Privacy Commissioner
Te Mana Mātāpono Matatapu

PRIVACY COMMISSIONER'S KEYNOTE SPEECH

PRIVACY FORUM 9 MAY 2018

TE PAPA, WELLINGTON

Introduction

One of my favourite apps is Shazam. I'm not alone in that. It is one of the most popular apps in the world, making the top 10 list as far back as 2013, and even now ranking 12 in the top free apps of all time for iOS for iPhone.

If you haven't heard of it, it identifies music for you. If you are in a café, or taxi listening to a radio or a Spotify playlist, Shazam can tell you what is being played.

I love that it works, I love being able to do that. It fills a gap.

But what I love most about it is what it says about human creativity and it's infinite variety and uniqueness.

The fact that within a few notes or bars your combination of pitch, tone, tempo and instrumentation is so unique as to allow the very powerful back-end engine to identify it within seconds as distinct from millions of compositional efforts using the same scales, instruments, and pop music marketing consultants blows my mind.

Can we test it? I'll play you a tune, and you get your device out and ask it "what's that track", or Shazam it or whatever. I'll give you a minute or two, and when you've identified the track, put your hand up.

What did you get?

Upon This Tidal Wave Of Young Blood by Clap Your Hands and Say Yeah.

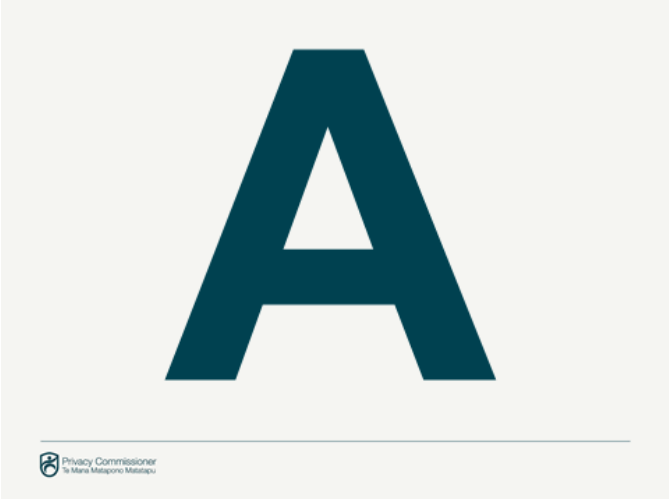
Shazam was bought by Apple Music last year, so I guess Shazam won the war for that space. But it was hotly contested for a while.

I remember back when I first started digitising my CD collection in the mid 2000's iTunes had a feature called "Get track listing".

You'd rip your CD into your machine, and hit that button, and enough of the digital DNA of the tracks would be transmitted over the old dial up line to Cupertino or wherever, and by some miracle, back would come the names of all the songs on the CD, applied to the right track, together with the name of the album, and often the CD cover image. Well, most of the time.

When I hit “Get track listing” for the *Clap Your Hands and Say Yeah* album, something went wrong. I didn’t notice for ages, but it did seem weird to me that this album would share so many names in common with The Raconteurs *Broken Boy Soldiers*.

The song you identified as *Upon This Tidal Wave of Young Blood* has inexplicably been known in my system, across multiple devices, and many iterations of iTunes and iOS, as “A” ever since I ripped it.



My car has some kind of auto-play function that I haven’t figured out how to turn off or alter, so every time I plug my phone in, it plays the first alphabetically listed track in my iTunes “A”.

My family don’t like that track. I can’t think why

I think you know where I’m going. A case of mistaken musical identity, because of an algorithmic glitch, that has proven very difficult to budge, and causes ongoing irritation.

Okay, so this is a privacy conference, what the hell are you doing talking about CDs, and track listings and your auto-play that you’re too dumb to turn off?

