AI technologies | Maximising benefits, minimising potential harm

Associate Professor Colin Gavaghan
Professor James Maclaurin
University of Otago

Centre for AI and Public Policy
Centre for Law and Emerging Technologies

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND
SAPERE AUDE

CENTRE FOR
Artificial Intelligence and Public Policy
Te tari Rorohiko Atamai, Kaupapa Here Tūmatanui

The Law Foundation
NEW ZEALAND

AI technologies | Maximising benefits, minimising potential harm

In this talk…

- The relationship between AI and Data Science
- CAIPP as an in interdisciplinary centre
- Mapping the domain of the social, ethical and legal effects of AI
- Cases and strategies for maximising benefit and minimising harm

AI, Data and Data Science

- There are not simple agreed-upon definitions of either data science or AI.
- AI is changing data.

Data was…

- given for a purpose
- static
- able to be corrected or deleted

Data now…

- Data is given but it is also extracted
- Data is inferred
- I know less about what data others hold about me, what it's for, how it was constructed…

I have less control as a data subject

- Tyranny of the minority
- My data is 'exchanged' for essential services by effective monopolies
- It's hard to ask a company to correct or delete data if I don't know it exists or I don't understand what it means
- Data is a form of wealth that is very unevenly distributed

So for the individual

- Data has become much more dynamic, much more empowering, very efficiently harvested
- And I have less knowledge about it and less control over it than people used to

AI is changing business and government

- It is providing insights, new types of products and services.
- It is allowing us to assess intentions, risks… more accurately and on the fly.
- It is allowing us to target resources in ways we couldn’t before.

But…

- The information ecology can be as uncertain for governments and businesses as it is for individuals.
- inaccuracy, bias, lack of transparency are problems for organisations just as for individuals, but organisations have different levels of motivation to solve those problems.

IA is democratising data for both individuals and organisations

- I don’t have to be a statistician to use statistics for very complex tasks
- But at the same time I might not know very much about how or how well those tools are making those decisions.

