# ACCANect 2018

**Session 11: 1:30 – 1:50pm**

**Australian Human Rights Commission – Human rights and technology**

**Presenter: Edward Santow, Human Rights Commissioner**

**ACCAN is very pleased to welcome Edward Santow, Human Rights Commissioner to talk about The Australian Human Rights Commission’s major project on the relationship between human rights and technology.**

**1:50 – 2:00pm Q&A**

JULIE McCROSSIN: Ladies and gentlemen, we'll be starting in

a couple of minutes. Does anyone not have a lucky number or

they've lost their lucky number? I'm about to do more draws.

Ladies and gentlemen, I'm just a couple of minutes early, but while

the last people come in first of all welcome back from lunch and

thank you for responding to the bell. I'm still committed to us going

to our end of today's networking drinks early. That's part of my

commitment for your early response to my requests. I just wanted

to put in a plug I suppose for our final session that's going to be

looking at cybersafety issues within remote Indigenous communities

and there's a short film that's now owned by the E-Safety

Commissioner and it endeavours to communicate to young people

about cybersafety and I thought just as we get people in and to kick

off our next segment, it only runs for just on 2 minutes. If I could

ask Pete to show this film, it's called "Be Deadly Online". It will

come up in a moment.

VIDEO:

>> Check this.

>> Some of this, all right. (LAUGHTER)

>> Serves you right if you break your bloody neck you bunch of

idiots and don't you put me on the Internet thingy, people might

come and rob me.

>> Don't worry aunty you've got nothing worth stealing.

>> Don't give me cheek, I wiped your bum when you were a baby.

>> Okay, here we go.

>> Internet thingy, Internet thingy. Rob me, rob me. Internet

thingy. Don't put me on the Internet thingy.

>> Ryan, why did you post that video everywhere? Phillip will see

it for sure.

>> No, he loves it, he'll piss himself laughing.

>> This is hilarious borrow, living dangerously making fun of aunty

Em, though. What's with you drinking out of the slushy machine

without a cup. Hardly hi genetic.

>> Sorry, didn't think.

>> And wearing your work shirt when you're up to that dumb stuff.

Now you'll have plenty of time to think. I have to let you go. Keep

the vids coming though, they made me piss myself laughing.

SONG: # Learn all about your security settings #

>> Think twice about who can see your dumb stuff. Once it's

online your dumb business becomes everyone's business. For more

information about protecting your digital footprint, visit the

cybersmart website.

JULIE McCROSSIN: Do you want to give it a round of applause,

guys? I think it's just really interesting to see people's attempts to

communicate with people who aren't going to be reading

information and how do you get the messages across. Thank you

very much for that opportunity to see it. Look, it gives me great

pleasure to welcome the Human Rights Commissioner Edward

Santow who's going to give us insights into human rights and

technology. Please make him welcome. (APPLAUSE)

EDWARD SANTOW: Thank you very much for the intro, Julie. It's

a great pleasure to be here on behalf of the Human Rights

Commission I'd like to acknowledge the fact we're meeting on the

traditional land of the Gadigal people of the Eora Nation. That land

was never ceded and will always be Aboriginal land. ACCAN has

brought us together beautifully and particularly acknowledge Teresa

Corbin who has led this organisation so well for a little while and

hopefully a long while still to come. So I'm here to talk about

human rights and technology and I mean this, the rise of new

technology genuinely brings a wealth of opportunities that we're

excited about, that we should be rightly excited about. Some of the

things that we can see in medicine a genuinely going to make huge

differences to people's lives. They are already helping cure

illnesses. They're also making a big difference to people,

particularly in the area of disability. We have heard some wonderful

examples. For example, Microsoft using image recognition

technology for people who are blind or have a vision impairment.

