Panel Presentation by the Privacy Commissioner, John Edwards, to the Data Analytics Forum on Ethics and Algorithmic Transparency on 18 July 2018 in Wellington

Panellists: **John Edwards, Privacy Commissioner**
Paul Stone, Open Data Charter
Koren O'Brien, AI Forum

INTRODUCTION

Can big data be principled?

For the purposes of this discussion, I'm less interested in the 'big data' side of the question and more interested in the 'principled' side of the question.

TORTURE THE DATA

We've been thinking a lot about the risks that might be involved in extracting public or private value from data.

Increasingly, CEOs, Ministers are asking "can we automate this? What can we learn from this dataset that will inform policy or business strategy?"

The pressure to look to technology to provide answers to complex social problems is increasing, and is supported by consultancies, data scientists and software vendors.

But as renowned British economist Ronald Coase warned: "If you torture the data long enough, it will confess to anything."

MATH DESTRUCTION

In 2014, Wired magazine warned '*Algorithms are great but they can also ruin lives'.*

The article pointed out that an algorithm may falsely profile an individual as a terrorist.

This is something which in 2014 confronted about 1,500 unlucky airline travellers in the US each week.

As the computer security expert Bruce Schneier noted:

> *Finding terrorism plots is a needle-in-a-haystack problem, and throwing more hay on the pile doesn't make that problem any easier.*

Predictive risk modelling, algorithm-based decision making, machine learning techniques have been called "weapons of math destruction" (US mathematician Cathy O'Neil).

Luke Dormehl (*The Formula: How Algorithms Solve All Our Problems and Create More*) echoed a similar line:

> *A single human showing explicit bias can only ever affect a finite number of people.*
>
> *An algorithm, on the other hand, has the potential to impact the lives of exponentially more.*

BLACK BOXES

Many algorithmic assessment tools operate as 'black boxes' with a lack of transparency or understanding over how they operate.

This can lead to situations where decision-makers make bad decisions, and those subject to the decisions cannot appeal them – because of commercial sensitivity.

Lack of transparency is compounded when private commercial developers claim trade secrecy over their proprietary algorithms.

COMPAS

Compas - an algorithm developed by a US company Northpointe - calculates the likelihood of someone reoffending and suggests what kind of supervision an offender should receive in prison.

The results come from a survey of the offender and information about his or her past conduct.

The assessments are a data-driven complement to the written presentencing reports compiled by law enforcement agencies.

The company that created Compas says the algorithm's results are backed by research but it is secretive about its details.

That secrecy became the focus of a court case in the state of Wisconsin after a young black man, Eric Loomis, received an eight year and six month sentence for fleeing police in a stolen car.

The judge had concluded Mr Loomis was a 'high risk' to the community, based on his Compas score.

In appealing Mr Loomis' sentence, his lawyer had argued that his client should be able to review the algorithm and make arguments about its validity.

The Wisconsin Supreme Court has since ruled against Mr Loomis.

IMMIGRATION NZ

Immigration New Zealand recently confirmed it had scrapped data and predictive risk modelling work it had been doing to prioritise deportations.

An Immigration NZ document obtained by RNZ said “the ambition is to extend the harm model into a predictive model that can be used to predict the likelihood of long-term harm to NZ Inc, based on demographics and individual harm".

The High-Harm pilot model was implemented in July 2014. Its focus on targeting over-stayers led critics to say it was a form of racial profiling.

My office has said we will work with Immigration NZ if it developed technology or a similar initiative in future to ensure it was fit for purpose.

PRINCIPLES FOR USE

Statistics New Zealand and my office reviewed the use of predictive risk modelling by government agencies.

Our goal was to develop guidance for evaluating proposed analytical models.

The idea was to:

- consider the application of the privacy principles of the Privacy Act to the use of data and data analytics;
- assist in avoiding poorly developed analytical models which could compound bad outcomes; and lastly,
- promote best practice in this area.

We developed joint guidance based on six principles to underpin safe and effective data use in the public sector.

These principles are the first step in developing a detailed set of guidelines to support government agencies on best practice for analytical activities.

The *Principles for safe and effective use of data and analytics* guidance is available on our websites. [hard copies here]

I want to focus on two bits of advice in the guidance in particular.

> *Keep in mind the people behind the data and how to protect them against misuse of information.*
>
> *Analytical processes are a tool to inform human decision-making and should never entirely replace human oversight.*

HIDDEN BIAS

Keep in mind the people behind the data and how to protect them against misuse of information.

While algorithms have a reputation for being ‘neutral’, they can inherit biases. They are made by potentially biased people, and can use potentially unreliable or biased data.

For example, data from a criminal justice system often involves elements of historic or systemic racism.

The goal of 'predictive parity' – which is the characteristic of an algorithm where a formula generates equally accurate forecasts for all racial groups – can actually result in 'optimal discrimination'.

This is due to the unequal base rates or bias in the existing data.

A 2011 paper by researchers Faisal Kamiran and Toon Calders, *Data pre-processing techniques for classification without discrimination* sets out the trade-off between accuracy and discrimination in algorithms in the presence of biased historical data.

If you have a dataset that has a historically "favoured" group of people and a "discriminated" group of people, the more you rely on the historical data, the more the outcomes will discriminate against the disadvantaged group.

The researchers were able to demonstrate this algorithmic bias was due to the history of heightened scrutiny of US black neighbourhoods in what was known as broken windows policing, where black people more likely to be arrested for a given crime.