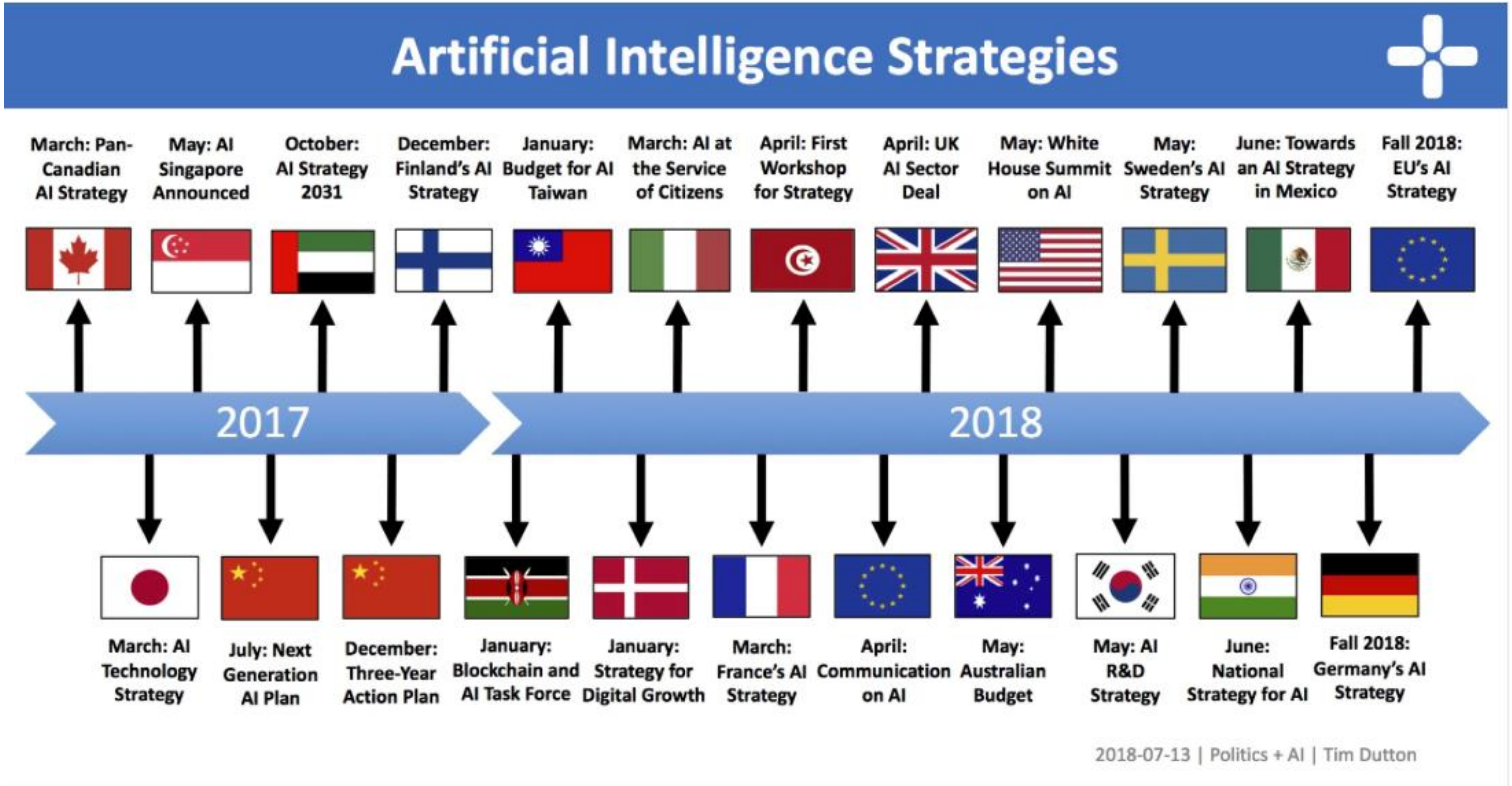
CENTRE FOR

Artificial Intelligence and Public Policy

Te tari Rorohiko Atamai, Kaupapa Here Tūmatanui

Now including computer and information science, law, philosophy, economics, education, zoology, statistics, linguistics, management, marketing, politics, psychology, sociology, social work…

The domain of social, ethical, legal research into AI


Artificial Intelligence Strategies
March: Pan-Canadian AI Strategy
May: AI Singapore Announced
October: AI Strategy 2031
December: Finland's AI Strategy
January: Budget for AI Taiwan
March: AI at the Service of Citizens
April: First Workshop for Strategy
April: UK AI Sector Deal
May: White House Summit on AI
May: Sweden's AI Strategy
June: Towards an AI Strategy in Mexico
Fall 2018: EU's AI Strategy
2017
2018
March: AI Technology Strategy
July: Next Generation AI Plan
December: Three-Year Action Plan
January: Blockchain and AI Task Force
January: Strategy for Digital Growth
March: France's AI Strategy
April: Communication on AI
May: Australian Budget
May: AI R&D Strategy
June: National Strategy for AI
Fall 2018: Germany's AI Strategy
2018-07-13 | Politics + AI | Tim Dutton

The domain of social, ethical, legal research into AI

Collection,
consent,
use of data

Fairness /
accuracy

Data Sovereignty

Effects on employment,
professions

Economic & social
inequality,
polarisation

Effects on productivity,
the economy…

Effects on politics,
democracy,
free speech

Human Rights

Equity of
access

Privacy,
surveillance

Explainability

Autonomy

Regulation,
liability,
institutions

Governance

Bias,
discrimination

Inclusion

Effects on:
Health, Education training,
Justice policing crime,
defence security…

Control,
human factors

Trust

Recreation,
family life,
social interaction

Business,
innovation

liability / responsibility

Effects on Māori

Effects on
wellbeing

Effects on

Effects on politics,
democracy,
free speech

Inclusion

Data Sovereignty

Human Rights

Regulation,
liability,
institutions

Trust

Effects on Māori

Effects on employment,
professions

Effects on productivity,
the economy…

Explainability

Governance

Recreation,
family life,
social interaction

Effects on
wellbeing

Privacy,
surveillance

Business,
innovation

Economic & social
inequality,
polarisation

Equity of
access

Effects on:
Health, Education training,
Justice policing crime,
defence security…

