

Artificial Intelligence

ISSUES
IN SOCIETY

Edited by Justin Healey

 **THE SPINNEY PRESS**

First published by



PO Box 438 Thirroul NSW 2515 Australia
www.spinneypress.com.au

© The Spinney Press 2020.

COPYRIGHT

All rights reserved. Other than for purposes of and subject to the conditions prescribed under the Australian Copyright Act 1968 and subsequent amendments, no part of this publication may in any form or by any means (electronic, mechanical, microcopying, photocopying, recording or otherwise) be reproduced, stored in a retrieval system or transmitted without prior permission. Inquiries should be directed to the publisher.

REPRODUCTION AND COMMUNICATION FOR EDUCATIONAL PURPOSES

The Australian Copyright Act 1968 (the Act) allows a maximum of one chapter or 10% of the pages of this work, whichever is the greater, to be reproduced and/or communicated by any educational institution for its educational purposes provided that the educational institution (or the body that administers it) has given a remuneration notice to Copyright Agency Limited (CAL) under the Act.

For details of the CAL licence for educational institutions contact:
Copyright Agency Limited, Level 11, 66 Goulburn Street Sydney NSW 2000
Telephone: (02) 9394 7600 Fax: (02) 9394 7601 Email: info@copyright.com.au

REPRODUCTION AND COMMUNICATION FOR OTHER PURPOSES

Except as permitted under the Act (for example a fair dealing for the purposes of study, research, criticism or review) no part of this book may be reproduced, stored in a retrieval system, communicated or transmitted in any form or by any means without prior written permission. All inquiries should be made to the publisher at the address above.

Title: Artificial Intelligence / edited by Justin Healey.

Series: Issues in Society, Volume 450.

ISBN 978-1-922274-00-7 (paperback)

ISBN 978-1-922274-01-4 (PDF)



A catalogue record for this
book is available from the
National Library of Australia

Cover images courtesy of iStock.

CHAPTER 1	ARTIFICIAL INTELLIGENCE: ETHICS AND REGULATION	
	Explainer: what is artificial intelligence?	1
	Artificial intelligence	3
	Artificial intelligence: 10 ways society will change by 2050	4
	Your questions answered on artificial intelligence	6
	History of artificial intelligence	11
	Artificial intelligence: Australia’s ethics framework	12
	Will we ever agree to just one set of rules on the ethical development of AI?	17
	How do we make artificial intelligence more humane?	19
	Protecting human rights in the context of AI	21
CHAPTER 2	FUTURE FEARS: THE RISE OF ROBOTS	
	Rise of the machines	23
	Five reasons why robots won’t take over the world	25
	Advances in AI and robotics	26
	How robots change the world	27
	Artificial intelligence may take your job, so political leaders need to start doing theirs	29
	Overcoming our mistrust of robots in our homes and workplaces	32
	We should learn to work with robots and not worry about them taking our jobs	34
	Ethics of AI: how should we treat rational, sentient robots – if they existed?	36
	Robot rights and humans	37
	Artificial intelligence is being trained to have empathy. Should we be worried?	38
	Careful how you treat today’s AI: it might take revenge in the future	40
	Before replacing a carer with a robot, we need to assess the pros and cons	42
	Driverless cars: everything you need to know about the transport revolution	44
	Driverless vehicles: pros and cons	46
	Controlling killer robots: how do we do it?	47
	Killer robots: why banning autonomous weapons is not a good idea	49
	Exploring issues – worksheets and activities	51
	Fast facts	57
	Glossary	58
	Web links	59
	Index	60

INTRODUCTION

Artificial Intelligence is Volume 450 in the 'Issues in Society' series of educational resource books. The aim of this series is to offer current, diverse information about important issues in our world, from an Australian perspective.

KEY ISSUES IN THIS TOPIC

Artificial intelligence (AI) has been quietly evolving from futuristic scenarios in sci-fi movie plots and into the factual present day. Put simply, AI is a computer system that can do tasks that humans need intelligence to do. AI is already transforming the way we live, work, travel and communicate. From the development of driverless cars, to the machine learning and algorithms that already power our mobile phones, voice-activated assistants and the Internet of Things, we are co-existing and engaging daily with AI-enabled technology, including robots.

However, along with the potential to improve our quality of life, AI risks compromising some human rights and freedoms. We all need to better understand how AI is being developed, used and regulated as it becomes more embedded and empowered in the world around us.

What is artificial intelligence, how does it work, and what are the benefits and ethical concerns? The robot revolution is happening right now – how much do you really know about the impacts of AI on your everyday life?

SOURCES OF INFORMATION

Titles in the 'Issues in Society' series are individual resource books which provide an overview on a specific subject comprised of facts and opinions.

The information in this resource book is not from any single author, publication or organisation. The unique value of the 'Issues in Society' series lies in its diversity of content and perspectives.

The content comes from a wide variety of sources and includes:

- Newspaper reports and opinion pieces
- Website fact sheets
- Magazine and journal articles
- Statistics and surveys
- Government reports
- Literature from special interest groups

CRITICAL EVALUATION

As the information reproduced in this book is from a number of different sources, readers should always be aware of the origin of the text and whether or not the source is likely to be expressing a particular bias or agenda.

It is hoped that, as you read about the many aspects of the issues explored in this book, you will critically evaluate the information presented. In some cases, it is important that you decide whether you are being presented with facts or opinions. Does the writer give a biased or an unbiased report? If an opinion is being expressed, do you agree with the writer?

EXPLORING ISSUES

The 'Exploring issues' section at the back of this book features a range of ready-to-use worksheets relating to the articles and issues raised in this book. The activities and exercises in these worksheets are suitable for use by students at middle secondary school level and beyond.

FURTHER RESEARCH

This title offers a useful starting point for those who need convenient access to information about the issues involved. However, it is only a starting point. The 'Web links' section at the back of this book contains a list of useful websites which you can access for more reading on the topic.

Explainer: what is artificial intelligence?

AI has jumped from sci-fi movie plots into mainstream news headlines in just a couple of years, according to this *ABC News* report by **Margot O'Neill**

And the headlines are often contradictory. AI is either a technological leap into greater prosperity or mass unemployment; it will either be our most valuable servant or terrifying master.

But what is AI, how does it work, and what are the benefits and the concerns?

WHAT IS ARTIFICIAL INTELLIGENCE?

AI is a computer system that can do tasks that humans need intelligence to do.

“An intelligent computer system could be as simple as a program that plays chess or as complex as a driverless car,” Mary-Anne Williams, professor of social robotics at the University of Technology, Sydney, said.

A driverless car, for example, relies on multiple sensors to understand where it is and what’s around it. These include speed, location, direction and 360-degree vision. Based on those inputs, among others, the “intelligent” computer system controls the car by deciding, like a human would, when to turn the steering and when to

accelerate or brake.

Then there’s machine learning, a subset of AI, which involves teaching computer programs to learn by finding patterns in data. The more data, the more the computer system improves.

“Whether it’s recognising objects, identifying people in photos, reading lung scans or transcribing spoken mandarin, if we pick a narrow task like that [and] we give it enough data, the computer learns to do it as well as, if not better, than us,” University of New South Wales professor of artificial intelligence Toby Walsh said.

AI doesn’t have to sleep or make the same mistake twice. It can also access vast troves of digital data in seconds. Our brains cannot.

An intelligent computer system could be as simple as a program that plays chess or as complex as a driverless car.



DO I ALREADY USE AI?

Yes, probably every day. AI is in your smart phone; it's there every time you ask a question of iPhone's Siri or Amazon's Alexa. It's in your satellite navigation system and instant translation apps.

AI algorithms recognise your speech, provide search results, help sort your emails and recommend what you should buy, watch or read.

"AI is the new electricity," according to Andrew Ng, former chief scientist at Baidu, one of the leading Chinese web services companies. AI will increasingly be all around you from your phone to your TV, car and home appliances.

WHY ARE WE TALKING ABOUT IT NOW?

Four factors have now converged to push AI beyond games and into our everyday lives and workplaces:

- Computer processing power is doubling every two years (known as Moore's Law)
- The amount of data being generated is doubling every year (AI algorithms are hungry for data)
- Recently, the amount of AI funding has also been doubling every two years
- There is now 50 years of established AI research, giving us better and better algorithms.

The term artificial intelligence was first coined in 1956 by US computer scientist John McCarthy. Until

recently, the public mostly heard about AI in Hollywood movies like *The Terminator* or whenever it defeated a human in a competition.

In 1997, IBM's Deep Blue computer beat Russian chess master Garry Kasparov. In 2011, IBM's super-computer Watson beat human players on the US game show *Jeopardy!*. Last year, Google's AlphaGo beat Go master Lee Sedol.

"We now have the computing power, the data, the algorithms and a lot of people working on the problems," Professor Walsh said.

CAN AI HELP US?

AI promises spectacular benefits for humanity, including better and more precise medical diagnosis and treatment; relieving the drudgery and danger of repetitive and dehumanising jobs; and super-charging decision making and problem solving.