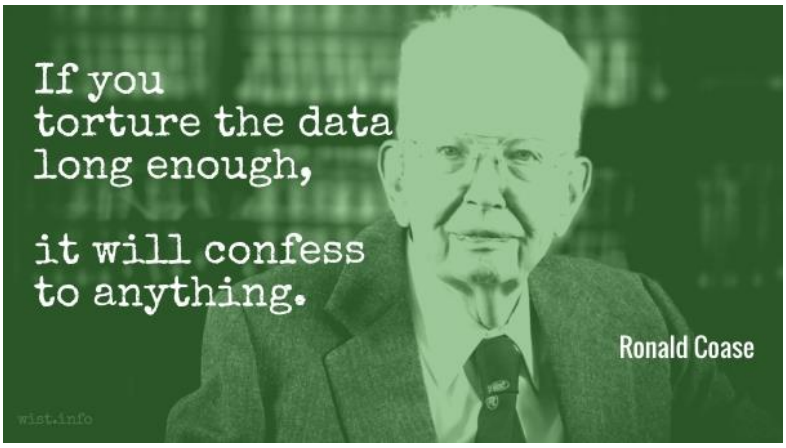
I have used the device of metaphor.

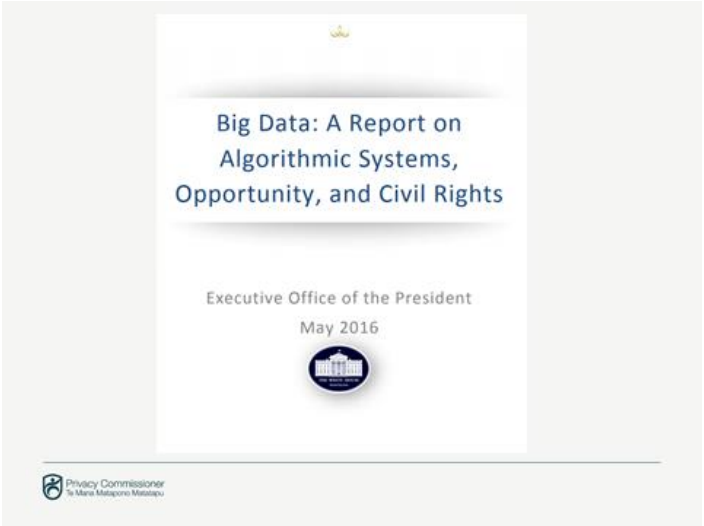
Torture the data

We’ve been thinking a lot about the risks that might be involved in extracting public or private value from data. The use of algorithms to run over your personal information in a dataset, to aid in decision making that might affect resource allocation, or the availability of goods or services to particular individuals in the economy, by commercial or government agencies.

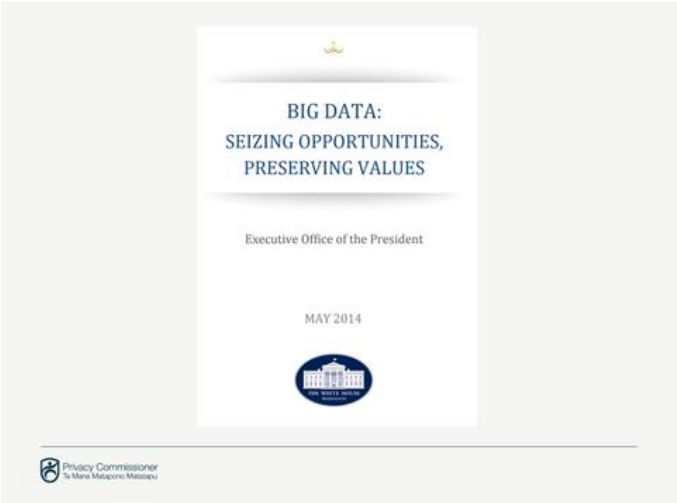
Increasingly managers, CEOs, Ministers are asking “can we automate this? What can we learn from this dataset that will inform policy, or business strategy?” The