There are wonderful things out there. But, it's my melancholy duty

at the commission to focus primarily on the risks and so today I'm

going to be focusing especially on the risks, because the promise

that this new age of technological development offers, that promise

cannot be realised unless we address the risks to our basic human

rights. So I'm particularly going to focus today about how our

personal information is increasingly being used in ways that we

might never have predicted and the consequences have quite

profound effects on our basic human rights. To date, we've

considered these issues primarily through the lens of privacy and

that's a really important consideration. We are right to be cautious

and concerned about privacy. But my point today is essentially this,

we need to expand our thinking, because it's not solely our privacy

that is at stake. A range of fundamental rights such as equality of

nondiscrimination, the right to a fair trial, free speech, all of these

rights are also engaged and we need to understand how and we also

need to understand how those rights must be protected. We should

start by considering how our personal information has become

a commodity. The graph behind me shows what I think all of you

already intuitively know and that is that most Australians, in fact the

vast majority, don't read the privacy policies or terms and

conditions before clicking through to access a service or product that

they want or need. 94 per cent according to the Consumer Policy

Research Centre, 94 per cent of Australians don't always read

privacy policies. That seems to me to be actually a surprisingly low

number. I would have thought it would be closer to 99 per cent of

Australians don't always read privacy policies and that 1 per cent

out there, wow, I'd love to have dinner with that person just once.

(LAUGHTER) So the fact that we don't necessarily read the privacy

policies, let alone understand those policies and make decisions on

the basis of that, what does that tell us? I think it tells us a few

things. The first is that often there's only an illusion of choice. We

may appear to be consenting to the use of our personal information

in a range of areas, but if we truly need to access a particular

service or product or application, then we don't really have a choice

as to whether to accept that. Secondly, it also I think shows that

we don't necessarily understand in advance how our personal

information might be used and I want to come back in particular to

that second point in a moment. One of the wonderful things about

ACCAN is it exists - and I'm quoting here from its mission

statement - "to empower consumers to make good choices about

products and services". I want to pause in particular on that term

"good choices". Early in the afternoon as you've just had a lovely

lunch, I'm not going to get too philosophical or abstract, but I do

need to make this point. We're at a juncture where we need to

question what we mean by good choices. So put it bluntly, if you're

faced with Hobson's choice, there is no good choice to make. As

new technology reshapes our world, we need to reconsider the role

of consumer choice. It is fundamentally important. It is crucial that

individuals in a democracy be able to exercise autonomy, that they

be able to exercise choice, but it has to be real choice. It has to be

choice where the scales aren't kind of tipped in a particular way. It

has to be informed, but it also has to be free and that too often is

something that is at threat. It's been said that our personal

information is the fuel that powers artificial intelligence. I, like

many people, am excited about the rise of artificial intelligence or AI

because it offers the prospect of being able to achieve things that

we humans simply can't, or don't do well. I, for example, am not

objectively speaking the best or most accurate driver. The prospect

that my children may never need to learn to drive because they rely

on autonomous cars and that we as a result are able to reduce the

carnage on our roads. I think that's something to be genuinely

excited about. In other words, there's a lot to I guess embrace

when it comes to AI. But we need to have appropriate guard rails in

place. And so at the commission, so far in our project - and we're

still early days - we're particularly worried in the AI area about three

related phenomena. The first is what I would call, again fairly

bluntly, as our personal information being used against us. And

that's quite different from a kind of narrow conception of privacy.