FACIAL RECOGNITION TECHNOLOGY

Analytical processes are a tool to inform human decision-making and should never entirely replace human oversight.

A study on bias in facial recognition software by Joy Buolamwini, a researcher at the MIT Media Lab published in the New York Times in February showed that:

> "Gender was misidentified in less than one percent of lighter-skinned males; in up to seven percent of lighter-skinned females; up to 12 percent of darker-skinned males; up to 35 percent in darker-skinned females.
>
> "Overall, male subjects were more accurately classified than female subjects and lighter subjects were more accurately classified than darker individuals."

Microsoft announced recently its facial recognition technology was now more accurate in identifying people of colour.

Microsoft said its improved system had reduced the error rates for darker-skinned men and women by "up to 20 times".

The company said its facial-recognition improvements were "part of our ongoing work to address the industry-wide and societal issues on bias".

LONDON METROPOLITAN POLICE

The London Metropolitan Police has come under fire recently over its trial use of automated facial recognition technology.

According to information released under Britain's Freedom of Information laws, the technology gave a 98 percent false positive rate.

The Metropolitan Police claimed that this figure is misleading because there is human intervention after the system flags up the match.

But the system hasn't been successful in its true positive identifications either with a watchdog NGO, Big Brother Watch, showing there have been two accurate matches – and neither person was a criminal.

ACHIEVING GREATER ACCURACY

For argument's sake, let's assume a predictive algorithmic system was able to achieve 98 percent accuracy.

While 98 percent accuracy might be a dream result, assumptions based on "close to 100 percent" can have significant effects on the 2 percent.

If you're determining if someone is a potential terrorist, or determining an individual's eligibility for benefits or programmes, 2 percent of 100,000 people is 2,000 people misidentified, mistakenly targeted or ineligible.

As more powerful analytical tools become available, there is also a greater capacity to analyse datasets and to re-purpose information collected for other purposes.

I'm very encouraged that the Ministry of Social Development has prioritised a piece of work called its *Privacy, Human Rights and Ethics Framework*.

Application of this framework will ensure that any possible future operational predictive risk modelling carried out by MSD complies with the Privacy Act 1993 - and balances privacy rights with other objectives.

INTRODUCING THE GDPR

It is with great interest we've seen the introduction of the European Union's General Data Protection Regulation.

The GDPR took effect on May 24th.

In amongst the raft of changes it brings with it, the GDPR also addresses automated decision making and profiling.

The GDPR's definition of 'profiling' is:

> "the automated processing of personal data consisting of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements".

The relevant articles of the GDPR include:

a) Article 13 incorporates transparency obligations for automated decision making, including profiling. A European agency must advise the individual if it uses automated decision-making and explain the logic, at the time an individual's information is collected.
b) Article 21 provides directly affected individuals with a right to object to certain automated decision-making (including profiling). This includes a right to object where the processing is in the public interest by a public body. The processing must stop unless the controller can demonstrate compelling legitimate grounds which override the individual's interests.
c) Article 22 is a more general right for individuals not to be subject to automated decision-making (including profiling), unless authorised by law (provided suitable safeguards apply), or subject to certain exceptions, including the individual's consent. There are safeguards for special categories of sensitive information, that generally cannot be used, and a right to seek human intervention in the decision and to contest it.

These protections are significant for Europe and in the context of international benchmark setting.

This type of data processing is considered to be high risk, requiring agencies to carry out a Data Protection Impact Assessment to identify and assess the risks involved.

In New Zealand, we call them Privacy Impact Assessments.

COUNCIL OF EUROPE

The Council of Europe Data Protection Convention has similar new provisions to the GDPR.

If you are unfamiliar with the Council of Europe, it is distinct from the 28-nation European Union and covers 47 member states as the continent's leading human rights organisation.

Its modernised Data Protection Convention 108 says every individual should have a right:

- not to be subjected to a decision significantly affected them based solely on an automated processing of data with having their views taken into consideration.
- to obtain on request knowledge of the reasoning underlying data processing where the results of such processing are applied to them.
- to object at any time to the processing of personal data concerning them unless the controller demonstrates legitimate grounds for the processing which override the individual's interests or rights and fundamental freedoms.

In the United States, the Electronic Privacy Information Centre – or EPIC - has done a lot of work in advocating to lawmakers and the public for algorithmic transparency and ending secret profiling programmes.

The president of EPIC, Marc Rotenberg, has said "knowledge of the algorithm is a fundamental human right".

PRIVACY ACT REFORM

Against this international backdrop, my office has made a submission on the Privacy Bill currently before Parliament.

There is currently a gap in our legislation relating to the use of algorithmic or automated decision-making and related tools.

I have recommended the Privacy Bill include a new principle to limit the harms arising from automated decision-making and to require “algorithmic transparency” in appropriate cases.

Agencies should not be able to hide behind machines when decisions are taken affecting individual New Zealanders.

END THOUGHT

In the book Weapons of Math Destruction, there’s a suggestion that data scientists, like doctors, should pledge an equivalent of the Hippocratic Oath - one that focuses on the possible misuses and misinterpretations of their data models.

Two financial engineers, Emanual Derman and Paul Wilmott, drew up one such oath in the aftermath of the 2008 global financial crisis. It begins:

> *I will remember that I didn’t make the world, and it doesn’t satisfy my equations.*

And concludes

> *I understand that my work may have enormous effects on society and the economy, many of them beyond my comprehension.*

It’s a reminder to avoid the hubris of assuming that technological solutions can be perfect – when all our instincts tell us the real world isn’t.