"Driverless cars could save many, many lives because 95 per cent of accidents are due to human error," Professor Walsh said.

"Many of the problems that are stressing our planet today will be tackled through having better decision making with computers that access and analyse vast troves of data," he said.

BUT CAN IT ALSO HURT US?

There are a range of concerns:

- That the AI and robotics revolution might create mass unemployment inside a generation
- That AI will further undermine privacy and democracy through greater mass surveillance by governments and companies
- That we will be more easily manipulated by personalised algorithms creating fake news

Machine learning, a subset of AI, involves teaching computer programs to learn by finding patterns in data. The more data, the more the computer system improves.



ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) is a broad term used to describe a collection of technologies that can solve problems and perform tasks to achieve defined objectives without explicit human guidance.

Central to AI are automation and machine learning that underpin applications such as natural language processing (Apple Siri or Amazon Alexa), computer vision (Tesla Autopilot), and optimisation and decision support (Google Maps). AI has the potential to automate repetitive or dangerous tasks, increase productivity and allow the development of innovative consumer products. It is forecast to add trillions of dollars to the global economy in the coming decades.

Examples include:

- Using advanced data analytics techniques to diagnose diseases at earlier and more treatable stages
- Using automated machines for hauling and drilling on mine sites, increasing productivity and reducing risks to workers
- Enabling greater use of smart forms that can tailor legal information to assist individuals to draft a will or settle financial arrangements following a break-up
- Tailoring content on entertainment platforms to meet user preferences.

There is considerable effort, both in Australia and internationally, focused on ensuring that AI is applied ethically and delivers broad societal benefits.

Department of Industry, Innovation and Science. *Understanding emerging technologies*. Retrieved from www.industry.gov.au on 9 July 2019.

- That algorithms will be biased but will be used to decide important issues in our lives such as insurance claims, job applications, loan applications and even judicial sentencing.

THAT ALL SOUNDS BAD. SO, WILL IT OVERTAKE HUMANITY?

Experts are famously split on this. Prominent tech entrepreneurs and scientists such as Elon Musk and Stephen Hawking, among others, warn that AI could reach and quickly surpass humans, transforming into super-intelligence that would render us the second most intelligent species on the planet.

Musk has compared it to “summoning the demon”. Scientists call it ‘singularity’, “where machines improve themselves almost without end,” Professor Walsh said.

Facebook’s Mark Zuckerberg accuses Musk of being alarmist. Professor Walsh says we don’t yet even fully understand all the facets of human intelligence and there may be limits to how far AI can develop.

He’s surveyed 300 of his AI colleagues around the world and most believe if AI can reach human level intelligence, it is at least 50 to 100 years away.

If it happens, humanity will likely have already solved most of the problems about whether the machines’ values are aligned with ours. “I’m not so worried about that,” he says.

AND WHO CONTROLS IT?

The recent push into AI came from big US tech companies such as Google, Facebook, Amazon, Microsoft and Apple. And the US military. What could go wrong?

There’s growing concern that these companies are too big and control too much data, which trains the AI algorithms.

China has now also joined the race with plans to dominate the world in AI development by 2030.

There’s presently very little national or international regulation around how AI is developed. The Big Tech

AI promises spectacular benefits for humanity, including better and more precise medical diagnosis and treatment; relieving the drudgery and danger of repetitive and dehumanising jobs; and super-charging decision making and problem solving.

companies have begun discussing the need for guiding principles to ensure AI is only used for public good.

“One of those is what is the point of AI? It has to be to augment people, to support people, not replace them,” Microsoft Australia national technology officer James Kavanagh says.

“Secondly, it has to be democratised. It can’t be in the hands of a small number of technology companies.

“Thirdly, it has to be built on foundations of trust. We need to be able to understand any biases in algorithms and how they make decisions.”

© ABC. Reproduced by permission of the Australian Broadcasting Corporation – Library Sales.

O’Neill, M (7 August 2017). ‘Explainer: What is artificial intelligence?’, *ABC News*. Retrieved from www.abc.net.au/news on 9 July 2019.

ARTIFICIAL INTELLIGENCE: 10 WAYS SOCIETY WILL CHANGE BY 2050

Leading Australian artificial intelligence scientist **Professor Toby Walsh** is warning that we are “sleepwalking” into an AI future in which billions of machines and computers will be able to think.

Professor Walsh, from the University of New South Wales, is calling for a national discussion about whether society needs to adopt clear boundaries and guidelines around how AI is developed and how it’s used in our lives.

In his book *It’s Alive: Artificial Intelligence from The Logic Piano to Killer Robots*, he has highlighted key questions in a series of predictions that describe how our future could be far better or far worse because of AI.

Here’s how he thinks society might change by 2050, thanks to artificial intelligence.

1. YOU ARE BANNED FROM DRIVING

Humans drive drunk, tired and distracted and cause 95 per cent of accidents. The roads will be much safer without human drivers and most likely far less congested, as networked vehicles service passengers 24/7.

Street parking and most car parks will disappear, transport will be cheaper and groups such the elderly and disabled will have greater personal mobility.

Most people won’t bother buying cars and will lose driving skills. And autonomous vehicles will arrive quickly – within 15-20 years.

“By 2050, the year 2000 will look as quaintly old-fashioned as the horse-drawn era of 1900 did to people in 1950,” Professor Walsh said.

2. YOU SEE THE DOCTOR DAILY

Your personal 24/7 AI doctor will know your gene sequence and vulnerabilities to particular diseases. It will continually monitor your blood pressure, sugar levels, sleep and exercise. It will process data from your toilet, which will automatically analyse your urine and stools.

Your future version of a smartphone or fitness watch will regularly take selfies to identify melanomas and eye disease. It will record your voice for signs of a cold, dementia or a stroke. It will call for help if you faint. It will also be a trillion-dollar global business.

“Our personal AI physician will have our life history, it will know far more about medicine than any single doctor, and it will stay on top of all the emerging medical literature,” Professor Walsh said.

3. MARILYN MONROE IS BACK IN THE MOVIES

Avatars will be programmed to act and talk like anyone we choose in interactive movies, including ourselves or celebrities from recent history. Where the story goes depends on what you do or say.

Hollywood and the computer games industry will

merge and immerse us in hyper-real worlds. But there will be increasing concern about the seductive nature of these unreal, alternate worlds. There may be an underclass of addicts who spend every waking moment in them. And some who behave in distasteful or illegal ways.

“This problem will likely trouble our society greatly,” Professor Walsh said.

“There will be calls that behaviours which are illegal in the real world should be made illegal or impossible in the virtual.”

4. A COMPUTER HIRES AND FIRES YOU

That’s just the beginning. AI systems will also increasingly take over managing how you work: scheduling your activities, approving holidays, monitoring and rewarding your performance.

“By 2050, the year 2000 will look as quaintly old-fashioned as the horse-drawn era of 1900 did to people in 1950.”

But should we hand over decisions like hiring and especially firing to a computer?

“We will have to learn when to say to computers: ‘Sorry, I can’t let you do that.’ It’s not enough for a machine to do a task better than a human. There are some decisions we simply should not allow machines to make.”

5. YOU TALK TO ROOMS

You will walk into a room and say “lights on” and “who won the football?” and one of the many AI devices in your house will recognise your voice and understand you well enough to know which football code you follow.

A few people will resist and determinedly follow a disconnected 20th century life. But most of us will take advantage of having just about everything in our lives connected: fridges, toasters, baths, door locks, windows, bicycles and pot plants.

AI will operate through the so-called Internet of Things using conversation instead of typing.

“Our privacy, diversity and democracy will be challenged,” Professor Walsh said.

“[Government] intelligence [agencies] can’t wait for every room to be listening to us. Marketers, too, would love all this data about our everyday lives.

“So, the next time you get asked to check your privacy settings, think long and hard about what you may be giving up.”

6. A ROBOT ROBS A BANK

Cyber-crime to date has been relatively low-tech with phishing and malware attacks. But AI will surpass human hackers – and the only defence will be another AI program.

Warfare is also moving into cyberspace. But these technologies will also quickly find their way into the civilian sphere. One of the challenges will be that many advances in AI used to defend systems will be quickly turned around to attack systems.

“The supposed hacking by Russians in order to influence the 2016 US presidential election demonstrates the impact that such cyber attacks can have,” Professor Walsh said.

“Banks [and other companies and governments] will have no choice but to invest more and more in sophisticated AI systems to defend themselves from attack.”

7. WORLD SOCCER CHAMPIONS LOSE TO A ROBOT TEAM

Robots will have superior ball skills, including unfailing accuracy in passes and penalties. They will know precisely where all players are at all times and will know how to interpret that information because their AI system learned strategic play from watching every World Cup match ever recorded.

The human team will be soundly defeated. Even fans of the robots will call for the humans to be given a break. That’s why most sporting teams will stay human. But AI will change football and most other games with managers and players using AI to train and play better.

