Use of AI by New Zealand Government

Colin Gavaghan

Artificial Intelligence and Law in New Zealand

A three-year project to evaluate legal and policy implications of artificial intelligence (AI) for New Zealand. The project is based at the University of Otago, and funded by the New Zealand Law Foundation.

A two-year postdoc for the second phase of the project (AI and Employment Law) will be advertised at the end of 2017.

Project Aims

Artificial intelligence (AI) is coming at us before we fully understand what it might mean. Established ways of doing things in areas like transport regulation, crime prevention and legal practice are being challenged by new technologies such as driverless cars, crime prediction software and 'AI lawyers'.

AI technologies pose fascinating legal, practical and ethical challenges, which require interdisciplinary solutions.

In our project, we will investigate two topics that link AI and the law, and study their implications for New Zealand.

1. **Predictive AI technologies in the criminal justice system**
   AI systems can learn to use experience of the past to make predictions about the future. Such predictive systems can be used by police forces to help decide how to allocate resources on a given day, or even to target particular individuals. They can also be used in the courts, to assess the likelihood of a plaintiff reoffending, or even of an individual committing a first offence. These systems are already in use in some countries, but there are complex issues surrounding their adoption.
   - Can we ensure these systems' decisions are transparent and trustworthy?
   - Might these systems contain implicit bias towards certain groups?
   - Might human users become over-reliant on such systems?
2. **AI and Employment Law**
   There is much current discussion around the topic of 'technological unemployment' - the prospect that people will lose their jobs to intelligent machines. This prospect raises many legal questions.
   - How would such job losses fit within existing categories of redundancy and unfair dismissal? Are any changes needed to employment law to cater for this scenario?
   - In professions where employees have a social role as well as a practical function (e.g. law, medicine, education), is there a danger that replacing human employees with machines will erode an important component of the work? If so, how might this be safeguarded?
   - If intelligent machines are employed by companies, might we need legal mechanisms for defining their obligations and rights? Should they perhaps be regarded as 'legal persons', for some purposes?