It's not that our private self is being made public. It is that quite

literally, personal information about us is being used against our

interests. The second thing that we are concerned about is

problems with how algorithms wrangle personal information which is

a way of saying it is beginning to rise a new species of

discrimination. It's known as algorithmic bias. It's where

AI-powered algorithms end up having usually completely unintended

consequences that are discriminatory on the basis of things that we

can't control like your age or race, your disability, your gender and

so on. And then the third concern that we have is the fact that AI is

still very much in its infancy and so when deployed in the real world

there are still really significant problems. Again, that's probably like

a $3 way of saying something much simpler and that is that the

technology isn't very good yet and where it's not very good, where

it's unreliable we should be deeply, deeply worried. So starting with

the first of those phenomena, as a community we're only just

starting to understand that our personal information is being

used - sorry, is being widely collected, that that personal data is

being aggregated into big pools of information and that my personal

information is being further aggregated with everybody else's

personal information and that finally, algorithm s trawl through that

big data set with a view to getting insights about us and making

predictions about the future. Many predictions are benign and

innocuous. I personally think it's a good thing if my doctor fully

understands my personal medical history and can compare it with

other people's medical history so that she can provide me with

better medical treatment. I think that is a good thing. I would

embrace that, but I also think it's a choice, so that's my personal

choice. But other insights are much more problematic. We need

here to differentiate between what is mildly irritating and what is

deeply harmful. In the irritating category I'd put things like

advertising that is targeted at me based on an analysis of my

viewing habits on the Internet. That is annoying, but it's not going

to affect my absolute basic human rights. Much more harmful

would be something like a decision to deny me a loan, or to

conclude that I'm likely to commit a crime and before having done

anything wrong I then get detained. Now if those are the sorts of

conclusions that are increasingly being drawn from our personal

information, we're truly entering a world where our data can be and

is being turned against us. So I mentioned before the phenomenon

of algorithmic bias. To understand how that works, we first need to

understand how AI-powered applications work and where they fall

short. So a good example is image recognition or specifically facial

recognition. So AI is powering new technology that is used to

classify or identify people. So you can have a computer that sees

an image, a picture, a photo and says "That person is John Smith"

or "That person is a woman" or "That person appears to be of

a Caucasian background" or whatever and that has almost limitless

applications obviously in criminal justice, but also business and

elsewhere. Alan Broad has published an excellent book that deals

with these issues and it's called "Made by Humans the AI Condition"

and drew attention to the research project known as Gender Shades

and that research aimed to test the accuracy of three popular

commercially-available classifiers of photos of people to determine

what their gender is. One was from IBM, one was from Microsoft

and the third one was Face ++. Each of those classifiers is powered

by artificial intelligence. Let me explain to you how it works. The

researchers wanted to see how these classifiers performed in

respect of different skin colours. In other words they wanted to see

how accurate the classifier was depending on the colour of

a person's skin and so they started with what's known as "the pilot

parliaments benchmark". That's the picture that you might be able

to see behind me. It is a way of providing a fairly representative,

well, a very representative sample of the ethnic diversity of our

community so you can see the diversity in shades of skin tone over

the entire community. So they started with that to get a baseline

set of data. Ellen Broad then summarised the research findings.