“Data scientists will be some of the best paid members of football [and other sporting] clubs,” Professor Walsh said.

“Scouts will hang out at [top universities] to recruit young computer scientists.”

8. GHOST SHIPS, PLANES AND TRAINS CROSS THE GLOBE

The oceans, skies and railroads of the planet will be filled with autonomous ships, planes and trains transporting cargo without any people on board, as driverless car technology spreads to other industries. It will improve safety and efficiency. And children will no longer grow up wanting to be train drivers.

“Planes carrying people will probably continue to be piloted by humans,” Professor Walsh said.

“But after several decades of safe flights by cargo planes, the debate will begin whether humans should still be airline pilots.”

9. TV NEWS IS MADE WITHOUT HUMANS

Nearly every part of this prediction is already here – it’s just that no one has yet pulled all the pieces together. Computers now write simple sport and financial stories but as technology improves, AI will write more complex stories. Avatars and chatbots will play the role of presenters filmed by robotic cameras. And the



news you watch will be narrowcast, or tailored to your personal preferences.

“There will be ongoing debate about the biases of algorithms, especially when humans take no part in deciding what news we see,” Professor Walsh said.

“Our viewpoints are shaped by the lens through which we look at the world. Will algorithms challenge us enough? Will they understand lies and deception? Will they care about what we care about?”

10. HUMANS LIVE ON AFTER DEATH

It will be common to leave behind an AI chatbot that will talk like you, know the story of your life and comfort your family when you die.

Some people might give their chatbot the task of reading their will; settling old scores; or relieving grief through humour.

Digital doubles will also appear in place of the living. Celebrities will use bots to create social media; many of us will similarly use them to manage our diaries.

This digital outsourcing will fuel a lively debate.

Professor Walsh asked: “What redress do you have against an AI bot that pretends to be you? Do you have a right to know if you’re interacting with a computer rather than a real person? Should AI bots be prohibited from political discourse? Who can switch off your bot after you die? Do bots have freedom of speech? It will be an interesting future.”

© ABC. Reproduced by permission of the Australian Broadcasting Corporation – Library Sales.

Walsh, T (7 August 2017). ‘Artificial intelligence: Professor Toby Walsh on 10 ways society will change by 2050’, *ABC News*. Retrieved from www.abc.net.au/news on 9 July 2019.

YOUR QUESTIONS ANSWERED ON ARTIFICIAL INTELLIGENCE

*Artificial intelligence and robotics have enjoyed a resurgence of interest, and there is renewed optimism about their place in our future. But what do they mean for us? Following are answers to questions put to experts about AI and robotics, first published by **The Conversation***

Q1. How plausible is human-like artificial intelligence?

A. Toby Walsh, Professor of AI

It is 100% plausible that we'll have human-like artificial intelligence. I say this even though the human brain is the most complex system in the universe that we know of. There's nothing approaching the complexity of the brain's billions of neurons and trillions of connections. But there are also no physical laws we know of that would prevent us reproducing or exceeding its capabilities.

A. Kevin Korb, Reader in Computer Science

Popular fiction AI from Issac Asimov to Steven Spielberg is plausible. What the question doesn't address is: when will it be plausible?

Most AI researchers (including me) see little or no evidence of it coming anytime soon. Progress on the major AI challenges is slow, if real.

What I find less plausible than the AI in fiction is the emotional and moral lives of robots. They seem to be either unrealistically empty, such as the emotionless Data in *Star Trek*, or unrealistically human-identical or superior, such as the AI in Spike Jonze's *Her*.

All three – emotion, ethics and intelligence – travel together, and are not genuinely possible in some form

without the others, but fiction writers tend to treat them as separate. Plato's Socrates made a similar mistake.

A. Gary Lea, Researcher in Artificial Intelligence Regulation

AI is not impossible, but the real issue is: "how like is like?" The answer probably lies in applied tests: the Turing test was already (arguably) passed in 2014 but there is also the coffee test (can an embodied AI walk into an unfamiliar house and make a cup of coffee?), the college degree test and the job test.

If AI systems could progressively pass all of those tests (plus whatever else the psychologists might think of), then we would be getting very close. Perhaps the ultimate challenge would be whether a suitably embodied AI could live among us as J. Average and go undetected for five years or so before declaring itself.

Q2. Automation is already replacing many jobs. Is it time to make laws to protect some of these industries?

A. Jonathan Roberts, Professor of Robotics

Researchers at the University of Oxford published a now well-cited paper in 2013 that ranked jobs in order of how feasible it was to computerise or automate them. They found that nearly half of jobs in the USA could be at risk from computerisation within 20 years.

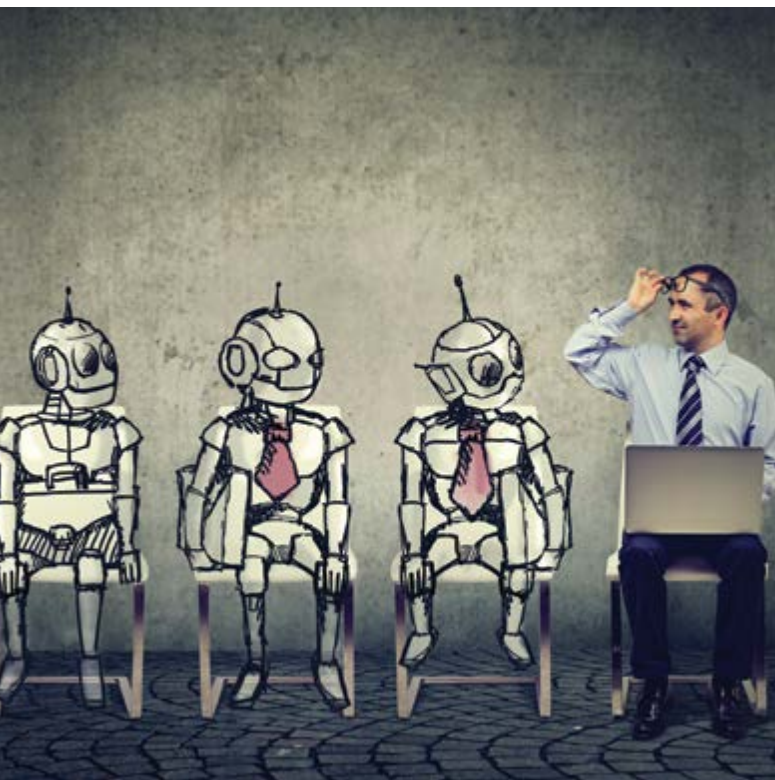
This research was followed in 2014 by the viral video hit, *Humans Need Not Apply*, which argued that many jobs will be replaced by robots or automated systems and that employment would be a major issue for humans in the future.

Of course, it is difficult to predict what will happen, as the reasons for replacing people with machines are not simply based around available technology. The major factor is actually the business case and the social attitudes and behaviour of people in particular markets.

A. Rob Sparrow, Professor of Philosophy

Advances in computing and robotic technologies are undoubtedly going to lead to the replacement of many jobs currently done by humans. I'm not convinced that we should be making laws to protect particular industries though. Rather, I think we should be doing two things.

First, we should be making sure that people are assured of a good standard of living and an opportunity to pursue meaningful projects even in a world in which many more jobs are being done by machines. After all, the idea that, in the future, machines would work so that human beings didn't have to toil used to



be a common theme in utopian thought.

When we accept that machines putting people out of work is bad, what we are really accepting is the idea that whether ordinary people have an income and access to activities that can give their lives meaning should be up to the wealthy, who may choose to employ them or not. Instead, we should be looking to redistribute the wealth generated by machines in order to reduce the need for people to work without thereby reducing the opportunities available to them to be doing things that they care about and gain value from.

Second, we should be protecting vulnerable people in our society from being treated worse by machines than they would be treated by human beings. With my mother, Linda Sparrow, I have argued that introducing robots into the aged care setting will most likely result in older people receiving a worse standard of treatment than they already do in the aged care sector. Prisoners and children are also groups who are vulnerable to suffering at the hands of robots introduced without their consent.

A. Toby Walsh, Professor of AI

There are some big changes about to happen. The #1 job in the US today is truck driver. In 30 years time, most trucks will be autonomous.

How we cope with this change is a question not for technologists like myself but for society as a whole. History would suggest that protectionism is unlikely to work. We would, for instance, need every country in the world to sign up.

But there are other ways we can adjust to this brave new world. My vote would be to ensure we have an educated workforce that can adapt to the new jobs that technology create.

We need people to enter the workforce with skills for jobs that will exist in a couple of decades time when the technologies for these jobs have been invented.

We need to ensure that everyone benefits from the rising tide of technology, not just the owners of the robots. Perhaps we can all work less and share the economic benefits of automation? This is likely to require fundamental changes to our taxation and welfare system informed by the ideas of people like the economist Thomas Piketty.