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

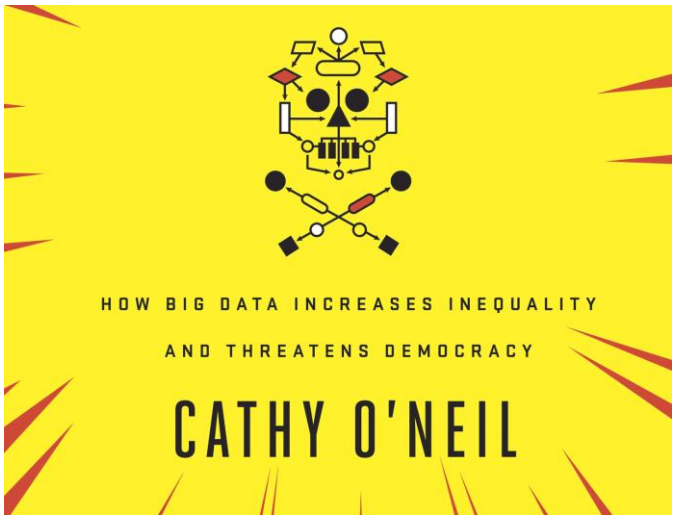




The White House put out a paper on it in May 2016 (yes, the other White House) a follow on from their 2014 smash hit *Big Data: Seizing Opportunities, Preserving Values* work.



That work was elaborated on and popularised in Cathy O'Neil's book, *Weapons of Math Destruction*, and is repeated in a variety of variations in a raft of others.



The 2016 White House paper identified two challenges:

Challenge 1: Inputs to an Algorithm

- Poorly selected data
- Incomplete, incorrect, or outdated data,
- Selection bias,
- Unintentional perpetuation and promotion of historical biases

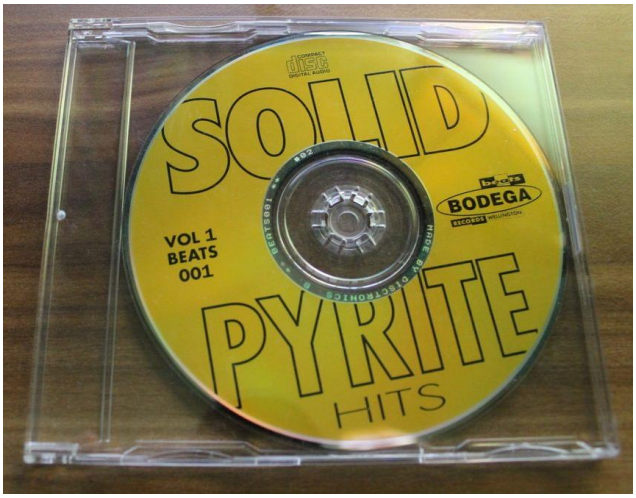
Challenge 2: The Design of Algorithmic Systems and Machine Learning

- Poorly designed matching systems
- Personalisation and recommendation services that narrow instead of expand user options
- Decision-making systems that assume correlation necessarily implies causation
- Data sets that lack information or disproportionately represent certain populations

Error rates

Can we go back to our metaphor?

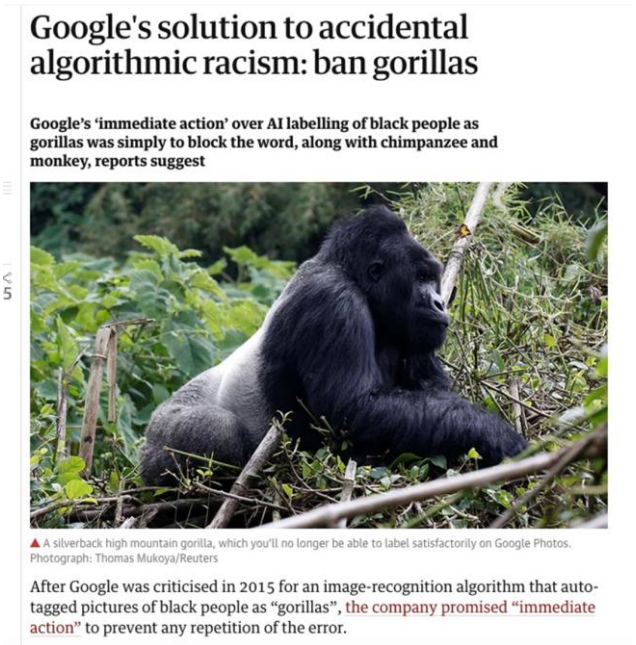
I must have ripped around 200 CDs, there were a couple it didn't know (an album of live sessions from Bar Bodega called *Solid Pyrite*) but only one the machine got completely wrong.