She said: "Well only 1 per cent of lighter skin males identified by

commercial during the testing, gender was misdiagnosed in 7 per

cent of lighter females." In other words the lighter a person's skin,

the worse the algorithm performed. Why is that? Usually the data

set that the AI is trained on is not representative of the broader

community. It doesn't have the kind of baseline data that I have on

the slide behind me with a full representation of the diversity of our

community. It is much, much more on the whole it is much, much

more I guess representative of Caucasian men than it is of anyone

outside of that category. As a Caucasian man I am conscious that

orange is not the only fruit, that I'm not the only type of person in

the community. But let me drill down on some of this data a little

bit more. So the next slide behind me has actually two graphs. The

stop graph shows the accuracy of these gender identifiers in

aggregate. So it was somewhere between 88 and just under 94 per

cent accurate in aggregate when you looked at the entire spectrum

of the community. What then is quite shocking is where you

differentiate out just by skin colour, not by gender just by skin

colour. So there you can see that if you are Caucasian or white or

certainly have lighter skin, then the accuracy of these AI-powered

programmes is very high. So it's somewhere between 95 and over

99 per cent accurate in respect of people who look like me, white

men. On the other hand, in respect of darker skinned people, it is

much, much less accurate. It's somewhere between 77 and 87 per

cent accurate. It's a huge difference. It gets worse and worse in

respect of other groupings of people. So let me give you just two

more examples. If you look at misgendering simply by reference to

skin colour, 93.6 per cent of the errors on the Microsoft application

are related to people with darker skin and then the final example

that I have behind me is if you look at it by reference to

women - people who are female - 95.9 per cent of the errors

occurred with respect to females on the Face ++ tool and so the

point I'm trying to make here is that these tools are not anywhere

near perfectly accurate, but the inaccuracy is not spread evenly

across the community. Those people who are more likely to have

an inaccuracy in respect to them are also more likely to experience

other problems because of historical injustice or discrimination or

whatever it happens to be. So let me try and apply this to

a practical situation. And the particular example I want to give is

the use of facial recognition in policing, which is increasingly

something that is happening all over the world and there have been

some pilot studies that have been run particularly in the UK. The

first example that I'm going to give in a moment is from the London

Metropolitan Police and what they tried to do was see whether the

facial recognition technology was able to identify people on their

alert list and see whether it was able to come up with who those

people were. They put 104 alerts through the facial recognition

system, or at least 104 of the ones they put through the facial

recognition system, the computer basically said "Yes, we can

identify who it is". The graph behind me shows how many of them

were wrong. It's 102 out of 104 were wrong. Over 98 per cent of

them were wrong, which on one view is useless or worse than

useless. The London police had an interesting argument and it's

not - it sounds very quirky and it is a bit quirky, but it's not

completely without merit - I'm not going to get too diverted - they

said "Well, look it doesn't matter if it keeps on identifying wrong

people, as long as it gets one or two correct alerts you can rely on

the humans to make sure they're going to pick up all the incorrect

ones and the two that were correctly identified, we might never

have caught those people anyway". I'm sceptical about that,

because that's not generally how humans interact with computers.

If the computer says "This person is your criminal" a police officer

generally quite understandably will go "Yeah, there's a kind of

confirmation bias that can come into play here and that can be

dangerous particularly when you consider the coercive powers in

respect of every police force around the world. There was a slightly

bigger study that was run by the South Wales police in the UK.

I hasten to note that I haven't said "NSW" this is the original South

Wales over in the UK. That was a bigger study. It involved 2,400

false positives. Only 234 of the alerts that they were able to run

actually had correct matches. So less than 10 per cent accuracy

there. Again, I would say that that means that we should not be

using this technology yet unless we can increase the accuracy

many, many, many fold, we shouldn't even be considering using this

technology particularly when we are aware of what the

consequences are of getting these things wrong. Which brings me

to my second last slide. Why should we be worried about this? I've

given a specific example of what's known as algorithmic bias that

applies in the criminal justice context. The slide behind me tries to

summarise some of the areas in which bias can arise and just for

the benefit of people who can't see the slide, it shows that this is

a very wide application. So anywhere where AI might be used to

make these sorts of assessments and particularly to make

predictions, there is a real risk of bias. So we've already seen, for

example, in the banking sector that predictions based on who might

be more likely to repay a loan, particularly overseas, the studies

have shown that they can particularly negatively affect people of

colour and even in situations where the algorithm are told "don't

consider race or ethnicity" there are proxies that tend towards

identifying someone on the basis of race or ethnicity. Particularly in

places like the US there are certain geographical areas where you

may have a high congregation of people of a particular ethnicity. So

if you take into account geography, then effectively you are also

sometimes taking into account ethnicity. That's just an indirect

consequence of doing that. And so the sorts of consequences that

arise from algorithmic bias are really pervasive and so we need to

be conscious of that and address that problem. So, my last slide.

