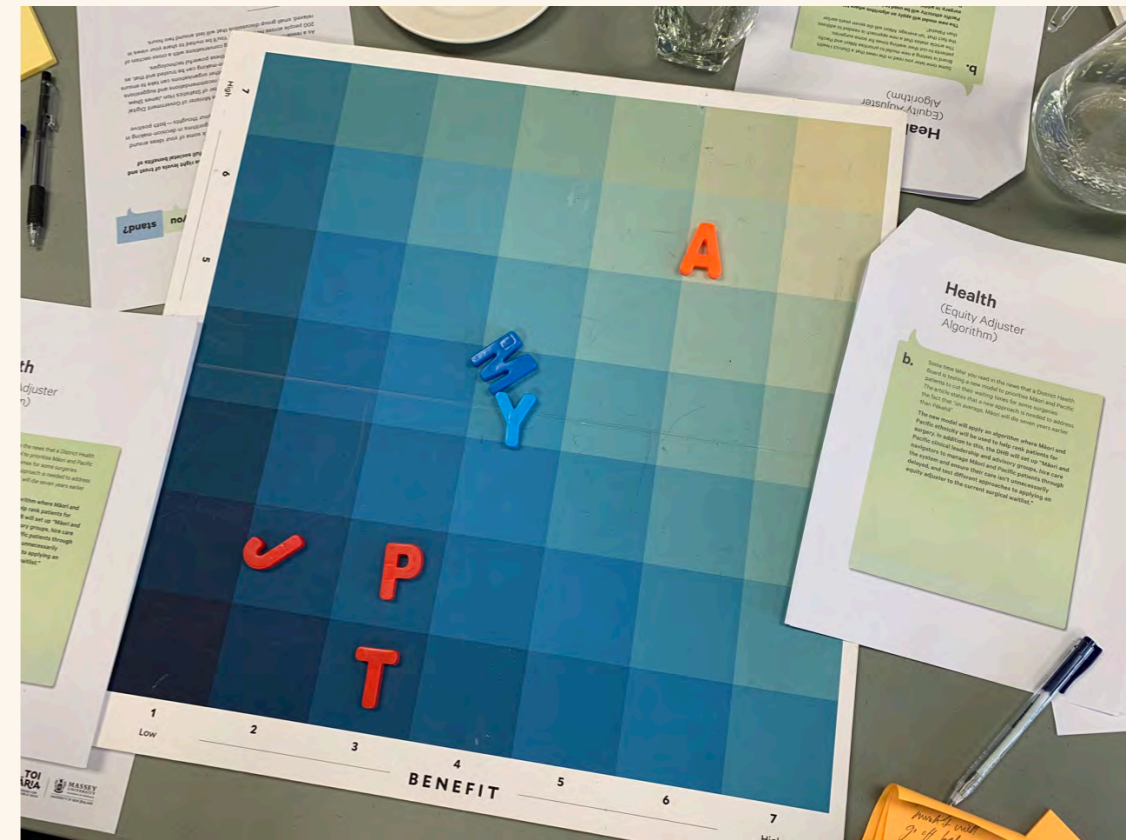
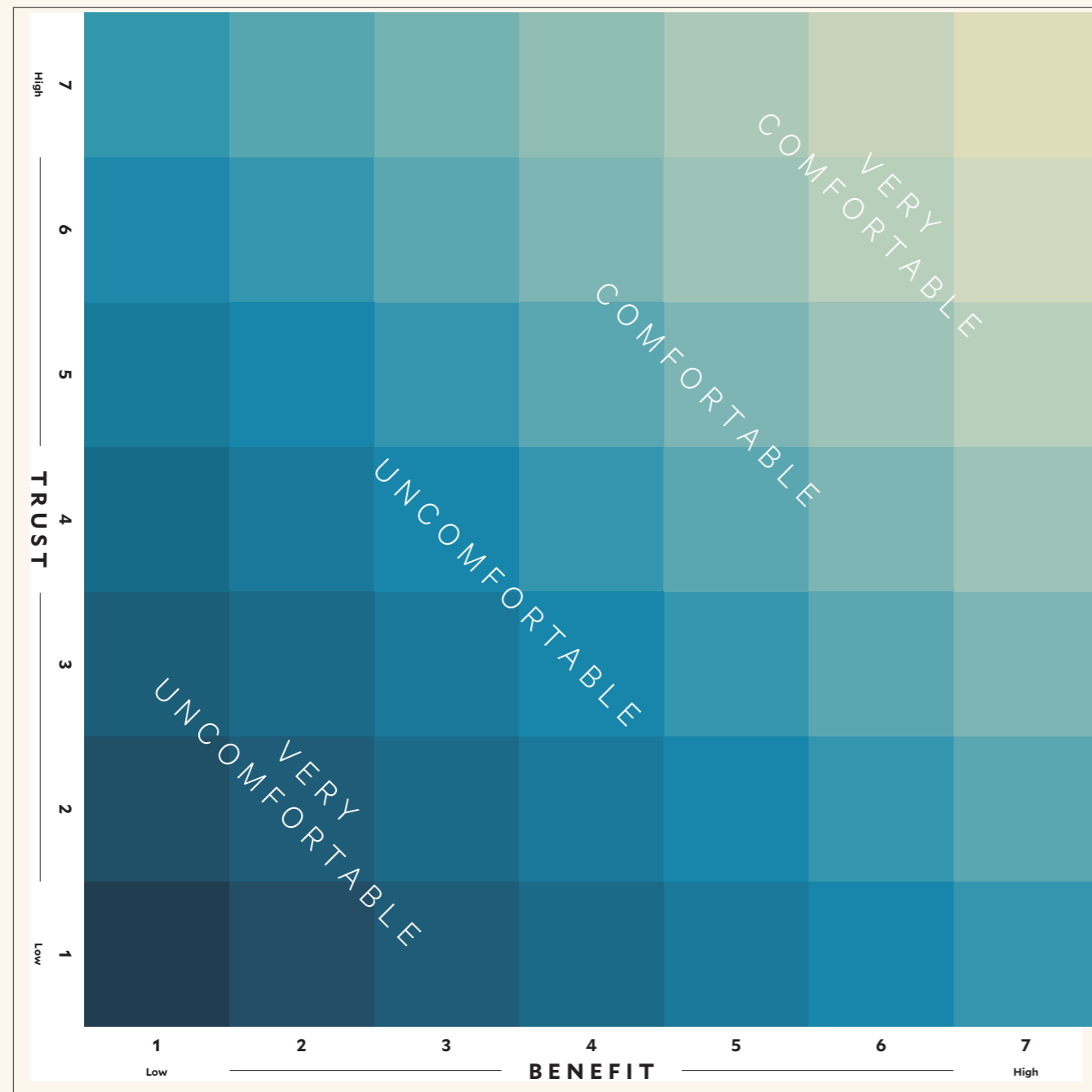
**CASE STUDY**  
**Whakatutuki Moemoeā — Whānau Fulfilling Dreams**  
 Pioneering tāne and whānau to work on their goals and participate in the redesign of services