A. Kevin Korb, Reader in Computer Science

Industrial protection and restriction are the wrong way to go. I'd rather we develop our technology so as to help solve some of our very real problems. That's bound to bring with it economic dislocation, so a caring society will accommodate those who lose out because of it.

But there's no reason we can't address that with improving technology as long as we keep the oligarchs under control. And if we educate people for flexibility rather than to fit into a particular job, intelligent people will be able to cope with the dislocation.

A. Jai Galliot, Defence Analyst

The standard argument is that workers displaced by



automation go on to find more meaningful work. However, this does not hold in all cases.

Think about someone who signed up with the air force to fly jets. These pilots may have spent their whole social, physical and psychological lives preparing or maintaining readiness to defend their nation and its people.

For service personnel, there are few higher-value jobs than serving one's nation through rendering active military service on the battlefield, so this assurance of finding alternative and meaningful work in a more passive role is likely to be of little consolation to a displaced soldier.

Thinking beyond the military, we need to be concerned that the Foundation for Young Australians indicates that as many as 60% of today's young people are being trained for jobs that will soon be transformed due to automation.

The sad fact of the matter is that one robot can replace many workers. The future of developed economies therefore depends on youth adapting to globalised and/or shared jobs that are increasingly complemented by automation within what will inevitably be an innovation and knowledge economy.

Q3. Where will AI be in five to ten years?

A. Toby Walsh, Professor of AI

AI will become the operating system of all our connected devices. Apps like Siri and Cortana will morph into the way we interact with the connected world.

AI will be the way we interact with our smartphones, cars, fridges, central heating system and front door. We will be living in an always-on world.

A. Jonathan Roberts, Professor of Robotics

It is likely that in the next five to ten years we will see machine learning systems interact with us in the form of robots. The next large technology hurdle that must be overcome in robotics is to give them the power of sight.

This is a grand challenge and one that has filled the research careers of many thousands of robotics researchers over the past four or five decades. There is a growing feeling in the robotics community that machine learning using large datasets will finally crack some of the problems in enabling a robot to actually see.

Four universities have recently teamed up in Australia in an ARC funded Centre of Excellence in Robotic Vision. Their mission is to solve many of the problems that prevent robots seeing.

Q4. Should we be concerned about military and other armed robots?

A. Rob Sparrow, Professor of Philosophy

The last thing humanity needs now is for many of its most talented engineers and roboticists to be working on machines for killing people.

Robotic weapons will greatly lower the threshold of conflict. They will make it easier for governments to start wars because they will hold out the illusion of being able to fight without taking any casualties. They will increase the risk of accidental war because militaries will deploy unmanned systems in high-threat environments, where it would be too risky to place a human being, such as just outside a potential enemy's airspace or deep sea ports.

In these circumstances, robots may even start wars without any human being having the chance to veto the decision. The use of autonomous robots to kill people

threatens to further erode respect for human life.

It was for these reasons that, with several colleagues overseas, I co-founded the International Committee for Robot Arms Control, which has in turn supported the Campaign to Stop Killer Robots.

A. Toby Walsh, Professor of AI

"Killer robots" are the next revolution in warfare, after gunpowder and nuclear bombs. If we act now, we can perhaps get a ban in place and prevent an arms race to develop better and better killer robots.

A ban won't uninvent the technology. It's much the same technology that will go, for instance, into our autonomous cars. And autonomous cars will prevent the 1,000 or so deaths on the roads of Australia each year.

But a ban will associate enough stigma with the technology that arms companies won't sell them, that arms companies won't develop them to be better and better at killing humans. This has worked with a number of other weapon types in the past like blinding lasers. If we don't put a ban in place, you can be sure that terrorists and rogue nations will use killer robots against us.

For those who argue that killer robots are already covered by existing humanitarian law, I profoundly disagree. We cannot correctly engineer them today not to cause excessive collateral damage. And in the future, when we can, there is little stopping them being hacked and made to behave unethically. Even used lawfully, they will be weapons of terror.

You can learn more about these issues by watching my TEDx talk on this topic, *How can you stop killer robots?*

A. Sean Welsh, Researcher in Robot Ethics

We should be concerned about military robots. How-



ever, we should not be under the illusion that there is no existing legislation that regulates weaponised robots.

There is no specific law that bans murdering with piano wire. There is simply a general law against murder. We do not need to ban piano wire to stop murders. Similarly, existing laws already forbid the use of any weapons to commit murder in peacetime and to cause unlawful deaths in wartime.

There is no need to ban autonomous weapons as a result of fears that they may be used unlawfully any more than there is a need to ban autonomous cars for fear they might be used illegally (as car bombs). The use of any weapon that is indiscriminate, disproportionate and causes unnecessary suffering is already unlawful under international humanitarian law.

Some advocate that autonomous weapons should be put in the same category as biological and chemical weapons. However, the main reason for bans on chemical and biological weapons is that they are inherently indiscriminate (cannot tell friend from foe from civilian) and cause unnecessary suffering (slow painful deaths). They have no humanitarian positives.

By contrast, there is no suggestion that “killer robots” (even in the examples given by opponents) will necessarily be indiscriminate or cause painful deaths. The increased precision and accuracy of robotic weapons systems compared to human operated ones is a key point in their favour.

If correctly engineered, they would be less likely to cause collateral damage to innocents than human-operated weapons. Indeed robot weapons might be engineered so as to be more likely to capture rather than kill. Autonomous weapons do have potential humanitarian positives.

Q5. How plausible is super-intelligent AI?

A. David Dowe, Associate Professor in Machine Learning and Artificial Intelligence

We can look at the progress made at various tasks once said to be impossible for machines to do, and see them one by one gradually being achieved. For example: beating the human world chess champion (1997); winning at *Jeopardy!* (2011); driverless vehicles, which are now somewhat standard on mining sites; automated translation, etc.

And, insofar as intelligence test problems are a measure of intelligence, I've recently looked at how computers are performing on these tests.

A. Rob Sparrow, Professor of Philosophy

If there can be artificial intelligence then there can be super-intelligent artificial intelligences. There doesn't seem to be any reason why entities other than human beings could not be intelligent. Nor does there seem to be any reason to think that highest human IQ represents the upper limit on intelligence.

If there is any danger of human beings creating such machines in the near future, we should be very scared. Think about how human beings treat rats. Why should machines that were as many times more intelligent

than us, as we are more intelligent than rats, treat us any better?

Q6. Given what little we know about our own minds, can we expect to intentionally create artificial consciousness?

A. Kevin Korb, Reader in Computer Science

As a believer in functionalism, I believe it is possible to create artificial consciousness. It doesn't follow that we can “expect” to do it, but only that we might.

John Searle's arguments against the possibility of artificial consciousness seem to confuse functional realisability with computational realisability. That is, it may well be (logically) impossible to “compute” consciousness, but that doesn't mean that an embedded, functional computer cannot be conscious.

A. Rob Sparrow, Professor of Philosophy

A number of engineers, computer scientists, and science fiction authors argue that we are on the verge of creating artificial consciousness. They usually proceed by estimating the number of neurons in the human brain and pointing out that we will soon be able to build computers with a similar number of logic gates.

If you ask a psychologist or a psychiatrist, whose job it is to actually “fix” minds, I think you will likely get a very different answer. After all, the state-of-the-art treatment for severe depression still consists of shocking the brain with electricity, which looks remarkably like trying to fix a stalled car by pouring petrol over the top of the engine. So I'm sceptical that we understand enough about the mind to design one.

Q7. How do cyborgs differ (technically or conceptually) from AI?

A. Katina Michael, Associate Professor in Information Systems

A cyborg is a human-machine combination. By definition, a cyborg is any human who adds parts, or enhances his or her abilities by using technology. As we have advanced our technological capabilities, we have discovered that we can merge technology onto and into the human body for prosthesis and/or amplification. Thus, technology is no longer an extension of us, but “becomes” a part of us if we opt into that design.

In contrast, artificial intelligence is the capability of a computer system to learn from its experiences and simulate human intelligence in decision-making. A cyborg usually begins as a human and may undergo a transformational process, whereas artificial intelligence is imbued into a computer system itself predominantly in the form of software.

Some researchers have claimed that a cyborg can also begin in a humanoid robot and incorporate the living tissue of a human or other organism. Regardless, whether it is a human-to-machine or machine-to-organism coalescence, when AI is applied via silicon microchips or nanotechnology embedded into prosthetic forms like a dependent limb, a vital organ, or a replacement/

additional sensory input, a human or piece of machinery is said to be a cyborg.

There are already early experiments with such cybernetics. In 1998 Professor Kevin Warwick named his first experiment Cyborg 1.0, surgically implanting a silicon chip transponder into his forearm. In 2002 in project Cyborg 2.0, Warwick had a one hundred electrode array surgically implanted into the median nerve fibres of his left arm.

Ultimately we need to be extremely careful that any artificial intelligence we invite into our bodies does not submerge the human consciousness and, in doing so, rule over it.

Q8. Are you generally optimistic or pessimistic about future of artificial intelligence and its benefits for humanity?