AI POLICY – UNITED NATIONS

OECD.AI
Policy Observatory

European Commission › Strategy › Shaping Eurc

Shaping Europe's digital future

POLICY

Artificial Intelligence

WORLD ECONOMIC FORUM

Reimagining Regulation for the Age of AI:
New Zealand Pilot Project

WHITE PAPER
JUNE 2020

The Digital Council for Aotearoa

MITCHELL PHAM

ROGER DENNIS

MARIANNE ELLIOTT

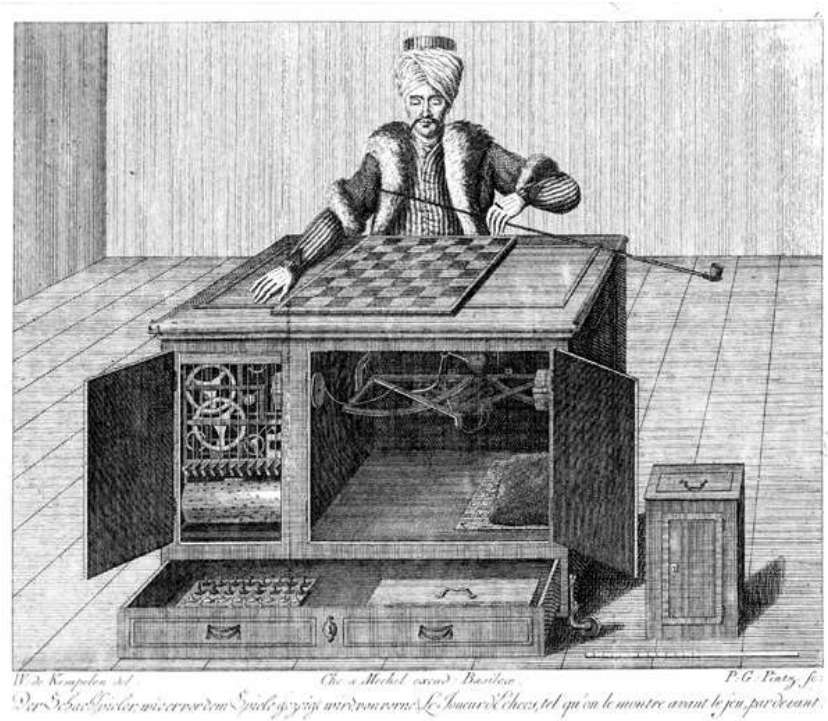
KENDALL FLUTEY

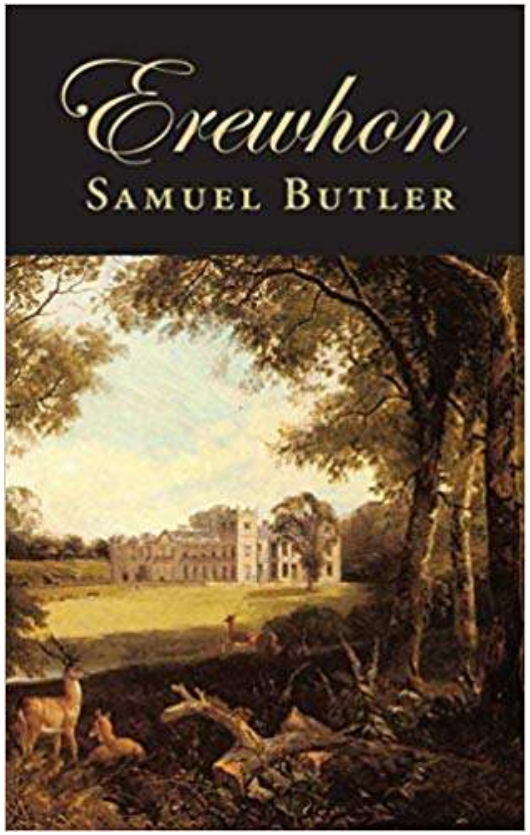
RACHEL KELLY

NIKORA NGAROPO

COLIN GAVAGHAN

AI: When did it all begin?





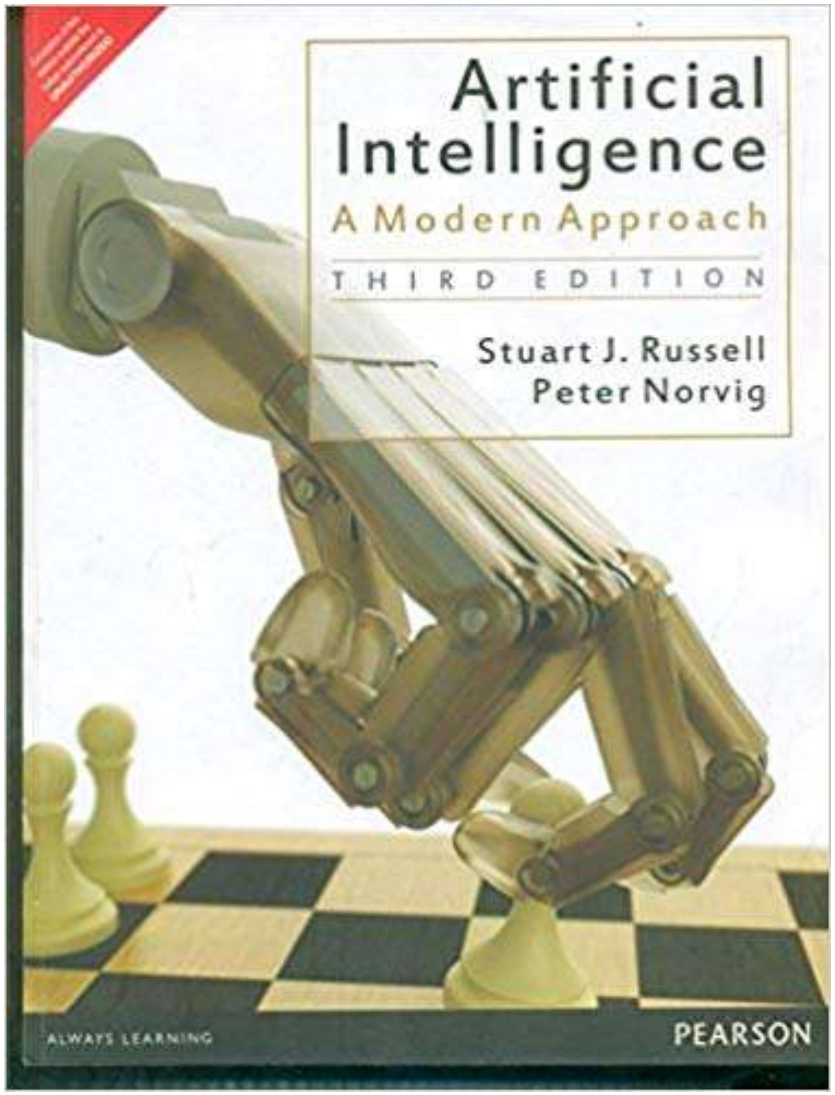
When did it all *really* begin?