Which means its error rate is only 0.5 percent right? If you went to your boss and said “we can automate our decision making with 99.5 percent accuracy” you’d win employee of the month right, and soon get offered a job with Google or Amazon?

Except, that success rate is not evenly distributed across all music. Early Shazam found classical music more difficult. Its success rate was a fraction of the rate it got with so-called popular music. I’m sure it would still struggle with locally produced indie labels, or bootlegs of live shows.

Similarly, we’ve all seen the examples of Silicon Valley’s data input problem. African Americans tagged as gorillas in image recognition software.



Google's solution to accidental algorithmic racism: ban gorillas

Google's 'immediate action' over AI labelling of black people as gorillas was simply to block the word, along with chimpanzee and monkey, reports suggest

▲ A silverback high mountain gorilla, which you'll no longer be able to label satisfactorily on Google Photos. Photograph: Thomas Mukoya/Reuters

After Google was criticised in 2015 for an image-recognition algorithm that auto-tagged pictures of black people as “gorillas”, the company promised “immediate action” to prevent any repetition of the error.

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; and up to 35 percent in darker-skinner females.
>
> “Overall, male subjects were more accurately classified than female subjects and lighter subjects were more accurately classified than darker individuals,”

If your accuracy is based on measurements of only one input group, you might have to restate your confidence levels to the boss, or funder!

But let’s assume you are going to reach Get Track Listing’s 99.5 percent accuracy rate. If you’ve only got pop music in your system, it doesn’t matter that the algorithm can’t distinguish between one jazz track and another. I might be straining the metaphor, but for example if you are wanting to use an algorithm to identify from your dataset the most needy in the community, for example, those who you haven’t been able to reach with your social programme, building a model around data about people you do happen to have on hand, that is, people who you have had contact with might not help with that objective?

This was one of the problems we foresaw when we reported on our reservations about the now canned plan to tie MSD funding to data in the NGO sector. If you drive away the most vulnerable, their data won’t be in your system. They won’t be reported on, and therefore they’ll be less likely to be targeted for assistance. On the other hand, no data, no problem!

And while 99.5 percent accuracy might be a marketer’s idea of paradise, when you are working with big numbers, assumptions based on “close to 100 percent” so it’s okay to intervene in individuals lives, or determine their eligibility for benefits or programmes, can have significant effects on the .05 percent. Half a percent of the 120,000 registered unemployed in January for example, is 600 people misidentified or ineligible, which might be a lot of disruption and grief for an already vulnerable group, depending on what you are going to do with that data.

A Wired magazine article in 2014 titled '*Algorithms are great but they can also ruin lives'* pointed out that in the US an algorithm may falsely profile an airline traveller as a terrorist 1,500 times each week. Imagine being that guy!

We've started doing some work in this area, focussing on predictive risk modelling to start us off. The first thing we found was that New Zealand government analysts seem to be the only people in the world who call this sub-genre of Big Data by that term.

What we mean by that here is the attempt to determine future outcomes or likelihood of risk by analysing the characteristics associated with those outcomes in historical cases.

The more accurate predictive risk modelling techniques are the more use they have as a way to increase the timeliness of preventative responses – such as in cases of protecting vulnerable children from harm. And you don't have to be as accurate as my old track listing app or Shazam to get the go ahead. We have a national breast screening programme based around an 85 percent predictive success rate.

If you'd told me when I started ripping my CD collection that it was "only" 99.5 percent accurate, I wouldn't have said "oh well, I'll just manually type in the track names then". I would have been happy to accept that my family might start to abuse me every time I went to charge my phone. Your tolerance for the inaccuracy should depend on the consequences, and what steps you can take to mitigate the adverse ones.

Likewise, the possibility (actually the inevitability) that your algorithmic aids have limitations, and might be based on flawed input, excluded data, or sloppy matches doesn't mean you should ditch them. But we need to design in steps to ensure the effect is continually monitored and to check that it is not embedding and amplifying the flaws of your earlier manual system.