A. Toby Walsh, Professor of AI

I am both optimistic and pessimistic. AI is one of humankind's truly revolutionary endeavours. It will transform our economies, our society and our position in the centre of this world. If we get this right, the world will be a much better place. We'll all be healthier, wealthier and happier.

Of course, as with any technology, there are also bad paths we might end up following instead of the good ones. And unfortunately, humankind has a track record of late of following the bad paths.

We know global warming is coming but we seem unable not to follow this path. We know that terrorism is fracturing the world but we seem unable to prevent this. AI will also challenge our society in deep and fundamental ways. It will, for instance, completely change the nature of work. Science fiction will soon be science fact.

A. Rob Sparrow, Professor of Philosophy

I am generally pessimistic about the long-term impact of artificial intelligence research on humanity. I don't want to deny that artificial intelligence has many benefits to offer, especially in supporting human beings to make better decisions and to pursue scientific goals that are currently beyond our reach. Investigating how brains work by trying to build machines that can do what they do is an interesting and worthwhile project in its own right.

However, there is a real danger the systems that AI researchers come up with will mainly be used to further enrich the wealthy and to entrench the power of the powerful. I also think there is a risk that the prospect of AI will allow people to delude themselves that we don't need to do something about climate change now. It may also distract them from the fact that we already know what to do, but we lack the political will to do it.

Finally, even though I don't think we've currently got much of a clue of how this might happen, if engineers do eventually succeed in creating genuine AIs that are smarter than we are, this might well be a species-level extinction threat.

A. Jonathan Roberts, Professor of Robotics

I am generally optimistic about the long-term future of AI to humanity. I think that AI has the potential to radically change humanity and hence, if you don't like change, you are not going to like the future.

I think that AI will revolutionise health care, especially diagnosis, and will enable the customisation of medicine to the individual. It is very possible that AI GPs and robot doctors will share their knowledge as they acquire it, creating a super doctor that will have access to all the medical data of the world.

I am also optimistic because humans tend to recognise when technology is having major negative consequences, and we eventually deal with it. Humans are in control and will naturally try and use technology to make a better world.

A. Kevin Korb, Reader in Computer Science

I'm pessimistic about the medium-term future of humanity. I think climate change and attendant displacements, wars etc. may well massively disrupt science and technology. In that case progress on AI may stop.

If that doesn't happen, then I think progress will continue and we'll achieve AI in the long-term. Along the way, AI research will produce spin-offs that help economy and society, so I think as long as it exists AI tech will be important.

A. Gary Lea, Researcher in Artificial Intelligence Regulation

I suspect the long-term future for AI will turn out to be the usual mixed bag: some good, some bad. If scientists and engineers think sensibly about safety and public welfare when making their research, design and build choices (and provided there are suitable regulatory frameworks in place as a backstop), I think we should be okay.

So, on balance, I am cautiously optimistic on this front – but there are many other long-term existential risks for humanity.

Toby Walsh is Professor of AI, Research Group Leader, Optimisation Research Group, Data61. **David Dowe** is Associate Professor, Clayton School of Information Technology, Monash University. **Gary Lea** is Visiting Researcher in Artificial Intelligence Regulation, Australian National University. **Jai Galliot** is Research Fellow in Indo-Pacific Defence, UNSW. **Jonathan Roberts** is Professor in Robotics, Queensland University of Technology. **Katina Michael** is Associate Professor, School of Information Systems and Technology, University of Wollongong. **Kevin Korb** is Reader in Computer Science, Monash University. **Robert Sparrow** is Professor, Department of Philosophy; Adjunct Professor, Centre for Human Bioethics, Monash University. **Sean Welsh** is Doctoral Candidate in Robot Ethics, University of Canterbury.

THE CONVERSATION

Walsh, T, Dowe, D, Lea, G, Galliot, J, Roberts, J, Michael, K, Korb, K, Sparrow, R, and Welsh, S (16 November 2015). *Your questions answered on artificial intelligence*. Retrieved from <http://theconversation.com> on 9 July 2019.

HISTORY OF ARTIFICIAL INTELLIGENCE

A BRIEF HISTORY OF AI FROM THE QUEENSLAND BRAIN INSTITUTE

Modern AI began in the 1950s with the view to solving complex mathematical problems and creating 'thinking machines'.

From the outset, there were two competing approaches. One used formal rules to manipulate symbols, a logic-based approach not at all based on biology. This became known as 'good old-fashioned artificial intelligence': GOFAL. The other camp took inspiration from how the brain works and created 'artificial neural networks' loosely inspired by our brains. These still had to be trained using certain procedures to solve problems.

In the first 20 years, GOFAL was the more successful approach, leading to much hype and significant government funding. But in real-world settings GOFAL didn't achieve its outcomes. Artificial neural networks also struggled, and in the 1970s funding dried up, research slowed and the AI community shrank.

In the 1980s, improvements were made in both the rules-based GOFAL systems and biologically-inspired neural networks. Previously difficult problems became achievable and AI seemed promising once again. However, the hope and hype exceeded reality, and by

the 1990s AI research again diminished.

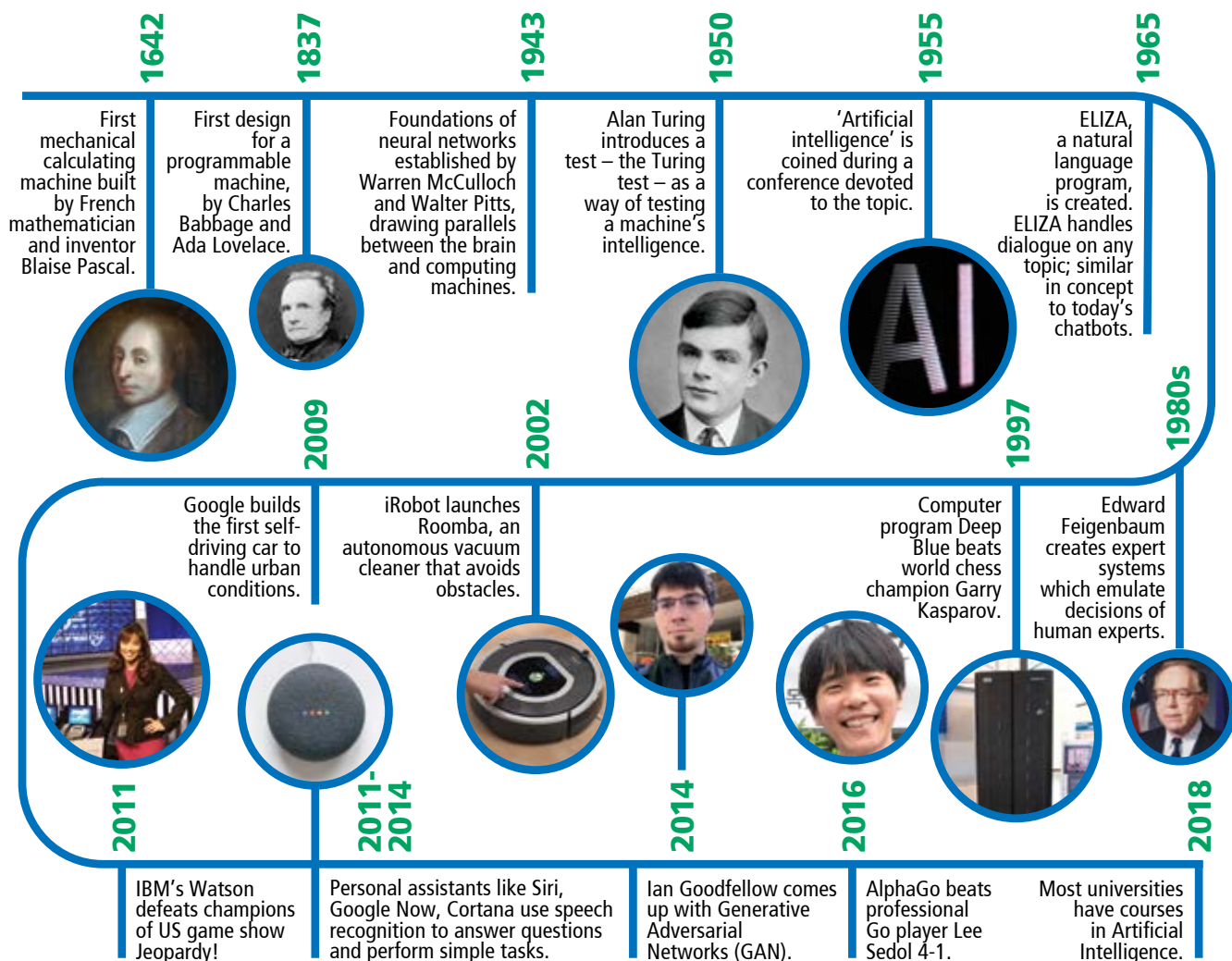
The latest surge of interest comes off the back of the power of deep learning, a type of biologically-inspired neural network that harnesses the huge amounts of data now available, and the massive computational power and speed of today's computers.