## Deliberative decision-making

*'In Participatory Design the people destined to use the system play a critical role in designing it.'*

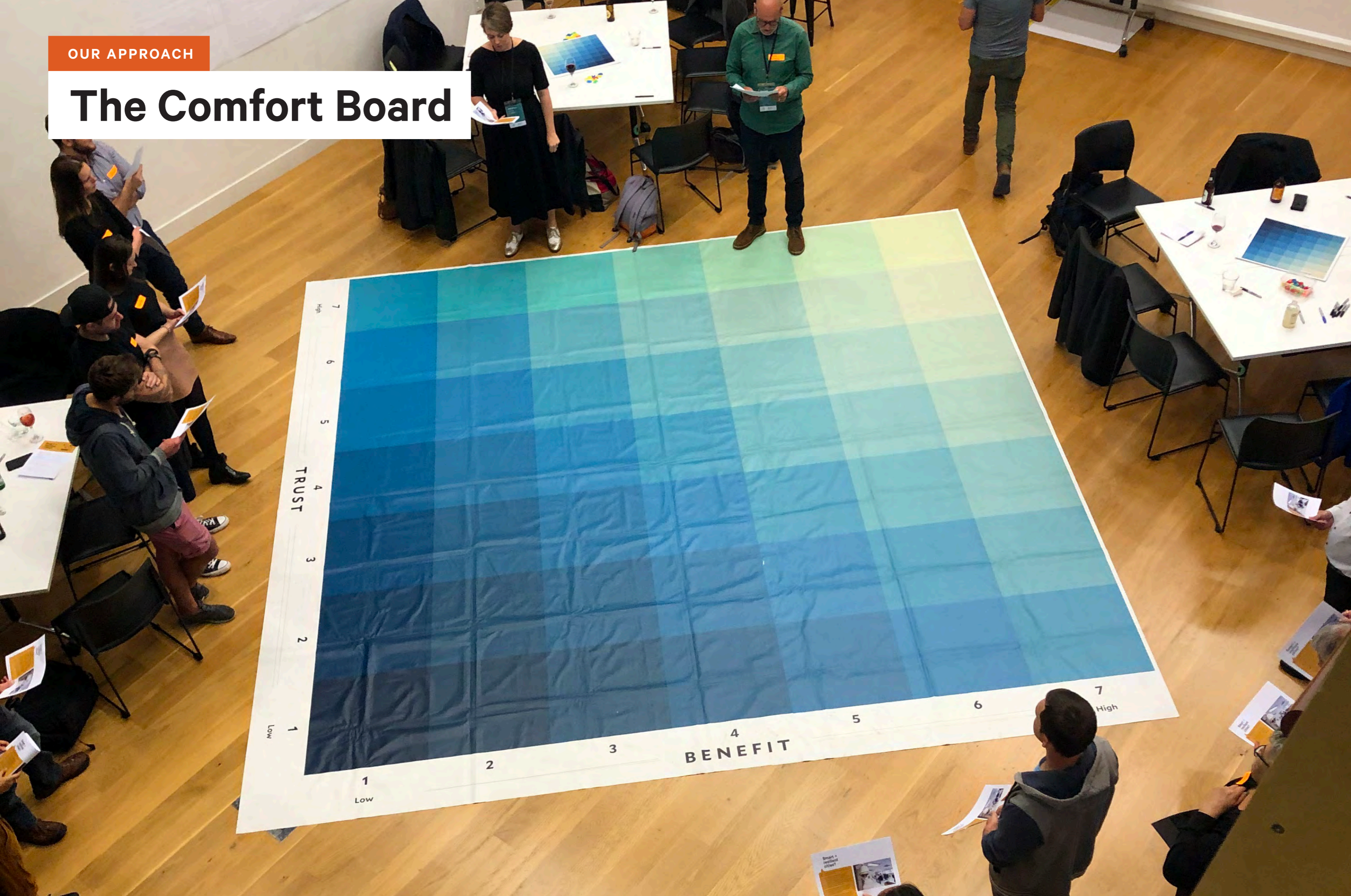
— SCHULER & NAMIOKA, 1993

# The Comfort Board: A participatory methodology for effective citizen engagement





# The Comfort Board





# The Comfort Board





# The Comfort Board





# Who we spoke to

## Workshop attendees included:

- » blind and vision impaired
- » ethnic community leaders and youth
- » Māori and Pacific youth
- » women with migrant and refugee backgrounds
- » Whānau Ora navigators (Māori health advocates)
- » young people with care experience
- » and members of the general public.



