Australian Government

Chief Scientist

DR ALAN FINKEL AO

Committee for Economic Development of Australia (CEDA)

Speech to a Public Event Lunch

Artificial Intelligence: A Matter of Trust

May 18th May 2018

Four Seasons Hotel
SYDNEY

Here's a question: do you consider yourself to be a trusting person?

Or let me put it another way: would you put your life in the hands of a total stranger?

The latter question probably makes you feel uneasy, but the reality is that you *do* put your life, and the lives of the people you love, in the hands of total strangers, hundreds if not thousands of times every day.

Take me, for example.

This morning I woke up. I switched on the light – trusting that I wouldn't be electrocuted by a faulty lamp, or cord, or socket.

I ate breakfast in a café – trusting that I wouldn't be poisoned by salmonella in my toasted sandwich.

I got into a car – trusting that every one of thousands of components was sound, that the mechanic had serviced it properly, and that other drivers wouldn't kill me.

I walked from my hotel across Elizabeth Street in peak hour. Hundreds of cars bearing down on me.

Sydney drivers.

And nothing to protect me except a red light and a white line.

Scary.

But if I were to rationalise these decisions, they would all make sense.

First, because I've done these things before, and didn't suffer.

But more importantly, because I know that I live in a society where human behaviour is governed by conventions and rules.

So I don't need to personally know the people in the cars on Elizabeth Street to trust that they will stop.

They will stop because the law says they have to.

They will stop because our society says it's good manners to wait your turn.

They will stop because breaking those rules would have social consequences, and employment consequences, and economic consequences, far more inconvenient than the temporary inconvenience of waiting for me to cross.

That capacity to trust in unknown humans, not because of a belief in our innate goodness, but because of the systems that we humans have made, is the true genius of our species.

We can work together – because we can trust.

We can trade – because we can trust.

We can express opinions – because we can trust.

***

Now let's replace a fellow human in these day to day interactions with artificial intelligence: AI.

What would it take for you to put the same level of trust in AI as you would extend to a human?

To chat with your child?

To drive your taxi?

To scan a battlefield and zero in on a human target?

To scan the crowd at a concert to match faces to those from a crime database?

To monitor the key strokes and eye movements and concentration levels of every single one of thousands of employees in your workplace, and decide who gets to keep their jobs next week?

Every one of these things can be done today. *Is* being done today.

Show of hands: does anyone feel entirely comfortable?

***

What is it about AI that unnerves us?

I suspect it's a combination of two things.

First, we lack information. The knowledge seems concentrated in a priesthood: a cabal. The number of AI experts is tiny – an estimated 22,000 worldwide are qualified to the level of a PhD.

People with rare skills in high demand come at an eye-watering price.

That makes it very difficult to keep them in universities and public agencies.

It's not unknown for technology developers to buy up IT faculties, holus-bolus.

And whether these experts work in industry, or the public sector, or universities, there are often commercial or security reasons to keep quiet about their activities.

So what is this priesthood capable of creating?

Not knowing makes us uneasy.

And so does finding out.

Consider Google – which this month renamed its entire research division “Google AI”. Presumably, in case the message “AI is everything and everything is AI” needed to be made any clearer.

Google routinely astonishes the world by bulldozing supposedly hard limits of AI capability.

We had another example last week.

In footage beamed around the world, Google debuted an AI named Google Duplex that makes phone calls on your behalf and chats with the human who answers. Google Duplex chats in a human voice.

With a rising inflection and the occasional “mm-hmm”, to make it very unlikely that the real human would ever twig that the other voice belonged to a machine.

Now for nearly 70 years the world has been waiting for an AI that could fool us into thinking it was human.

We call it the Turing Test, named after the scientist who proposed it in 1950, Alan Turing.

He began with the question “can machines think?”

And since we can’t agree on a single test for “thinking”, he replaced it with a thought experiment he called the “imitation game”: can we build a machine that would pass as human, to humans, if we couldn’t see it, and judged it by its words?

Hearing Google Duplex book a haircut and a table at a restaurant, some observers say the Turing test was met.

And as a result, your world has changed. Every phone call you make or receive is going to have a niggling doubt that wasn’t there before. *That could be a machine*.

So, we lack knowledge of developments that can affect us immediately and directly.

***

Second, we lack foreknowledge. We give up our data today without knowing what others might be able to do with it tomorrow.

For example: when you uploaded your photos to FaceBook, did you expect that FaceBook would be able to scan them, name all the people, identify your tastes in food and clothing and hobbies, diagnose certain genetic disorders, like Down

Syndrome, decide on your personality type, and package all that information for advertisers?

Probably not.

But we can't unpick our choices.

And it could be that we and our children can't escape the implications.

***

One response to these questions would be to conclude that the only safe way forward is to ban AI.

But that would be a tragic mistake for Australia.

It wouldn't halt progress, because I find it difficult to believe that China and France and the United Kingdom and the United States and every other nation that has staked its future on AI would now step back from the race.

Why would they, when they see that for all the risks, and all the mistakes we will inevitably make, the future has infinite promise?

Healthcare, available to all: tailored precisely to the individual.

Your personal chauffer, a privilege previously only available to the super wealthy.

An AI assistant for each of us, to manage our appointments and remind us of the things we would doubtless forget.

We want those benefits in Australia.

A ban would simply discourage research and development in the places where we most want to see it: reputable institutions, like CSIRO's Data61, and our universities.

Of course, the fastest way to end up with a total ban is to allow a free-for-all. A free-for-all that allows unscrupulous and unthinking and just plain incompetent people to do their worst.

No: we want rules that allow us to trust AI, just as they allow us to trust our fellow humans.

So my question again: what would it take for you to extend your trust?