Term coined in 1956 by John McCarthy at Dartmouth Summer Research Project on Artificial Intelligence

At that time, Herbert Simon predicted, “machines will be capable, within twenty years, of doing any work a man can do”.

"I could feel – I could smell – a new kind of intelligence across the table." Garry Kasparov, after losing to IBM's Deep Blue in 1997

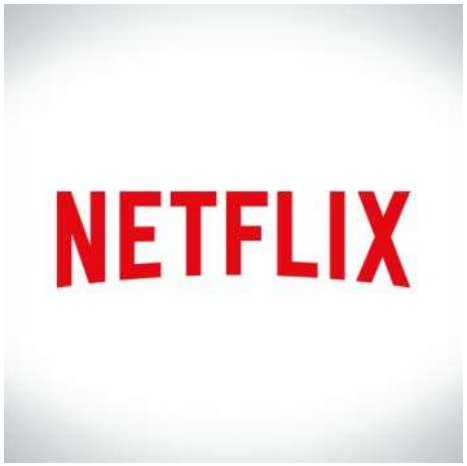
But what exactly *is* it?

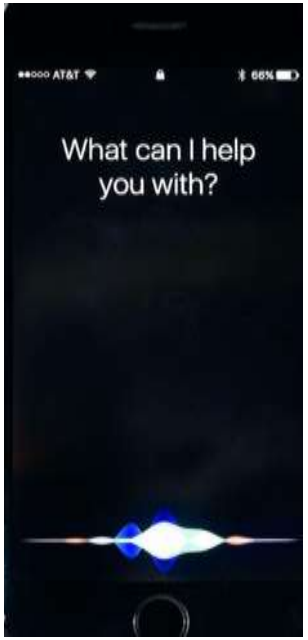


"artificial intelligence" is often used to describe machines or computers that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving"

- Narrow AI systems – task-specific.

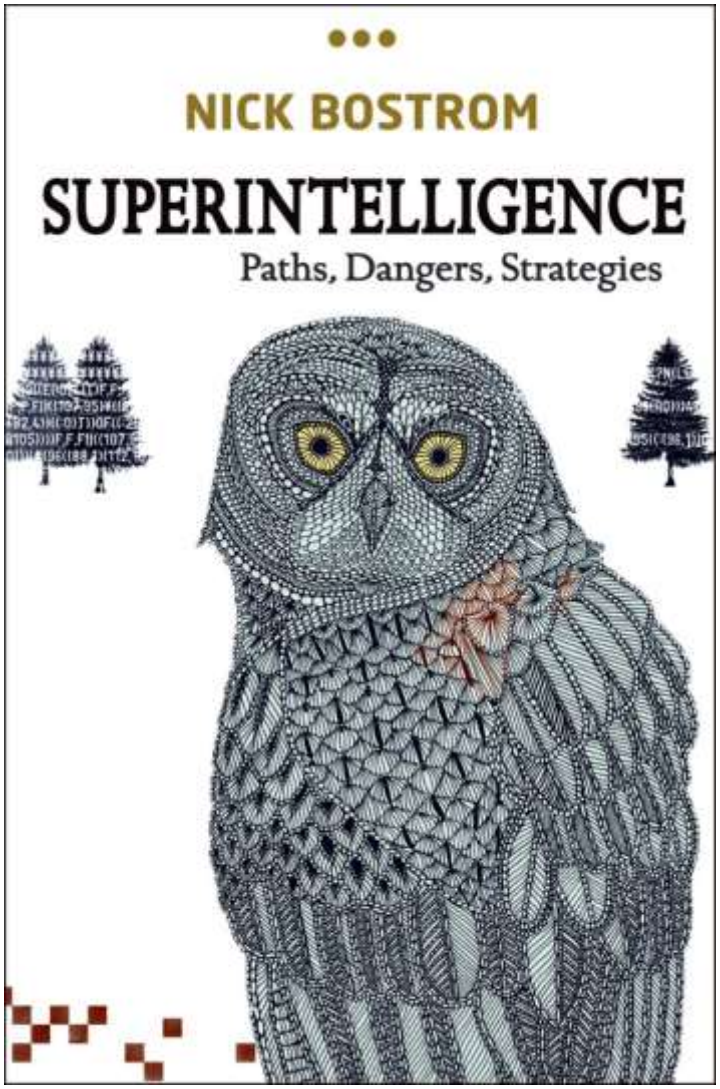
- Artificial general intelligence - capacity to learn any intellectual task that a human being can do.