# Workshop prototype





# Workshop with Pacific Youth



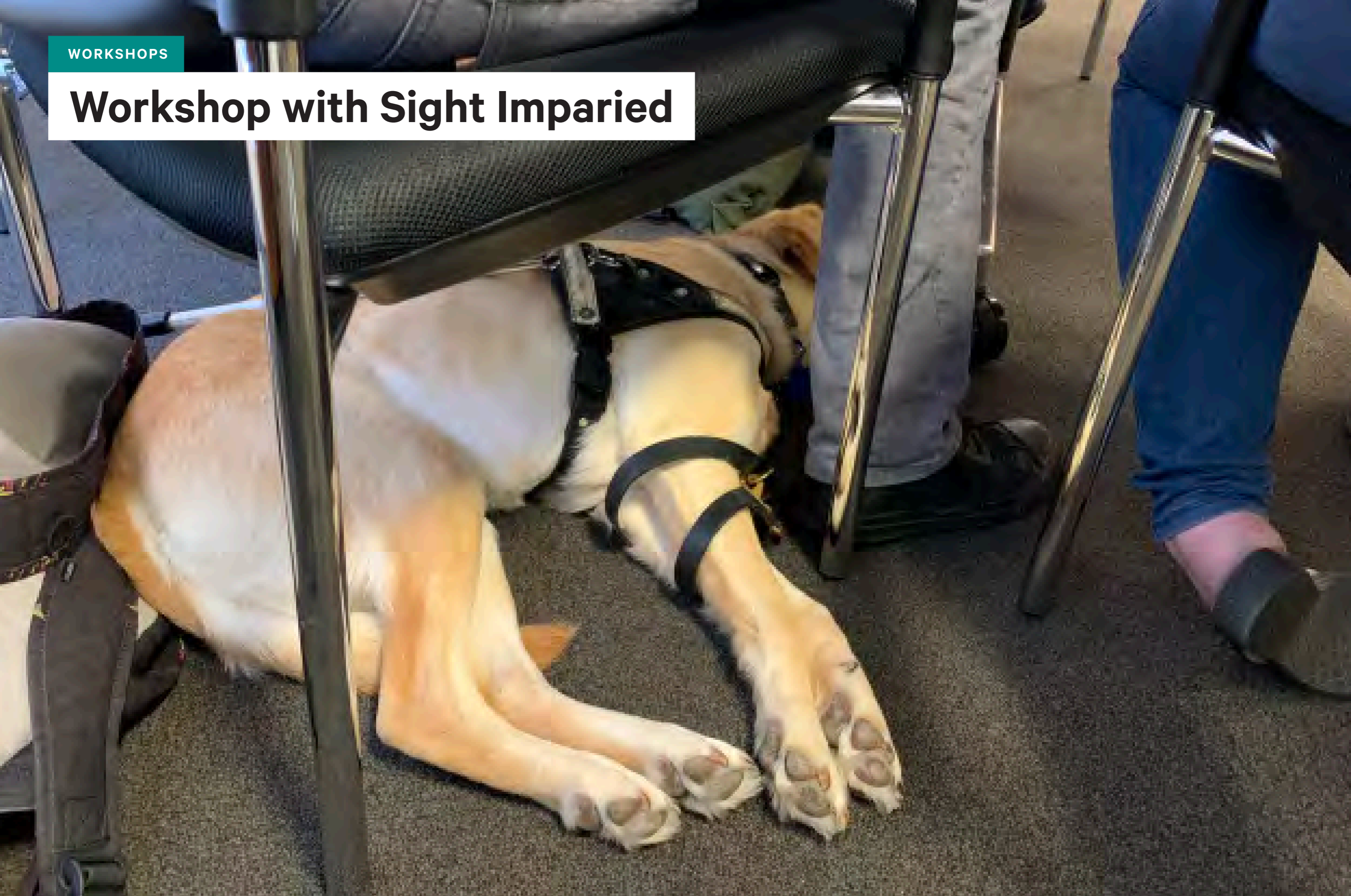


# Workshop with Māori Youth





# Workshop with Sight Impaired





# Scenarios

## Recruitment scenario: Search engine and CV filtering algorithms

### Recruitment

(Search Ranking  
Algorithm)

**a.**

You have been employed at your job for a number of years and feel you're ready to step up into a new role. You go online and pull up a search engine (e.g. Google or Yahoo). In the search bar you type in relevant keywords, for example 'truck driver jobs hamilton', 'senior corporate jobs wellington', or 'health worker job vacancy christchurch'. Within seconds the results are pages of listings for job and career websites, recruitment agencies and vacancies from various organisations. Ads feature first and are followed by general listings.

**The search engine uses an algorithm to sift through millions of websites to give you useful and relevant listings. It considers factors such as keywords, site-speed, your location, and usability to determine which sites are likely to give you the answer you are looking for.**

In the process of searching for a job you visit a number of different websites for information on how to write a CV, locations of organisations, tips on job hunting, and how to prepare for an interview. **This activity is automatically recorded by the search engine. It also records any information you provide whenever you sign up to a digital account such as your name, email and date of birth.**

### Recruitment

(Talent Selection  
Algorithm)

**b.**

You see a vacancy online that you are interested in and submit a CV with the organisation. As part of their internal filtering process **the organisation uses an algorithm to help them efficiently process the 167 applications they receive and ensure they get the right candidate for the job.**

**The algorithm not only analyses the information on your CV but also phrases and words, including sentence length, paragraphs, and key words, in order to delve deeper into the analysis of you as a candidate. From here it makes a recommendation on your likely initial match for the role which will determine if you are put forward to the next stage of the recruitment process.**

You are identified as a desirable candidate and are successful in getting through to the next stage of the employment process.