With enormous data sets, modern AI neural networks can often exceed human performance in many tasks, including pattern recognition and playing games like Go, previously very difficult for AI. Importantly, these systems can learn from experience, unlike GOFAL.

AI's ubiquity might now appear like it's not far off reaching human-level intelligence. But AI needs massive amounts of data to learn, unlike our brains, which can learn from a single experience.

Some researchers argue that for AI to advance further, more needs to be understood about the basic principles of how our brains function, and the kinds of biological shortcuts our brains take to complete tasks.

Queensland Brain Institute (30 January 2019). *History of Artificial Intelligence*. Retrieved from <http://qbi.uq.edu.au> on 9 July 2019.



This e-book is subject to the terms and conditions of a non-exclusive and non-transferable LICENCE AGREEMENT between THE SPINNEY PRESS and: Camberwell High School, Canterbury, library@camhigh.vic.edu.au

ARTIFICIAL INTELLIGENCE: AUSTRALIA'S ETHICS FRAMEWORK

EXECUTIVE SUMMARY FROM A DISCUSSION PAPER BY THE CSIRO AND DATA61

The ethics of artificial intelligence are of growing importance

Artificial intelligence (AI) is changing societies and economies around the world. Data61 analysis reveals that over the past few years, 14 countries and international organisations have announced AU\$86 billion for AI programs. Some of these technologies are powerful, which means they have considerable potential for both improved ethical outcomes as well as ethical risks. This report identifies key principles and measures that can be used to achieve the best possible results from AI, while keeping the wellbeing of Australians as the top priority.

Countries worldwide are developing solutions

Recent advances in AI-enabled technologies have prompted a wave of responses across the globe, as nations attempt to tackle emerging ethical issues (Figure 1).

Germany has delved into the ethics of automated vehicles, rolling out the most comprehensive government-led ethical guidance on their development available¹. New York has put in place an automated decisions task force, to review key systems used by government agencies for accountability and fairness².

The UK has a number of government advisory bodies, notably the Centre for Data Ethics and Innovation³. The European Union has explicitly highlighted ethical AI development as a source of competitive advantage⁴.

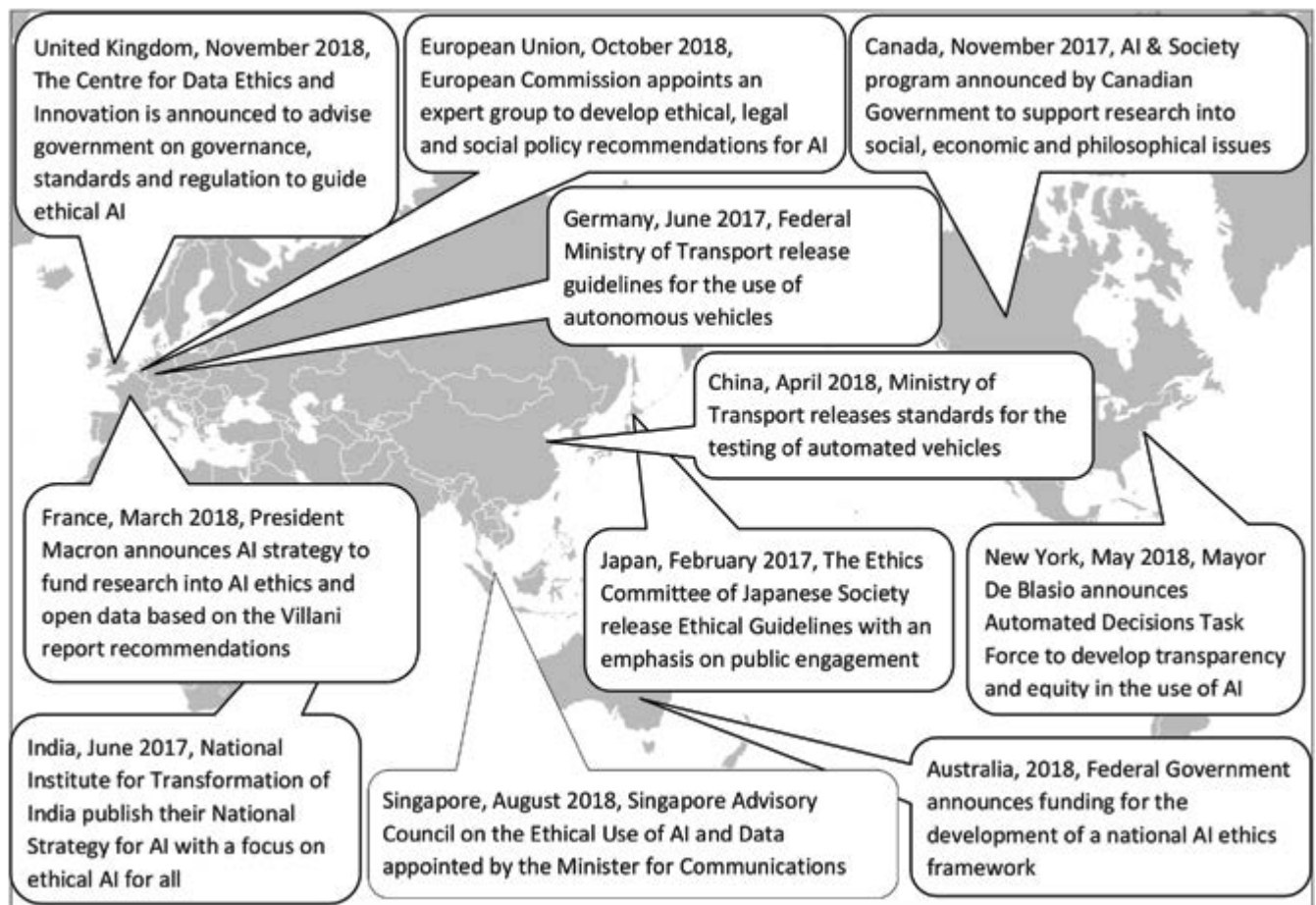
An approach based on case studies

This report examines key issues through exploring a series of case studies and trends that have prompted ethical debate in Australia and worldwide (see Figure 2).

Artificial intelligence (AI) holds enormous potential to improve society

While a “general AI” that replicates human intelligence is seen as an unlikely prospect in the coming few decades, there are numerous “narrow AI” technologies which are already incredibly sophisticated at handling specific tasks⁵. Medical AI technologies and autonomous vehicles are just a few high profile examples of AI that have potential to save lives and transform society.

FIGURE 1: MAP OF RECENT DEVELOPMENTS IN ARTIFICIAL INTELLIGENCE ETHICS WORLDWIDE



Data sources: Pan-Canadian AI Strategy⁵, Australian Federal Budget 2018-19⁶, German Ministry of Transport and Digital Infrastructure¹, National Institute for Transformation of India⁷, The Villani Report⁸, Reuters⁹, Japanese Society for Artificial Intelligence¹⁰, European Commission¹¹, UK Parliament¹², Singapore Government¹³, China’s State Council¹⁴, New York City Hall².

This e-book is subject to the terms and conditions of a non-exclusive and non-transferable LICENCE AGREEMENT between THE SPINNEY PRESS and: Camberwell High School, Canterbury, library@camhigh.vic.edu.au

FIGURE 2: TABLE OF KEY ISSUES EXAMINED IN CHAPTERS, CASE STUDIES AND RELEVANT PRINCIPLES

	Examples of case studies	Most relevant principles
Data governance and AI	<p>Identifying de-identified data</p> <p>In 2016, a dataset that included de-identified health information was uploaded to data.gov.au. It was expected that the data would be a useful tool for medical research and policy development. Unfortunately, it was discovered that in combination with other publicly available information, researchers were able to personally identify individuals from the data source. Quick action was taken to remove the dataset from data.gov.au.</p>	<p>Privacy protection</p> <p>Fairness</p>
Automated decisions	<p>Houston teachers fired by automated system</p> <p>An AI was used by the Houston school district to assess teacher performance and in some cases fire them. There was little transparency regarding the way that the AI was operating. The use of this AI was challenged in court by the teacher’s union, as the system was proprietary software and its inner workings were hidden. The case was settled and the district stopped using it¹⁵.</p>	<p>Fairness</p> <p>Transparency and explainability</p> <p>Contestability</p> <p>Accountability</p>
Predicting human behaviour	<p>The COMPAS sentencing tool</p> <p>COMPAS is a tool used in the US to give recommendations to judges about whether prospective parolee will re-offend. There is extensive debate over the accuracy of the system and whether it is fair to African Americans. Investigations by a non-profit outlet have indicated that incorrect predictions unfairly categorise black Americans as a higher risk. The system is proprietary software¹⁶⁻¹⁹.</p>	<p>Do no harm</p> <p>Regulatory and legal compliance</p> <p>Privacy protection</p> <p>Fairness</p> <p>Transparency and explainability</p>

Data sources: Office of the Australian Information Commissioner²⁰, US Senate Community Affairs Committee Secretariat¹⁵, ProPublica^{16,18,19}, Northpointe¹⁷.