How AI affects individuals

Fairness / accuracy

Data Sovereignty

Economic & social inequality, polarisation

Human Rights

Equity of access

Privacy, surveillance

Autonomy

Bias, discrimination

Regulation, liability, institutions

Inclusion

Trust

Recreation, family life, social interaction

Effects on wellbeing

Data-centric research

Collection,
consent,
use of data

Fairness /
accuracy

Effects on employment,
professions

Data Sovereignty

Effects on productivity,
the economy…

Human Rights

Equity of
access

Privacy,
surveillance

Inclusion

Regulation,
liability,
institutions

Governance

Bias,
discrimination

Trust

Recreation,
family life,
social interaction

Business,
innovation

liability / responsibility

Algorithm-centric research

Fairness / accuracy

Explainability

Privacy, surveillance

Regulation, liability, institutions

Governance

Bias, discrimination

Control, human factors

Trust

Business, innovation

liability / responsibility

The domain of social, ethical, legal research into AI

Collection, consent, use of data

Fairness / accuracy

Effects on employment, professions

Economic & social inequality, polarisation

Data Sovereignty

Effects on productivity, the economy…

Effects on politics, democracy, free speech

Human Rights

Equity of access

Privacy, surveillance

Explainability

Autonomy

Bias, discrimination

Regulation, liability, institutions

Governance

Inclusion

Effects on:
Health, Education training,
Justice policing crime,
defence security…

Control, human factors

Trust

Recreation, family life, social interaction

Business, innovation

Effects on Māori

liability / responsibility

Effects on wellbeing

Artificial Intelligence and Law in New Zealand

Fairness / accuracy

Effects on employment, professions

Economic & social inequality, polarisation

Effects on productivity, the economy…

Explainability

Regulation, liability, institutions

Bias, discrimination

’

Justice policing crime,

Control, human factors

The domain affected by GDPR

Collection,
consent,
use of data

Effects on politics,
democracy,
free speech

Autonomy

Inclusion

Control,
human factors

liability / responsibility

Fairness /
accuracy

Data Sovereignty

Human Rights

Regulation,
liability,
institutions

Trust

Effects on Māori

Effects on employment,
professions

Effects on productivity,
the economy…

Explainability

Governance

Recreation,
family life,
social interaction

Effects on
wellbeing

Privacy,
surveillance

Bias,
discrimination

Business,
innovation

Economic & social
inequality,
polarisation

Equity of
access

Effects on:
Health, Education training,
Justice policing crime,
defence security…

The domain affected by GDPR

**Collection,
consent,
use of data**

Human Rights

Privacy,

Explainability

Regulation,

**Bias,
discrimination**

Trust

AI technologies | Maximising benefits, minimising potential harm

So we know the question we want to answer—

How do we use data in a way that is fair, for public benefit, and trusted.