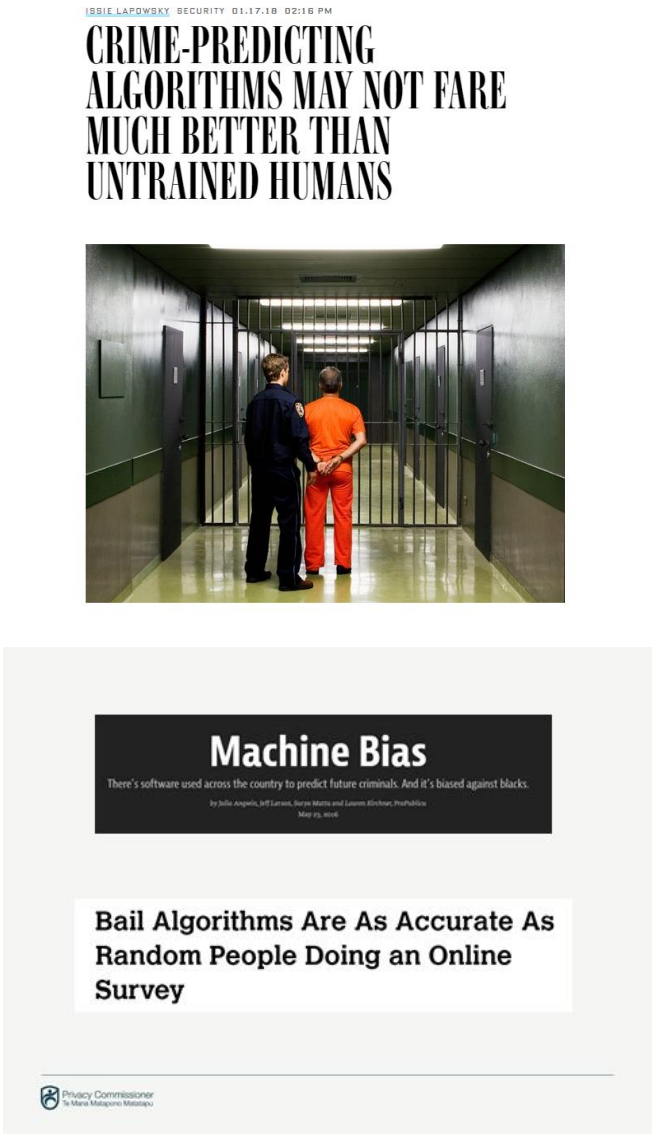
One of the problems we see emerging is the lack of transparency as to how these, increasingly proprietary tools operate. In the US, the tension between an individual's right to know what factors are influencing an outcome that might send them to jail, comes up against the commercial imperative to protect proprietary code, the intellectual property in the algorithm.

Case of Compas

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

A young black man, Eric Loomis, received an eight-and-a-half year sentence for driving a stolen car and fleeing police. The judge had concluded Mr Loomis was a 'high risk' to the community based on his Compas score. But Compas wouldn't

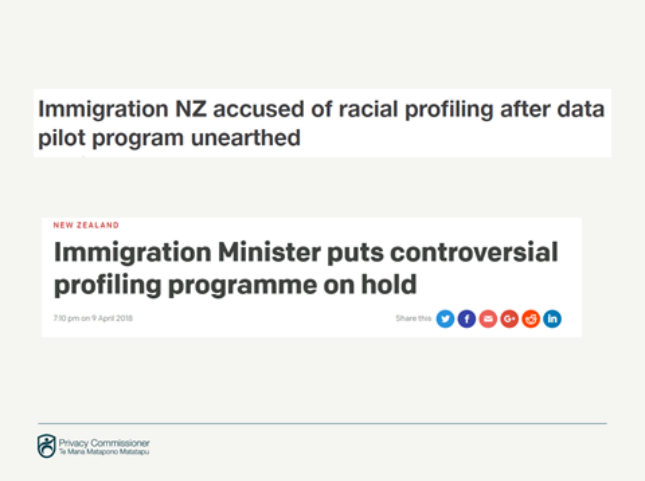
reveal the basis of the score because of commercial secrecy. The Wisconsin Supreme Court backed Compas.