The benefits come with risks

Automated decision systems can limit issues associated with human bias, but only if due care is focused on the data used by those systems and the ways they assess what is fair or safe. Automated vehicles could save thousands of lives by limiting accidents caused by human error, but as Germany’s Transport Ministry has highlighted in its ethics framework for AVs, they require regulation to ensure safety¹.

have spent centuries developing ethical concepts, culminating in the human-rights based framework used in international and Australian law. Australia is a party to seven core human rights agreements which have shaped our laws²¹.

An ethics framework for AI is not about rewriting these laws or ethical standards, it is about updating them to ensure that existing laws and ethical principles can be applied in the context of new AI technologies.

Existing ethics in context, not reinvented

Philosophers, academics, political leaders and ethicists

Data is at the core of AI

The recent advances in key AI capabilities such as deep

CORE PRINCIPLES FOR AI

- 1. Generates net benefits.** The AI system must generate benefits for people that are greater than the costs.
- 2. Do no harm.** Civilian AI systems must not be designed to harm or deceive people and should be implemented in ways that minimise any negative outcomes.
- 3. Regulatory and legal compliance.** The AI system must comply with all relevant international, Australian local, state/territory and federal government obligations, regulations and laws.
- 4. Privacy protection.** Any system, including AI systems, must ensure people’s private data is protected and kept confidential plus prevent data breaches which could cause reputational, psychological, financial, professional or other types of harm.
- 5. Fairness.** The development or use of the AI system must not result in unfair discrimination against individuals, communities or groups. This requires particular attention to ensure the “training data” is free from bias or characteristics which may cause the algorithm to behave unfairly.
- 6. Transparency and explainability.** People must be informed when an algorithm is being used that impacts them and they should be provided with information about what information the algorithm uses to make decisions.
- 7. Contestability.** When an algorithm impacts a person there must be an efficient process to allow that person to challenge the use or output of the algorithm.
- 8. Accountability.** People and organisations responsible for the creation and implementation of AI algorithms should be identifiable and accountable for the impacts of that algorithm, even if the impacts are unintended.

This e-book is subject to the terms and conditions of a non-exclusive and non-transferable LICENCE AGREEMENT between THE SPINNEY PRESS and: Camberwell High School, Canterbury, library@camhigh.vic.edu.au

learning have been made possible by vast troves of data. This data has to be collected and used, which means issues related to AI are closely intertwined with those that relate to privacy and data.

The nature of the data used also shapes the results of any decision or prediction made by an AI, opening the door to discrimination when inappropriate or inaccurate datasets are used. There are also key requirements of Australia's *Privacy Act* which will be difficult to navigate in the AI age²².

Predictions about people have added ethical layers

Around the world, AI is making all kinds of predictions about people, ranging from potential health issues through to the probability that they will end up re-appearing in court¹⁶. When it comes to medicine, this can provide enormous benefits for healthcare. When it comes to human behaviour, however, it's a challenging philosophical question with a wide range of viewpoints²³.

There are benefits, to be sure, but risks as well in creating self-fulfilling prophecies²⁴. The heart of big data is all about risk and probabilities, which humans struggle to accurately assess.

AI for a fairer go

Australia's colloquial motto is a "fair go" for all. Ensuring fairness across the many different groups in Australian society will be challenging, but this cuts right to the heart of ethical AI.

There are different ideas of what a "fair go" means. Algorithms can't necessarily treat every person exactly the same either; they should operate according to similar principles in similar situations. But while like goes with like, justice sometimes demands that different situations be treated differently. When developers need to codify fairness into AI algorithms, there are various challenges in managing often inevitable

trade-offs and sometimes there's no "right" choice because what is considered optimal may be disputed.

When the stakes are high, it's imperative to have a human decision-maker accountable for automated decisions – Australian laws already mandate it to a degree in some circumstances²⁵.

Transparency is key, but not a panacea

Transparency and AI is a complex issue. The ultimate goal of transparency measures are to achieve accountability, but the inner workings of some AI technologies defy easy explanation. Even in these cases, it is still possible to keep the developers and users of algorithms accountable²⁶.

An analogy can be drawn with people: an explanation of brain chemistry when making a decision doesn't necessarily help you understand how that decision was made – an explanation of that person's priorities is much more helpful.

There are also complex issues relating to commercial secrecy as well as the fact that making the inner workings of AI open to the public would leave them susceptible to being gamed²⁶.

Black boxes pose risks

On the other hand, AI "black boxes" in which the inner workings of an AI are shrouded in secrecy are not acceptable when public interest is at stake. Pathways forward involve a variety of measures for different situations, ranging from explainable AI technologies²⁷, testing, regulation that requires transparency in the key priorities and fairness measures used in an AI system, through to measures enabling external review and monitoring²⁶.

People should always be aware when a decision that affects them has been made by an AI, as difficulties with automated decisions by government departments have already been before Australian courts²⁸.

Justifying decisions

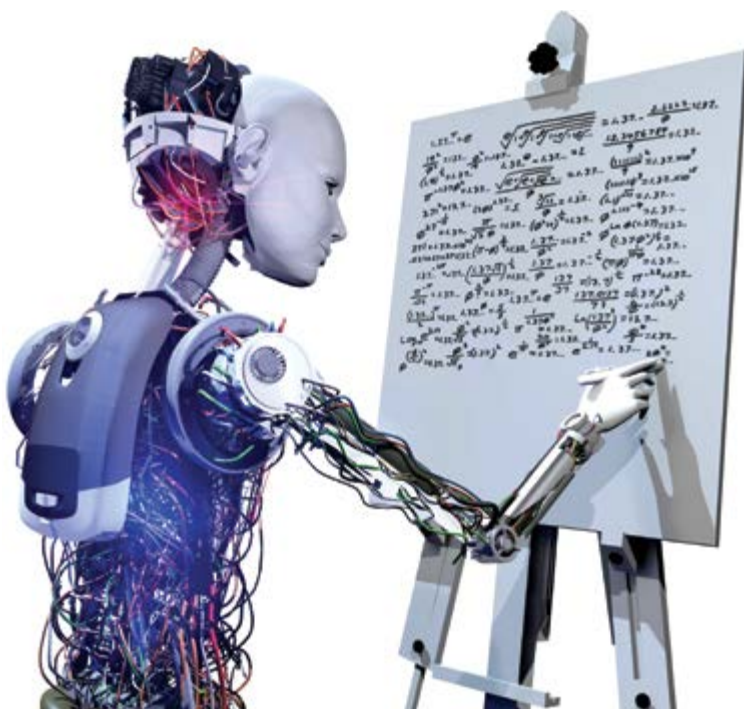
The transparency debate is one component feeding into another debate: justifiability. Can the designers of a machine justify what their AI is doing? How do we know what it is doing?

An independent, normative framework can serve to inform the development of AI, as well as justify or revise the decisions made by AI. This document is part of that conversation.

Privacy measures need to keep up with new AI capabilities

For decades, society has had rules about how fingerprints are collected and used. With new AI-enabled facial recognition, gait and iris scanning technologies, biometric information goes well beyond fingerprints in many respects²⁹.

Incidents like the Cambridge Analytica scandal demonstrate how far-reaching privacy breaches can be in the modern age, and AI technologies have the potential to impact this in significant ways. We may need to further explore what privacy means in a digital world.



Keeping the bigger picture in focus

Discussions on the ethics of autonomous vehicles tend to focus on issues like the “trolley problem” where the vehicle is given a choice of who to save in a life-or-death situation. Swerve to the right and hit an elderly person, stay straight and hit a child, or swerve to the left and kill the passengers?

These are important questions worth examining³⁰, but if widespread adoption of autonomous vehicles can improve safety and cut down on the hundreds of lives lost on Australian roads every year, then there is a risk that lives could be lost if relatively far-fetched scenarios dominate the discussion and delay testing and implementation. The values programmed into autonomous vehicles are important, though they need to be considered alongside potential costs of inaction.

AI will reduce the need for some skills and increase the demand for others

Disruption in the job market is a constant. However, AI may fuel the pace of change. There will be challenges in ensuring equality of opportunity and inclusiveness³¹.

An ethical approach to AI development requires helping people who are negatively impacted by automation transition their careers. This could involve training, reskilling and new career pathways. Improved information on risks and opportunities can help workers take proactive action. Incentives can be used to encourage the right type of training at the right times. Overall, acting early improves the chances of avoiding job loss or ongoing unemployment.

AI can help with intractable problems

Long-standing health and environmental issues are in need of novel solutions, and AI may be able to help.

Australia’s vast natural environment is in need of new tools to aid in its preservation, some of which are already being implemented³². People with serious



disabilities or health problems are able to participate more in society thanks to AI-enabled technologies³³.