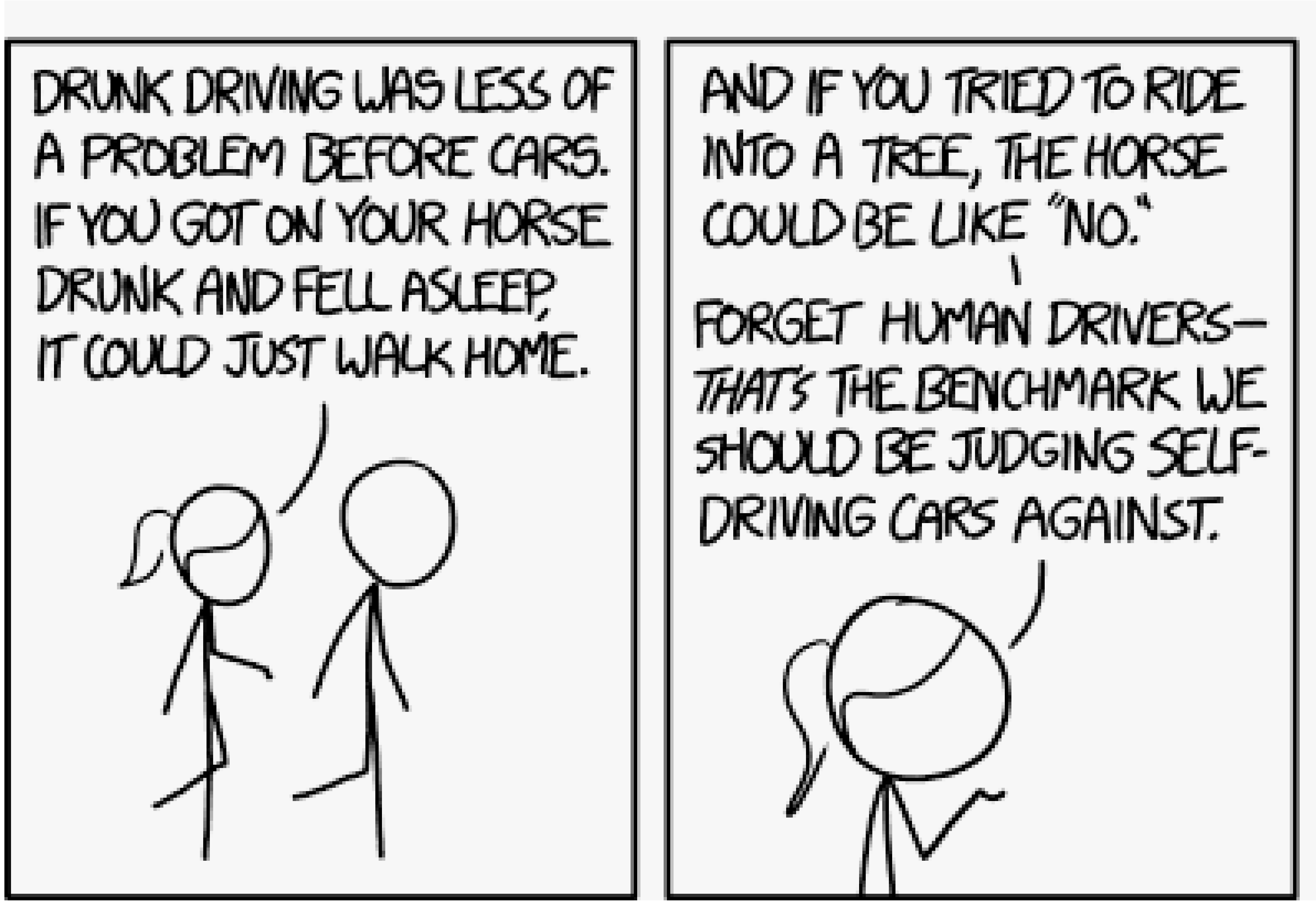
Regulation and AI

Of, by or for AI?

Do we need ‘AI law’?

‘the policy discussion should start by considering whether the existing regulations already adequately address the risk, or whether they need to be adapted to the addition of AI.’ (US National Science and Technology Council)

Not all problems are (entirely) new problems



DRUNK DRIVING WAS LESS OF A PROBLEM BEFORE CARS. IF YOU GOT ON YOUR HORSE DRUNK AND FELL ASLEEP, IT COULD JUST WALK HOME.
AND IF YOU TRIED TO RIDE INTO A TREE, THE HORSE COULD BE LIKE "NO."
FORGET HUMAN DRIVERS— THAT'S THE BENCHMARK WE SHOULD BE JUDGING SELF-DRIVING CARS AGAINST.

Driverless car makers could face jail if AI causes harm

AI technologies which harm workers could lead to their creators being prosecuted, according to the British government.

16:33, UK,
Tuesday 25 September 2018

Responding to a written parliamentary question, government spokesperson Baroness Buscombe confirmed that existing health and safety law "applies to artificial intelligence and machine learning software".

"I'm sceptical both that industry's own tests will be deep and comprehensive enough to catch important issues, and that the regulator is expert enough to meaningfully scrutinise them for rigour," said Michael Veale, researcher in responsible public sector machine learning at University College London.

Right to reasons

Official Information Act 1982

Section 23 (1): where a department or Minister of the Crown makes a decision or recommendation in respect of any person in his or its personal capacity, that person has the right to be given a written statement of… (c) *the reasons for the decision or recommendation.*