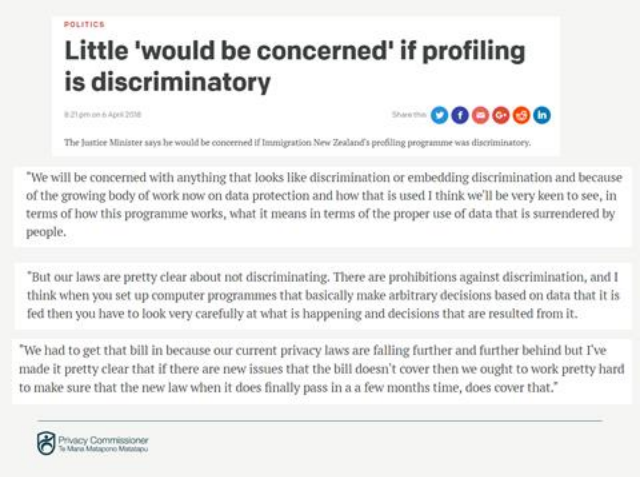


An article which included the Eric Loomis case by the investigative journalism newsroom ProPublica revealed that black defendants were far more likely than white defendants to be incorrectly judged to be at a higher rate of recidivism.

And as far as I know, the state of Wisconsin has yet to complete a statistical validation study of the Compas tool.

What if your dataset uses data based on discriminatory factors, or proxies for those? Could it happen here? Well there is nothing to stop it at present, although when we've looked at two reported instances at ACC and MBIE, we've found that in fact the systems weren't operating as feared or reported.

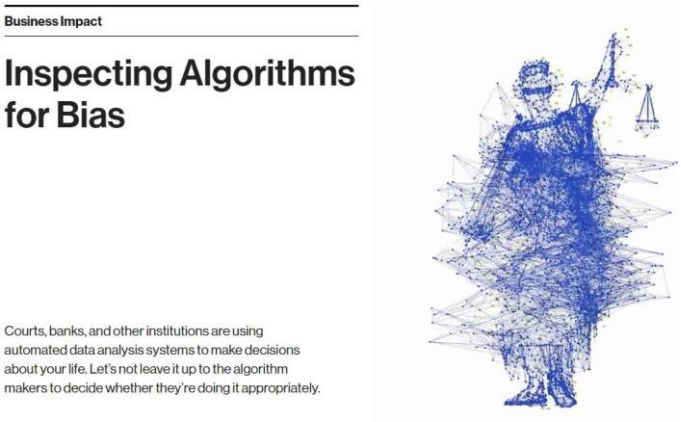




We are very pleased that the government ministers took this seriously, and recognised that this might be something Parliament should look at regulating.

Getting it right

Our next speaker will be able to talk about how the Ministry of Social Development is responding to the current gap in this space. I'm very encouraged that it has prioritised a piece of work called its *Privacy, Human Rights and Ethics Framework* to govern its operational use of predictive modelling and other data use. That is a laudable attempt to fill a lacuna in the law, but is it enough? We wouldn't be the first to regulate to enforce a precautionary approach.



GDPR

Other speakers are going to talk about the European Union's General Data Protection Regulation which takes effect on 24 May. It addresses automated decision making and artificial intelligence, and gives individuals the right to human intervention in cases of adverse decisions.

Article 22 (1): "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her".

Article 22 (3): "The data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision".

These protections are significant in the context of international benchmark setting. They are an influential signpost to future regulatory settings on automated decision making for greater transparency.

I'm going to draw them to Parliament's attention when we make our submission on the Privacy Bill and suggest that they might make a useful addition to the regulatory framework here, now that we have that rare opportunity.

Principles for safe and effective use

In the meantime, we're working with Statistics New Zealand on producing guidance to avoid some of the traps reported elsewhere.

They are still under development, but our working principles say data projects should only proceed where:

1. Deliver clear public benefit to New Zealanders
2. Fit for purpose
3. Focus on people
4. Retain human oversight

5. Transparency is essential
6. Understand the limitations

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. As an antidote to hubris, 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.*

And with that, we might move directly to the next session for people who have thought more deeply about this topic than how algorithmic failure can disrupt the harmony of a family holiday.

Privacy Commissioner John Edwards
9 May 2018