***

Think back to the web of rules that protected me when I stepped out onto Elizabeth Street this morning.

We could think of that web of rules as a spectrum.

On the extreme left, we have light touch rules: manners and customs.

As we move to the right, the rules become more binding.

We next see codes that govern behaviour in organisations, standards applied to manufactured products and regulations that apply to the practice of medicine, finance and trade.

Further along, the criminal law.

To the extreme right, international prohibitions against chemical and biological weapons.

Different human behaviours are regulated at different points, depending on their capacity for harm.

The challenge is to develop a similar spectrum for AI: not expecting a single solution, but instead an evolving web.

At the left-hand, light-touch end, consumer expectations for things like digital assistants.

At the other extreme, global governance for weapons of war: military devices that can select and kill their targets without a human in the loop.

At that extreme, Australians are already leading the discussion.

It was an Australian who put the issue on the global agenda in 2015: Professor Toby Walsh.

Through his network, he initiated a letter to the United Nations, calling for a global ban of autonomous weapons. It was signed by more than a thousand leaders of the global tech community, including Elon Musk and Stephen Hawking.

This year, Professor Walsh led the global boycott of a South Korean university that was planning to develop AI weapons in partnership with a defence manufacturer. The boycott was reported around the world. And it achieved its aim: the University's President agreed to bar all research directed to these ends.

It is an important reminder that we should never discount how influential Australian voices can be.

But I'll leave that particular discussion to Toby Walsh and others.

Instead, I've been thinking about the role that we could play at the left-hand end of the spectrum, the end of most immediate concern to you and me.

Our day to day interactions with technology providers, making the sort of thing that you might encounter in a workplace or a family home.

I look out and I see an audience of enthusiastic technology adopters.

Who among you believes your relationship with social media companies is built on trust and mutual respect?

The sector is working very hard to develop it. Google has an ethics board. Microsoft published a 150-page book on human-friendly AI development. The word of the moment in the tech sector is “responsible”.

These measures are surely positive, as far as they go. But who’s defining or auditing the standards? Are they making *you*, for example, feel any better?

Maybe a government department would vet the provider of a service that, for example, replaces social workers with algorithms that predict the likelihood that a particular child will be violent in school.

For the record, that’s not fiction: a child-violence predicting algorithm has been trialled.

But how many consumers are going to have the knowledge or the time to individually vet every AI they encounter?

When was the last time you read the terms and conditions before clicking “Accept”?

What we need is an agreed standard and a clear signal, so we individual consumers don’t need expert knowledge to make ethical choices, and so that companies know from the outset how they are expected to behave.

So my proposal is a trustmark. Its working title is ‘the Turing Certificate’, in honour of Alan Turing.

Companies can apply for Turing certification, and if they meet the standards and comply with the auditing requirements, they can display the Turing Stamp.

Then consumers and governments could use their purchasing power to reward and encourage ethical AI.

It works in other domains. I am sure everybody in this audience would prefer to purchase coffee that carries the ‘Fairtrade’ logo. That logo tells you an independent auditor has verified that the farmers received a fair price, and the coffee was produced without using child or slave labour.

And the Turing Certificate would do the same for ethical AI.

Independent auditors would certify the AI developers’ products, their business processes, and their ongoing compliance with clear and defined expectations.

***

The first response to a certification scheme is always “but that costs money”.

I understand that response, because I thought that way myself when I was building my company Axon Instruments in San Francisco.

I was making a device that was designed to be surgically inserted into people's brains. Living people's brains.

I expected to face a long and tortuous process to meet the exacting international ISO 9000 standards for good manufacturing practice.

What I discovered was that the standards were the scaffold I needed to build a competitive company.

They baked in the expectation of quality from the start.

True quality is achieved by design, not by test and reject.

We maintained these exacting design and business practices for our non-medical products, too, because they made us a better company and gave us a commercial edge.

Done right, the costs of securing certification should then be covered by increased sales, from customers prepared to pay a premium.

For example: the government.

Government agencies are already building AI into staff recruitment.

Others are exploring its potential in decision-making and service delivery.

Those contracts are likely to be extremely valuable for the companies that supply these capabilities.

So imagine if government demanded a Turing Stamp.

It would send a powerful signal.

***

Would the Turing Stamp be granted to organisations, or products?

Both.

It is the model that has long been accepted in the manufacturing sector.

And as you know, if you buy a washing machine, it will still be the same washing machine in five years' time. But if you download an app, it may well be radically different in five weeks.

So when you trade with an AI developer you expect to have an ongoing relationship.

I think you would want the company that stores your data and develops your upgrades to be ethical through and through.

Of course, in the manufacturing sector, the standards are both mandatory and enforceable: manufacturing is a highly visible process.

For AI, mandatory certification would be cumbersome. The voluntary Turing system would allow responsible companies to opt in.

A voluntary system does not mean self-certification. It means that the companies would voluntarily submit themselves to an external process. Smart companies, trading on quality, would welcome an auditing system that weeded out poor behaviour. And consumers would rightly insist on a stamp with proven integrity.

***

Is this a global measure that Australia could help to foster?

Surely, we have more to gain than most.

Where we compete in the global market, we compete on quality.

It's true in agriculture: our reputation secures a premium price.

It's true in higher education: our reputation is the foundation of our third biggest export industry.

And it should be true in technology.

A system that rewards quality and prioritises ethics will reward Australia.

Last week's Federal Budget made what I have described as a "promising first instalment".

$30 million has been allocated for AI, including an AI roadmap and a national AI ethics framework.

I hope we can use our influence to shape a responsible direction for the world.

***

So it may well be that in a few years' time, when I step onto Elizabeth Street, I'll be walking into the path of a self-driving car.

Let's make it a step that we can all take with trust.

THANK YOU