Elements of reasons

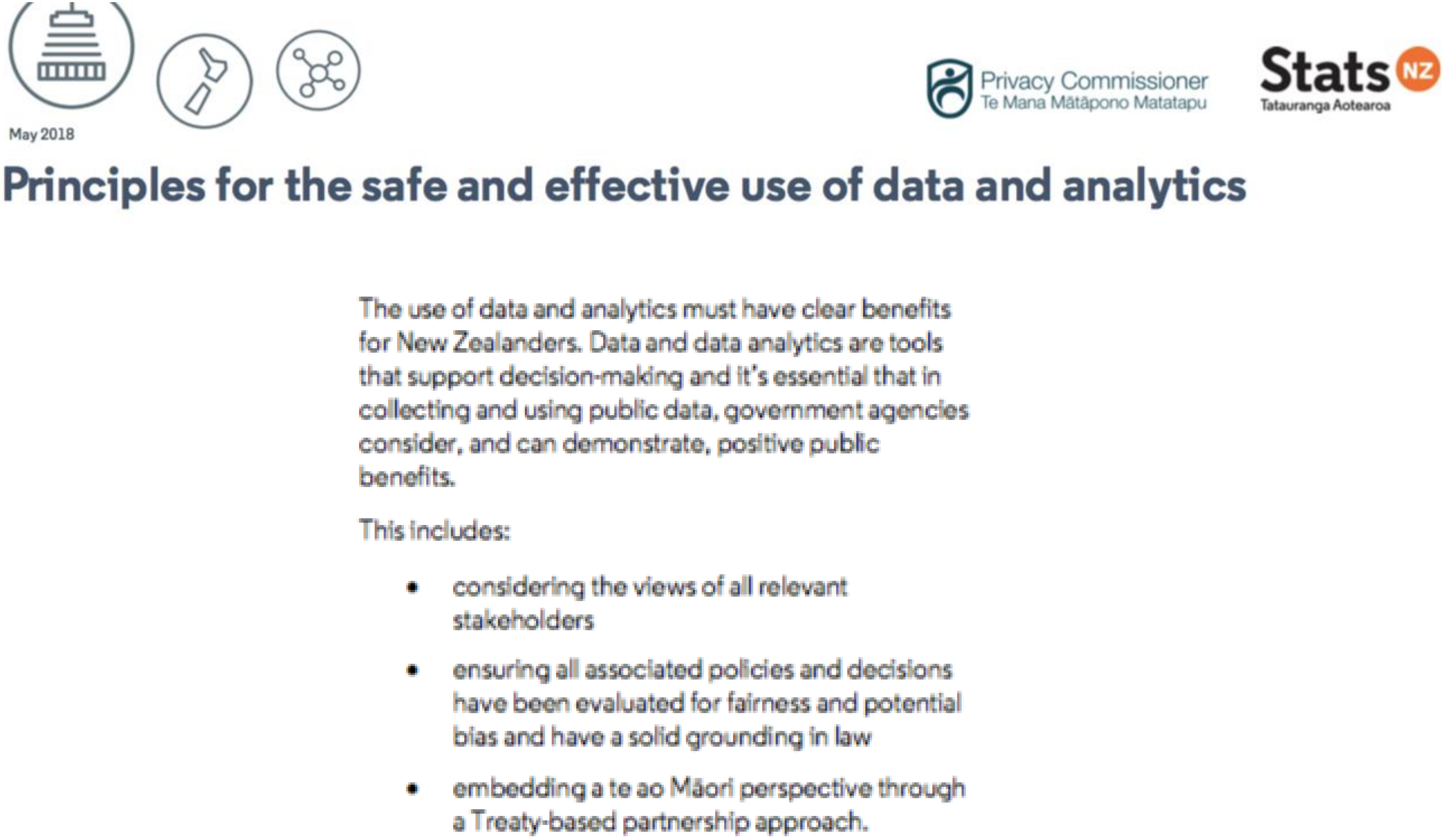
- System functionality – ex ante
- Specific decision – ex post

- Experts in how the software works
- Experts in the sort of decision being made (criminologists, social scientists, etc)
- Non-experts!

Explanation not bafflegab

‘The resulting systems can be explained mathematically, however the inputs for such systems are abstracted from the raw data to an extent where the numbers are practically meaningless to any outside observer.’

Dr Janet Bastiman, evidence to UK Parliament Science and Technology Ctte (2017)



May 2018

Privacy Commissioner
Te Mana Mātāpono Matatapu

Stats NZ
Tatauranga Aotearoa

Principles for the safe and effective use of data and analytics

The use of data and analytics must have clear benefits for New Zealanders. Data and data analytics are tools that support decision-making and it's essential that in collecting and using public data, government agencies consider, and can demonstrate, positive public benefits.

This includes:

- considering the views of all relevant stakeholders
- ensuring all associated policies and decisions have been evaluated for fairness and potential bias and have a solid grounding in law
- embedding a te ao Māori perspective through a Treaty-based partnership approach.

Accuracy and validation

The *Daubert* test (q.v. *Calder* in NZ)