I have been setting out what I realise is mostly something of a tale

of gloom... of "doom" I should say, but it's not all doom. The

reason we have established the project on human rights and

technology at the commission is because we genuinely believe that

AI offers and some of the other technologies, offer enormous

potential good. The term that we tend to use is "responsible

innovation". We don't want to stand in the way of innovation, but

what we do want to do is make sure that innovation happens in

a way that is protective of people's basic human rights which is I

don't think a huge ask. To give an analogy it's a bit like if we were

having this conversation a century ago saying, "Well, we've invented

this new thing called a car. It's very exciting, particularly for horses

and we're just going to roll it out, we're not going to worry about

people getting hurt in new ways, we're just going to roll it out and

everyone will all have a car and everything will be fine". That's not

the approach that we took then and it's not the approach we should

take now. Yes, there are huge opportunities here, but we need to

guard against the risks and so we in July, released an issues paper

that essentially set some of the background information that I've

been talking to today out, so you don't have to be an expert in

technology or human rights in order to engage with some of these

issues and it asks ten key questions that we hope to get information

from ordinary members of the public and people with particular

expertise as well that will help us understand some of the problems

and hopefully also understand some of the solutions, as well, be

able to identify some of the solutions. So a little plug, submissions

close on 2 October. If you want to see me in a particular abject

form you can ask me to beg for a submission and I will beg. We're

already receiving lots of submissions, but you're the sorts of people,

the people at this conference are particularly the sorts of people we

want to hear from. We want to hear from the diversity of the

community and it's really, really important to us and a submission

can be anything from a formal written document with lots of lovely

fonts and everything to some writing on the back of a napkin, I give

that specific example, because I've had a submission before written

on the back of a napkin. It was actually quite insightful. So

anything we'll take and that will be very, very valuable information

for us. The specific project web site we've established as you can

see on the screen is tech.humanrights.gov.au. It gives you

information about everyone in the project. Thank you very much.

JULIE McCROSSIN: Thanks heaps, Ed. A question or comment

guys. Let's have a couple. Who's already made a submission just

out of interest? No wonder he's here talking to you! There's

brainiacs in this room Ed, they've been talking all morning.

EDWARD SANTOW: You've got plenty of time. As a student you

never wanted to do something too early before the deadline.

JULIE McCROSSIN: Any question or comment? I'll nick through

here, if I may.

>> Paul from the ACMA, what is Hobson's choice?

EDWARD SANTOW: Hobson's choice is where you basically don't

have a real choice. All of your options are bad, essentially. Or did

you mean Hobson's choice specifically in this context?

>> I'm interested in the background.

JULIE McCROSSIN: Ed, we've had a running - could you put Lethe

up on the screen for me - we've had an ancient world theme. If

I could draw your attention, this is Lethe, we've heard someone

talking about the right to be forgotten or deleted. You can imagine

in this context what that meant and there's young people here. I'm

frightened they don't have a classical education and I think you

went to a school where that was offered. This is the Godess of

Oblivion, forgetfulness personified. When you went into the

underworld there was a river and if you drank of its waters you

forgot your life and that was evidently seen as a positive. I think

we've got a second image. There she is, a beautiful statue. What

this gentleman was doing to throwing me a lead to do work.

EDWARD SANTOW: I think it has an Australian background.

JULIE McCROSSIN: I think Robin Williams on ABC Radio national

had a program about it.

>> Thanks, I enjoyed your presentation. Ellie from RMIT

university. I'm wondering in terms of the submission and the work

you're doing to what extent does this end up "we need better

technology and more accurate data". It can kind of end up there.

Police stuff, in particular. They just haven't got there yet. Are you

kind of getting that response?

EDWARD SANTOW: A little bit, but not too much. I'll give you

another example from the policing context and people may be

aware of this example, because it's probably the most infamous.

It's known as the "compass example". So that involved an

algorithm in the US used to determine whether or not someone

should get bail and length of sentence and it turned out that

algorithm gave twice as long sentences often to African-Americans

than to no one African-American and twice as likely to deny bail with

basically comparable situations. The question is why. Again, that

particular algorithm was instructed "Don't consider race" but instead

it relied on a huge data set in those US State criminal justice

systems. That data set itself is problematic, because it relies on

convictions, many of which were of African-American people with

all-white juries where we know people of colour in the US have

traditionally been overpoliced as compared with

non-African-American people. We know all of those problems.