## Participant responses:

**“If it is just matching hard skills, then I trust it because there isn’t too much room for interpretation. Like software or if you’re prepared to work in a specific location. But I don’t trust it if it is judging softer skills.”**

**“The algorithm kind of just shuts out a certain group of people and really lets in a certain group of people. So even when they get to the interview process you’re interviewing the same people rather than a different diverse group of people and taking employees or people from different parts of society that would really benefit the system. You’re missing opportunities.”**



# Scenarios

**Youth Support scenario:**  
Based on Not in Education, Employment or Training (NEET) algorithm currently used by Ministry of Social Development to identify and provided support to young people.

## Youth Support

(Not in Education, Employment or Training: NEET Algorithm)

**a.**

Your 16 year old daughter Nikki, who has recently left school, receives a call from a local community worker offering her help as an early school leaver. The worker says Nikki has been flagged as being highly likely to benefit from help in accessing education, training or work-based learning. Nikki is being offered the services of a youth coach to assess her needs and develop a plan including education and training. It's up to Nikki as to whether she accepts this assistance or not.

You and Nikki check out the website provided.

The community worker is part of the Ministry of Social Development's (MSD) youth service called NEET — 'Not in Education, Employment or Training' — designed to proactively identify young people who are missing out on basic education and training opportunities, and to offer them support.

**These young school leavers are identified by MSD using an algorithm which draws on a range of information including Nikki's education records (achievement, truancy history and reason for leaving school), you and your partner's MSD benefit histories, and any Oranga Tamariki notifications. The algorithm produces a risk score linked to the likelihood of Nikki needing to be on a benefit as a young person. Young people with high risk scores are referred to community providers like the one that contacted your daughter.**

## Youth Support

(Not in Education, Employment or Training: NEET Algorithm)

**b.**

Some time later you read in the news that the NEET scheme is checked regularly to measure how accurately it identifies young school leavers who are actually in need of support.

**A recent check reported that there is a 79% chance that the risk score reflects the actual level of risk. MSD describes 79% as a level of algorithmic accuracy which is 'generally accepted as good accuracy'.**

An evaluation of the programme writes: "We find that NEET raised the educational retention of participants in the first year, by up to 9%. This positive impact is sustained for around one year. Further qualification achievements however are quite modest, at 2%. There is no improvement in their likelihood of being employed, and the number of times they accessed the benefit were slightly raised."

## Participant responses: Youth support

**“I think that there’s a risk with government-collected data — particularly when you’re dealing with Māori or Pacific communities or any communities that have intergenerational distrust of the crown — that the data that they have will contain lies because people don’t trust the agents of the crown enough to tell them the truth.”**

**“I’d like every single level of the various Ministries to have a person in that role who is Maori, and that person has equal decision-making power and funding and they also have their own team of researchers who work from a Kaupapa Māori perspective when collecting data.”**



# Scenarios

**Immigration scenario:**  
Based around a risk analysis algorithm used by Immigration New Zealand to inform decisions about visa applications.

## Immigration

(Risk Assessment Algorithm)

**a.**

Your friends, a sister and brother, are hoping to move to New Zealand to live. Both are university educated, neither have criminal records, and they are well off. They apply to Immigration New Zealand in their home country to immigrate to New Zealand.

**New Zealand's immigration system applies an automatic triage system (a risk score algorithm) which assesses all visa applicants using risk rules. Regardless of the risk score, every application has a case manager involved in assessing it. The risk score determines how much work a case manager needs to do — for instance, a high score may mean more in-depth verification of the applicant's documents when determining whether or not to issue a visa. The algorithm also reduces the time it takes to process a visa application.**

The sister gets an entry visa and the brother is refused entry to New Zealand.

## Immigration

(Risk Assessment Algorithm)

**b.**

The brother appeals to the New Zealand Immigration and Protection Tribunal, providing them with detailed information about his specific situation.

**The Tribunal assigns a case manager who assesses the brother's appeal and checks if Immigration New Zealand's algorithm has assessed him correctly. The criteria the algorithm applies is deemed correct, however, the Tribunal finds there are good reasons to make an exception and approve the brother's appeal.**

The brother is told to resubmit his appeal application and provide Immigration New Zealand with the additional information. His visa is subsequently approved.