Superintelligence

"any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest"



Regulation *of* AI

ELON MUSK SAYS HUMANS COULD END UP RULED BY AN IMMORTAL AI DICTATOR

BY **NINA GODLEWSKI** ON 4/9/18 AT 4:45 PM

“AI – what will it mean? Helpful robots washing and caring for an ageing population? Or pink-eyed terminators sent back from the future to cull the human race?”

What could possibly go wrong?

"foolproof and incapable of error"

Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian

Tempe police said car was in autonomous mode at the time of the crash and that the vehicle hit a woman who later died at a hospital

Sam Levin *and* **Julia Carrie Wong** *in San Francisco*

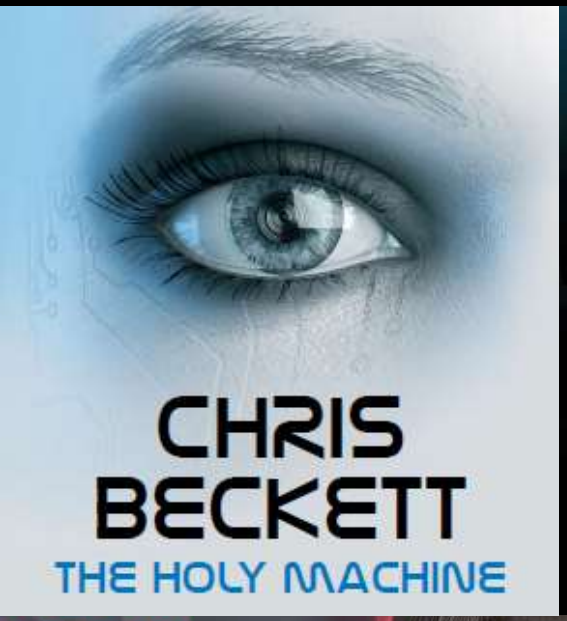
Mon 19 Mar 2018 22.48 GMT

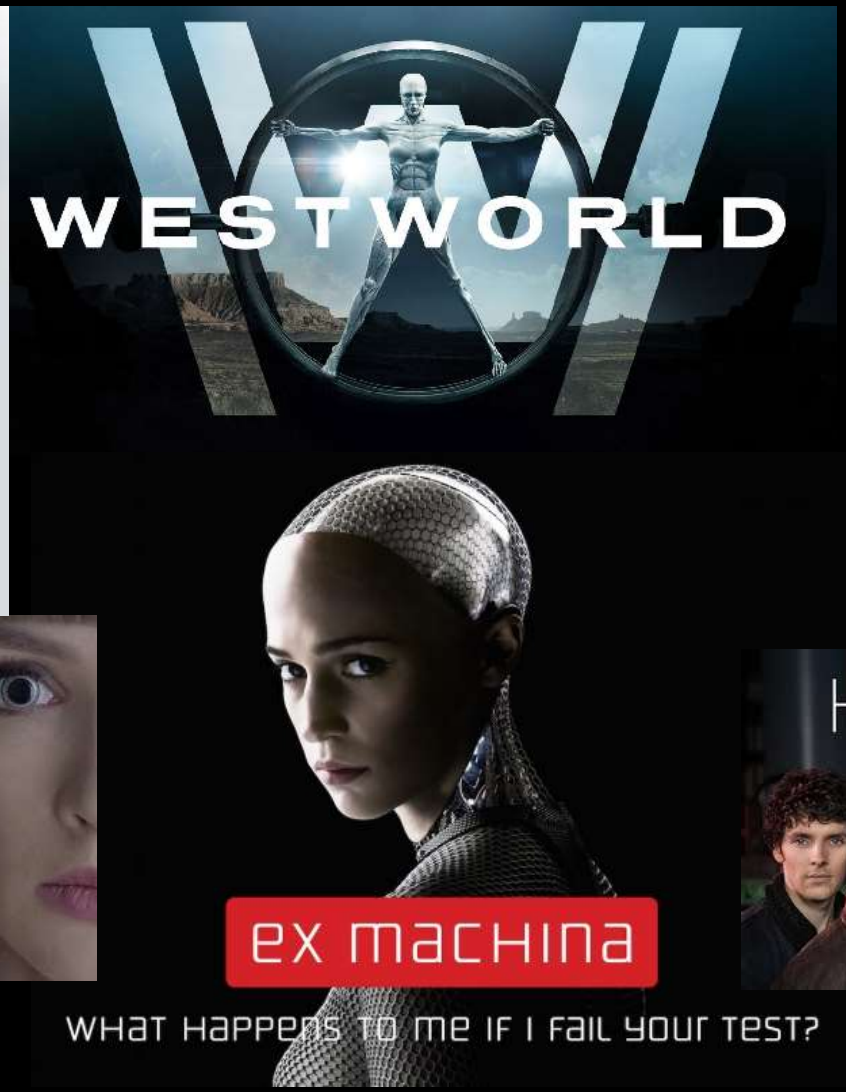
Learning bad habits

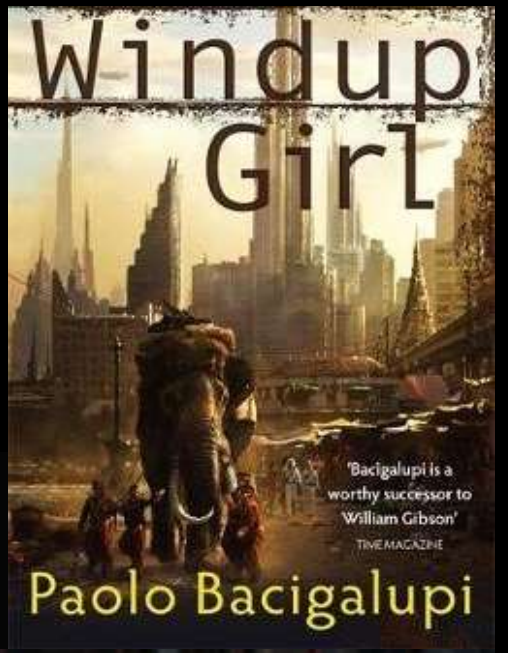
Released on twitter by Microsoft on 23 March 2016

Withdrawn by Microsoft on 24 March 2016

Regulation *for* AI


CHRIS
BECKETT
THE HOLY MACHINE


WESTWORLD
ex machina
WHAT HAPPENS TO ME IF I FAIL YOUR TEST?


Windup
Girl
'Bacigalupi is a worthy successor to William Gibson'
TIME MAGAZINE
Paolo Bacigalupi


NETFLIX
BETTER
THAN
US


HUMANS

Regulation *by* AI

NEW ZEALAND

Privacy and profiling fears over secret ACC software

15 Sep, 2017 5:00am

5 minutes to read

ACC have been using a secret model to predict which claimants will pose a higher risk in terms of their duration of claim.