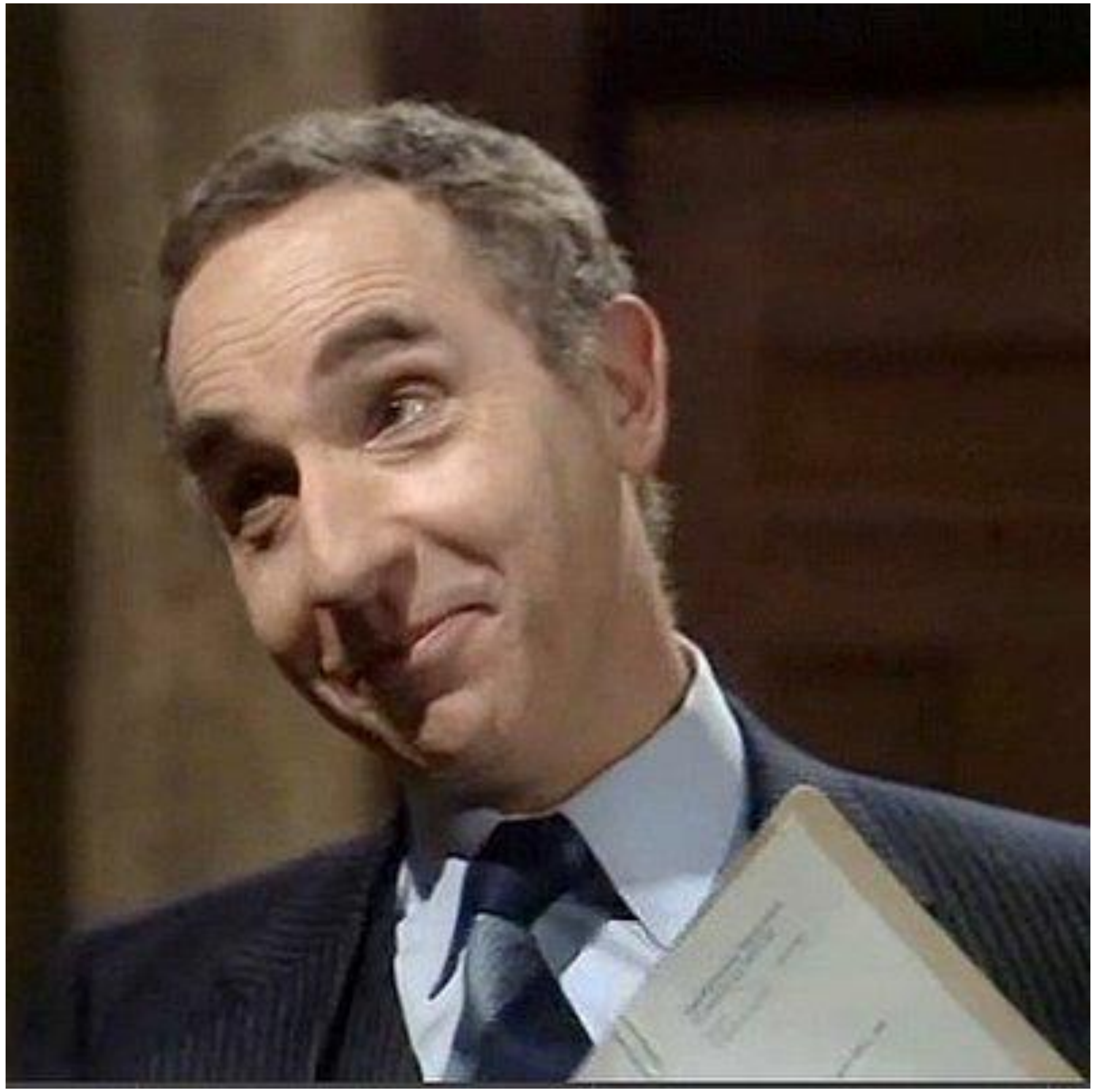
- Relevant and reliable?
- Scientifically valid and applicable to the facts in issue?
- Known and potential error rate?
- Published and peer-reviewed?

Not all errors are equal

ProPublica

- ‘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 quick and easy fixes



- The Politician’s Syllogism
- We must do something
- 'This' is something
- Therefore we must do 'this'

Keeping a human in the mix

‘When it comes to decisions that impact on people’s lives – judicial decisions etc- then a human should be accountable and in control of those.’

Noel Sharkey, Moral Maze, 18 Nov 2017

Belt and braces, or false reassurance?

- Supervisor vs driver reaction time
- Inert but alert?
- Decisional atrophy

“Automation bias” or “algorithmic aversion”

‘It remains to be seen, however, how an algorithm might influence custody officer decision-making practices in future. Might some (consciously or otherwise) prefer to abdicate responsibility for what are risky decisions to the algorithm, resulting in deskilling and ‘judgmental atrophy’? Others might resist the intervention of an artificial tool. Only future research will determine this.’

- Oswald, Grace, Urwin and Barnes. ‘Algorithmic risk assessment policing models’ *Information & Communications Technology Law* (2018)

Real empowerment, or passing the buck?

- Individual data subjects are not empowered to make use of the kind of algorithmic explanations they are likely to be offered
- 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.’

Impossible standards, or settling for too little?