# Participant responses: Immigration

**“If it is a simple yes/no task, I want the algorithm to do a perfect job. But if it is required to consider nuanced stuff, I want a human in the mix.”**

**“I want to know what’s going in the algorithm. I trust an algorithm, it is just a computer making decisions that a human would make from a checklist. It’s the decisions that I’m dubious about not the fact that a computer’s making them.”**



# Scenarios

**Criminal Justice scenario:**  
Based on the risk of ReConviction x Risk of Imprisonment algorithm used by the Department of Corrections, which uses a risk score to inform a range of decisions.

## Criminal Justice

(Risk of ReConviction  
x Risk of Imprisonment  
Algorithm — ROC\*ROI)

**a.**

You read an article from your local news provider about a system the Department of Corrections uses to assess the risk of re-offence and imprisonment of criminal offenders.

**The ROC\*ROI algorithm (Risk of ReConviction x Risk of Imprisonment) has been used in New Zealand since 2001 to produce a risk score about the probability of an offender re-offending within a certain period following their offence and is based “on the behaviour of the individuals in the underlying data set”. The score uses variables such as age, sex, frequency of offence, severity of crime and time spent in prison to produce a risk scale of low, medium or high for an offender. It does not use ethnicity data.**

ROC\*ROI has been used because research has shown that even simple risk scales (i.e. a checklist of risk factors) invariably outperform the clinical or professional judgements of trained experts and experienced correctional staff when making predictions about future offending.

The risk score alone does not decide the fate of the offender but allows Corrections and Parole staff to use this information alongside other data to make decisions about future prison costs, eligibility for rehabilitation programmes and whether an offender should receive bail or parole.

## Criminal Justice

(Risk of ReConviction  
x Risk of Imprisonment  
Algorithm — ROC\*ROI)

**b.**

Later that week you read a response to the same article written by a collective of social workers and researchers who review the ROC\*ROI tool against common definitions of fairness. They point out that the **statistical fairness of the tool can be challenged because “if some groups are subject to more surveillance and conviction rates than others, then this skews the data used to inform the predictions.”**

The article also discusses the social fairness of the ROC\*ROI tool, suggesting that **“while ethnicity is not used in the ROC\*ROI algorithm, every other variable such as age at first offence, frequency of conviction, and number of convictions will over-identify Māori as being at high risk.”**

The article concludes by questioning whether the ROC\*ROI tool is a statistically and socially fair way of informing decisions about eligibility for bail, parole and rehabilitation programmes.

# Participant responses: Criminal Justice

**“I think that data, whether it’s negative or positive, can always tell you something and you can always draw a conclusion based on that data. However, everybody has the potential to change. I just think that data can be used to help but data shouldn’t be used to weigh so heavily on the decisions of people’s lives.”**

**“I wouldn’t trust this to make a decision about someone going to prison. It’s just comparing them to other people who have done the same thing. But everyone’s different. They should assess you on who they are, not the crime you’ve done. I don’t just trust the Justice system. They are not making this personal. You can’t get the system out of this thing — it’s part of it.”**



# Scenarios

## Health scenario:

Based around a nationally-recognised algorithm used by District Health Boards to inform decisions about waiting list priority.

## Health

(Waiting List Priority Scoring System)

**a.**

You are on the waiting list for a kidney transplant.

You have been informed by your local District Health Board **that operations such as yours are prioritised by a nationally-recognised algorithm which gives you a priority score based on how urgently you need the surgery and how much you will benefit from it compared to other people.**

You have recently advised your specialist at the hospital that your condition has worsened, but you're still waiting for an operation date. You also know a family friend who is awaiting the same procedure — and you're both surprised when they receive confirmation of their operation, because their situation does not seem as urgent as yours.

## Health

(Waiting List Priority Scoring System)

**b.**

Some time later you read in the news that a District Health Board is testing a new model to prioritise Māori and Pacific patients to cut their waiting times for some surgeries. The article states that a new approach is needed to address the fact that “on average, Māori will die seven years earlier than Pākehā”.