Being able to know those problems and being able to solve them are

two different things. It's not simply a question of going "Okay, we

can't rely on some of that older historical data", because some of

those problems continue until today. It's more complex than that.

It's being able to identify those biases and being able to correct to

them. If I could add one thing on a personal note, Julie I think

rightly alluded to my very fortunate upbringing. I did go to a school

where they taught Latin. I didn't partake in it myself.

JULIE McCROSSIN: I was horrified. I'm trying to keep Latin alive!

EDWARD SANTOW: My dad was a judge. We fortunately for my

family, we didn't talk a lot about law or judging or whatever, but

I did have one conversation with him not long before he died where

I said to him "What do you reckon makes a good judge? " His

answer really surprised me. A good judge is someone who can

suppress their prejudices. What's bound up in that idea is that we

all - all of us - are motivated at least in part by some negative

brings, it's not always good. We all have prejudice built into us.

The key is to look for those prejudices and then say "I can't take

those things into account in my decision-making". It's true of

humans, but also of machines and I guess the new tool we need to

develop is how do we identify those prejudices and how do you

correct for them? Identifying them is actually easier than correcting

for them.

JULIE McCROSSIN: I'm so sorry, there's one more question and

we've got to be very quick. I've just been given information about

Hobson's choice, but I'll come back to that in a second and just race

up here.

>> G'day, Tim Holbin. In 2000 I had - many years ago I had this

idea of people having an information bank and some now 18 years

on I've been working with others where the apparatus to build that

means for people to own their data is technically present, in fact it

powers Facebook. We for some reason still have thermally printed

receipts. You talk about the very difficult job of your father as being

a judge. We need evidence to be able to present something to

a court and as we change from having paper records to databases

that are administered by agents within companies with increasingly

less information that can be trusted by a court to be provided on

behalf of citizens, how is it that given the technology exists to be

able to change this, the Human Rights Commission is thinking about

the way in which personhood extends into our info-sphere.

EDWARD SANTOW: That's not an easy question, but it's an

important one. Let me just try and say two things really quickly.

The first is this notion of owning your own personal information

I think is a really important one and I'm pleased to see change and

reform in that area. I think that's fundamentally a good thing, but

I also think it's not going to solve the problem in its entirety. That

doesn't mean it's a useless thing, it's just not the full problem. And

let me give an example or maybe an analogy. About 60 years ago,

you had companies that basically would write contracts with people

that said "Look, even if this new car that you've just bought is

absolutely just shocking and was negligently designed and built and

it blows up and it blows you up with it you're never going to be able

to sue us because you've signed away your rights". What we

decided in countries like Australia in pretty much all liberal

democracies is you can't contract out of certainly obligations.

Similarly I would say here that we need to be clear, it's not just

about putting all the responsibility on individuals to say "Okay,

you've got to read reams of privacy policies and you need to take

care of your own personal information". That's part of it and we

should be assisted in doing that, but it's also about saying, "Well

there are certainly things we just don't think should be allowed to

happen" particularly when we're talking about using a person's

personal information against their interests. That's something

where we need to put I think some hard-edged limits and I'm not

talking about really wild ones, I'm really just talking about limits

that protect people's basic human rights. I'm not making some

radical claim here. It's really basic stuff here, much as we've done

in consumer law more generally, so maybe before I get ahead,

I might leave my answer at that.

JULIE McCROSSIN: Thank you very much. I've got some initial

information on Hobson and I'll come back with more. Thomas

Hobson was a stable owner in Cambridge England died in 1861. He

offered customers the choice of taking the horse in his stable that

was closest to the door, or none at all. Interesting, I wonder how

he stayed in business, but that is the origin of Hobson's choice.

Please give Edward Santow a round of applause.