By: **Kirsty Johnston**

We need to know what makes robo-taxman tick

10 Apr, 2018 5:08pm

3 minutes to read

If machines will rule us, we need to know how they work. Picture / 123RF

The Bulletin: Critics hammer Immigration NZ's racial profiling algorithm

Alex Braae | Staff Writer

Partner content

Police setting up $9m facial recognition system which can identify people from CCTV feed

1:27 pm on 31 August 2020

Share this

Phil Pennington, Reporter
@pjppenn phil.pennington@rnz.co.nz

Police have been quietly setting up a $9 million facial recognition system that can take a live feed from CCTV cameras and identify people from it.

Otago experts to work with Government on AI framework

JULIA GABEL
MAY 07, 2018

Artificial intelligence (AI) experts from the University of Otago have been invited to work with the Government to form an AI and predictive analytics framework.

The invitation comes shortly after the Minister for Government Digital Services and Broadcasting, Communications and Digital Media, Clare Curran, announced the *Artificial Intelligence: Shaping a Future New Zealand* report.

In the announcement, Curran supports the establishment of a framework, stating, "An ethical framework will give people the tools to participate in conversations about AI and its implications in our society and economy."

She adds, "As a first step and because of the importance of ethics and governance issues around AI, I will be formalising the government's relationship with Otago University's NZ Law Foundation Centre for Law and Policy in Emerging Technologies."

Our project

- Phase 1: government use of AI
- Phase 2: impact of AI on work and jobs

Use cases

CASE STUDY

Visa triage

Challenge: Immigration New Zealand (INZ) processes more than 800,000 visa applications a year from offices around the world. As part of the Vision 2015 transformation programme, investments were made by INZ to improve global consistency of their processes, including risk assessment.

Solution: INZ developed a triage system, including software that assigns risk ratings to visa applications. The risk rating provides a guide to the level of verification to be performed by an Immigration Officer on an application, but does not determine whether an application is approved or declined. An Immigration Officer still assesses and decides every application.

The risk rating applied to a visa application is determined by the application of multiple risk rules working together. The risk rules are developed using a range of qualitative and quantitative information and data. For example, one of the "high risk" rules applies if the applicant does not hold an acceptable recognised travel document.

Risk rule changes are overseen by a tiered governance model. A Triage Reference Group assesses risk rule changes, and refers any significant changes to the rules or to the triage model to the Operational Systems Integrity Committee (OSIC). OSIC reports to the Immigration Leadership Team.

Outcome: All temporary entry visa applications are assessed in the triage system and immigration officers follow verification guidelines based on the risk rating to assess applications. This has increased consistency across visa processing offices, improved processing times, and allowed attention to be focused on higher-risk applications. This allows staff to identify new and emerging risks, and see where risks are no longer present.

CASE STUDY

Young people not in employment, education or training

Challenge: The unemployment rate for young people (15–24-year-olds) who leave school, but do not enter employment, education or training is more than double the rate of the next highest age group – those aged 25 to 34. Young Māori and Pasifika are particularly represented in this group.

Solution: Established in 2012, Work and Income's Youth Service, NEET, uses an algorithm to help identify those school leavers who may be at greater risk of long-term unemployment, and proactively offers them support in terms of qualifications and training opportunities.

The algorithm considers factors such as:

- demographic information
- whether a young person's parents were on a benefit
- the school history of a young person (including educational achievement, reason for leaving school, and truancy history)
- whether a young person has ever been the subject of a notification to Oranga Tamariki.

Each of these factors has been shown to affect whether a young person may need support. The algorithm produces risk indicator ratings for school leavers: high (top 10 percent), medium (next 10 percent), low (next 20 percent), or very low (final 60 percent). The rating indicates the level of support they might require and determines the funding for providers.

A young person with more of these factors, or where one or more factors has a higher value (such as multiple truancy or multiple notifications), will have a higher risk indicator rating.

The algorithm refers the high, medium and low risk (40 percent) school leavers to NEET providers who make contact and offer assistance. The data collected by these providers is incorporated into the model to improve future accuracy.

Outcome: Since 2012, more than 60,000 young people have accepted assistance from the service. One-third of these have been offered the service through the algorithm that has automated the referral system.