**The new model will apply an algorithm where Māori and Pacific ethnicity will be used to help rank patients for surgery. In addition to this, the DHB will set up “Māori and Pacific clinical leadership and advisory groups, hire care navigators to manage Māori and Pacific patients through the system and ensure their care isn't unnecessarily delayed, and test different approaches to applying an equity adjuster to the current surgical waitlist.”**



# Participant responses: Health

**“If it’s real-time data then I’d trust an algorithm over a human. People can get stressed, tired, emotional and irrational, and doctors are no exception.”**

**“It feels like an ethical dilemma. What happens when you have two people with the same condition, and one person is 80 years old and one is 10 years old. How does it decide? What are the criteria, and who decides those? It comes down to serious ethical and human rights principles.”**



# Some key themes emerged

## Participants told us they:

- » think of algorithms as just one part of much larger systems
- » see clear benefits of ADM for perceived 'low risk' tasks at speed and scale or acting as an 'assistant' to get work done
- » have lower comfort when algorithms were used to inform complex decisions that had major impacts on people's lives
- » think algorithm systems can make existing problems like bias and discrimination worse.
- » **Many participants had low trust in organisations using algorithms, especially government agencies.**

# Increasing comfort levels

## Suggestions for increasing comfort included:

- » understanding and addressing the specific needs of affected communities in the development and use of ADM systems
- » involving people with a diversity of experiences, cultures and world views in the design, development and deployment of algorithm systems,
- » being transparent including:
  - transparency of the wider system
  - transparency of data use
  - transparency of criteria
  - transparency of the algorithm
- » clear, open communication
- » opportunities to ask questions, get clarification, and give feedback
- » personal, or kanohi ki te kanohi (face to face) communication and relationship-building as part of decision-making systems



# Increasing comfort levels

## **Additional comfort increasers included:**

- » clear rules, standards, legislation or frameworks to govern ADM use, as well as capacity to assess whether they are implemented properly and enforce them where required
- » treating data with care and consideration
- » making sure algorithms and wider decision-making systems are treated with care and ongoing attention — they should be effectively monitored, tested, and maintained
- » and building capability and skills around algorithms and related digital processes.

# Recommendations

## **The Council's seven recommendations fell into four broad groups:**

- » actions that will give effect to Te Tiriti o Waitangi across government ADM and data projects
- » on-the-ground projects that test new ways of doing and thinking about things, demonstrate value and build a knowledge base to enable successes to be scaled and replicated
- » projects that bring clarity and cohesion to work already underway on ADM projects and fill any gaps in frameworks and guidance that are needed
- » projects that focus on building digital skills and knowledge and increase the diversity of the digital workforce.



# Recommendations

**Recommendation 1:** Fund community groups to lead their own data and automated decision-making projects.

**Recommendation 2:** Fund and support a public sector team to test and implement automated decision-making best practice.

**Recommendation 3:** Establish a public sector automated decision-making hub.

**Recommendation 4:** work collaboratively to develop and implement private sector automated decision-making rules or best practice.

**Recommendation 5:** Build automated decision-making systems from te ao Māori perspectives.

**Recommendation 6:** Build a diverse digital workforce.

**Recommendation 7:** Increase the digital skills and knowledge of public sector leaders

# Limitations of this approach

## Issues with participatory approaches and impact

- » issues of public engagement processes vs reality of political cycle and routes for implementation in government systems (churn)
- » risk of engagement with no perceptible change or improved outcomes
- » risk of engagement fatigue or 'repeat' audiences, repeated work
- » challenge of building engagement to suit public audiences not government need
- » Issues of language: ADM → algorithms
- » issues of expert views (policy, technical) vs public views (lived experience expertise)
- » issues of positionality (who does the research?)



# Value and impact (direct or indirect)

## Impact of this work to date?

- » exemplar of participatory approaches and evidence of best practice for what increases comfort for communities
- » Recommendations not directly implemented as suggested, there are however some initiatives that address similar themes but many of the issues still persist
- » Interest from international counterparts and research groups:
  - Centre for Data Ethics and Innovation, UK
  - Essay in New\_Public magazine, USA: ‘Can New Zealanders trust an automated government?’
- » The Australia and New Zealand School of Government (ANZSOG) Digitisation, Design and Policy Master Class (2022-2024)
- » BWB ‘More Zeros and Ones: Digital Technology, Maintenance and Equity in Aotearoa New Zealand, including chapter on ‘Participatory Design for a Digital World’.

# What can we apply to today's AI environment?