NEET has proved to be most effective for those with a high-risk rating, resulting in improved education achievements and wellbeing, and less time on a benefit, compared with those who did not use the service.

The ‘up’ side

- More accurate
- Quicker
- Potential to make better decisions than humans

The ‘down’ side

- Fettering of discretion
- Transparency
- May replicate or reinforce bias (while giving impression of being unbiased.)

CONCLUSIONS AND RECOMMENDATIONS

- Human decisionmakers aren't perfect
- Errors aren't always evenly distributed
- Beware of regulatory placebos
- Procurement/development should prioritise transparency
- Value and limitation of individual rights models
- Ongoing checks required

CONCLUSIONS AND RECOMMENDATIONS

- Human decisionmakers aren't perfect
- Errors aren't always evenly distributed
- Beware of regulatory placebos
- Procurement/development should prioritise transparency
- Value and limitation of individual rights models
- Ongoing checks required

- Errors aren't always evenly distributed

- The COMPAS controversy: "Black defendants who did not reoffend… were nearly twice as likely to be misclassified as higher risk compared to their white counterparts (45 percent vs. 23 percent) … white defendants who reoffended… were mistakenly labeled low risk almost twice as often as black reoffenders (48 percent vs. 28 percent)".

- Beware of regulatory placebos

- ‘Human in the loop’
- “It had the hallmarks of automatic rejection based on circumscribed criteria rather than a proper exercise of discretion...” (*Christiansen v D-G of Health* [2020] NZHC 883, per Walker, J.)

- Procurement/development should prioritise transparency

- *Wisconsin v Loomis* (2015)
- "Northpointe, Inc., the developer of COMPAS, considers COMPAS a proprietary instrument and a trade secret. Accordingly, it does not disclose how the risk scores are determined or how the factors are weighed."

- Value and limitation of individual rights models

- Individuals mostly too time-poor, resource-poor, and lacking in the necessary expertise to meaningfully make use of these rights
- Individual rights approach not well suited when algorithms create societal harms, such as discrimination against racial or minority groups.

  – Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for.'

- Ongoing checks required

- Having the model continuously monitored will help to promptly detect whether its performance is worsening or deviating from the expected behaviour (e.g. unintentional discrimination), thus making it possible to take appropriate remediation measures, such as selecting new features or retraining the model. (European Banking Agency)

OVERSIGHT

Government agencies should adopt or develop in-house processes to evaluate proposals to develop or procure new predictive algorithms. These should also apply when it is proposed to apply existing algorithms to a new purpose. These processes should evaluate a range of considerations, including accuracy, transparency, privacy and human rights impacts.

Government should consider the establishment of a regulatory/oversight agency. This would work with individual government agencies who intend either to introduce a new predictive algorithm, or to use an existing predictive algorithm for a new purpose.

JULY 2020

New Zealand Government

Stats NZ
Tatauranga Aotearoa

ALGORITHM CHARTER FOR AOTEAROA NEW ZEALAND

Risk matrix

Likelihood

Risk rating

Low	*Moderate*	*High*
The Algorithm Charter could be applied.	The Algorithm Charter should be applied.	The Algorithm Charter must be applied.

Application and Commitment

The Charter will apply differently to each signatory. The risk matrix approach means that signatories can focus first on decisions that have a high risk and exclude most of the many business rules that government agencies use every day to give effect to legislative requirements and for business as usual activities. The intention is to focus on those uses of algorithms that have a high or critical risk of unintended harms for New Zealanders. This commitment will be reviewed in twelve months as part of the scope review.

Commitment:

Our organisation understands that decisions made using algorithms impact people in New Zealand. We commit to making an assessment of the impact of decisions informed by our algorithms. We further commit to applying the Algorithm Charter commitments as guided by the identified risk rating.

Algorithm Charter Commitments:

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, ETHICS AND HUMAN RIGHTS

- 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.

- Voluntary sign up
- Non-binding
- High level principles
- (Next stage: implementation strategy?)

Canadian Directive on Automated Decision-making

- Came into effect this year
- Binding on government agencies
- Requires Algorithmic Impact Assessment prior to the production of any Automated Decision System.
- Applies to systems in development - not retrospective