International coordination is crucial

Developing standards for electrical and industrial products requires international coordination to make devices safe and functional across borders³⁴. Many AI technologies used in Australia won’t be made here. There are already plenty of off-the-shelf foreign AI products being used³⁵.

Regulations can induce foreign developers to work to Australian standards to a point, but there are limits. International coordination with partners overseas, including the International Standards Organisation (ISO), will be necessary to ensure AI products and software meet the required standards.

A TOOLKIT FOR ETHICAL AI

- 1. Impact assessments:** Auditable assessments of the potential direct and indirect impacts of AI, which address the potential negative impacts on individuals, communities and groups, along with mitigation procedures.
- 2. Internal or external review:** The use of specialised professionals or groups to review the AI and/or use of AI systems to ensure that they adhere to ethical principles and Australian policies and legislation.
- 3. Risk assessments:** The use of risk assessments to classify the level of risk associated with the development and/or use of AI.
- 4. Best practice guidelines:** The development of accessible cross industry best practice principles to help guide developers and AI users on gold standard practices.
- 5. Industry standards:** The provision of educational guides, training programs and potentially certification to help implement ethical standards in AI use and development
- 6. Collaboration:** Programs that promote and incentivise collaboration between industry and academia in the development of ‘ethical by design’ AI, along with demographic diversity in AI development.
- 7. Mechanisms for monitoring and improvement:** Regular monitoring of AI for accuracy, fairness and suitability for the task at hand. This should also involve consideration of whether the original goals of the algorithm are still relevant.
- 8. Recourse mechanisms:** Avenues for appeal when an automated decision or the use of an algorithm negatively affects a member of the public.
- 9. Consultation:** The use of public or specialist consultation to give the opportunity for the ethical issues of an AI to be discussed by key stakeholders.

Implementing ethical AI

AI is a broad set of technologies with a range of legal and ethical implications. There is no one-size-fits all solution to these emerging issues.

There are, however, tools which can be used to assess risk and ensure compliance and oversight. The most appropriate tools can be selected for each individual circumstance.

Best practice based on ethical principles

The development of best practice guidelines can help industry and society achieve better outcomes. This requires the identification of values, ethical principles and concepts that can serve as their basis.

ABOUT THIS REPORT

This report covers civilian applications of AI. Military applications are out of scope. This report also acknowledges research into AI ethics occurring as part of a project by the Australian Human Rights Commission³⁶, as well as work being undertaken by the recently established Gradient Institute.

This work complements research being conducted by the Australian Council of Learned Academies (ACOLA) and builds upon the Robotics Roadmap for Australia by the Australian Centre for Robotic Vision.

From a research perspective, this framework sits alongside existing standards, such as the National Health and Medical Research Council (NHMRC) Australian Code for the Responsible Conduct of Research and the NHMRC's National Statement on Ethical Conduct in Human Research.

REFERENCES

1. Ethics Commission Federal Ministry of Transport and Digital Infrastructure. 2017. *Automated and connected driving*. German Government. Germany.
2. New York City Hall. 2018. *Mayor de Blasio announces first-in-nation Task Force to examine automated decision systems used by the city*. Mayor's Office.
3. House of Lords Select Committee on Artificial Intelligence. 2018. *AI in the UK: Ready, willing and able?* UK Parliament. United Kingdom.
4. European Commission. 2018. *Artificial Intelligence for Europe*.
5. Canadian Institute for Advanced Research. 2017. *Pan-Canadian artificial intelligence strategy*.
6. Australian Government. 2018. *2018-2019 Budget Overview*. Canberra.
7. National Institute for Transforming India. 2018. *National strategy for artificial intelligence*.
8. Villani C. 2018. *For a Meaningful Artificial Intelligence*.
9. Rosemain M and Rose M. *France to spend \$1.8 billion on AI to compete with U.S., China*. Reuters. 29 March 2018.
10. Japanese Society for Artificial Intelligence. 2017. *The Japanese society for artificial intelligence ethical guidelines*. Japan.
11. European Commission. 2018. *High-level expert group on artificial intelligence*.
12. UK Parliament. 2018. *World first Centre for Data Ethics and Innovation: Government statement*.
13. Infocomm Media Development Authority. 2018. *Composition of the Advisory Council on the Ethical Use of Artificial Intelligence and Data*. Minister for Communications and Information. Singapore.
14. The State Council: The People's Republic of China. 2018. *Guidelines to ensure safe self-driving vehicle tests*. *The State Council: The People's Republic of China*.
15. Senate Community Affairs Committee Secretariat. 2017. *Design, scope, cost-benefit analysis, contracts awarded and implementation associated with the Better Management of the Social Welfare System initiative*. Parliament of Australia.
16. Angwin J, Larson J, Mattu S et al. 2016. Machine bias risk assessments in criminal sentencing. *ProPublica*.
17. Dieterich W, Mendoza C, Brennan T. 2016. COMPAS risk scales: Demonstrating accuracy and predictive parity. Northpointe Suite.
18. Angwin J, Larson J. 2016. ProPublica responds to company's critique of machine bias story. *ProPublica*.
19. Angwin J, Larson J. 2016. Bias in criminal risk scores is mathematically inevitable, researchers say. *ProPublica*.
20. Office of the Australian Information Commissioner. 2018. *Publication of Medicare Benefits Schedule and Pharmaceutical Benefits Schedule data: Commissioner initiated investigation report*. Australian Government.
21. Australian Government. *International human rights system*. Attorney-General's Department.
22. Australian Government. 1988. *Privacy Act 1988, No. 119, 1988 as amended*. Australian Government. Canberra.
23. Corbett-Davies S, Pierson E, Feller A et al. 2016. A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. *The Washington Post*.
24. Moses L B, Chan J. 2016. Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and Society*, 28(7): 806-822.
25. Australian Government. 1999. *Social Security (Administration) Act 1999: No 191, 1999*. Federal Register of Legislation.
26. Reisman D, Schultz J, Crawford K et al. 2018. *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*.
27. Kleinman Z. *IBM launches tool aimed at detecting AI bias*.
28. Khadem N. Tax office computer says yes, Federal Court says no. *ABC*. 8 October 2018.
29. Human Rights Law Centre. 2018. *The dangers of unregulated biometric use: Submission to the Inquiry into the Identity-matching Services Bill 2018 and the Australian Passports Amendment (Identity-matching Services) Bill 2018*. Human Rights Law Centre. Australia.
30. Awad E, Dsouza S, Kim R et al. 2018. The Moral Machine experiment. *Nature*, 563(7729): 59-64.
31. Berg A, Buffie E, Zanna L-F. 2018. *Should we fear the robot revolution? (The correct answer is yes)*. International Monetary Fund.
32. CSIRO. 2015. *Robots to ResQu our rainforests*. CSIRO.
33. Ray S. 2018. Data guru living with ALS modernizes industries by typing with his eyes. *Microsoft News*.
34. International Electrotechnical Commission. *Functional Safety and IEC 61508*.
35. Redrup Y. 2018. Google to make AI accessible to all businesses with Cloud AutoML. *Australian Financial Review*.
36. Australian Human Rights Commission. 2018. *Human rights and technology issues paper*. Sydney.

© CSIRO (Commonwealth Scientific and Industrial Research Organisation) 2019.

Dawson, D, Schleiger, E, Horton, J, McLaughlin, J, Robinson, C, Quezada, G, Scowcroft, J, and Hajkowicz, S. *Artificial Intelligence: Australia's Ethics Framework*, Executive summary, pp. 4-9, Data61/CSIRO, Australia. Retrieved from <http://consult.industry.gov.au> on 9 July 2019.

Will we ever agree to just one set of rules on the ethical development of AI?

EVERYONE HAS THEIR OWN IDEA ON THE ETHICAL USE OF AI, BUT CAN WE GET A GLOBAL CONSENSUS? BY **MICHAEL GUIHOT**

Australia is among 42 countries that last week signed up to a new set of policy guidelines for the development of artificial intelligence (AI) systems. Yet Australia has its own draft guidelines for ethics in AI out for public consultation, and a number of other countries and industry bodies have developed their own AI guidelines.

So why do we need so many guidelines, and are any of them enforceable?

THE NEW PRINCIPLES

The latest set of policy guidelines is the Recommendation on Artificial Intelligence from the Organisation for Economic Co-operation and Development (OECD).

It promotes five principles for the responsible development of trustworthy AI. It also includes five complementary strategies for developing national policy and international cooperation.

Given this comes from the OECD, it treads the line between promoting economic improvement and innovation and fostering fundamental values and trust in the development of AI.

The five AI principles encourage:

1. Inclusive growth, sustainable development and wellbeing
2. Human-centred values and fairness
3. Transparency and explainability

4. Robustness, security and safety
5. Accountability.

These recommendations are broad and do not carry the force of laws or even rules. Instead they seek to encourage member countries to incorporate these values or ethics in the development of AI.

BUT WHAT DO WE MEAN BY AI?

It is hard to make specific recommendations in relation to AI. That is partly because AI is not one thing with a single application that poses singular risks or threats.

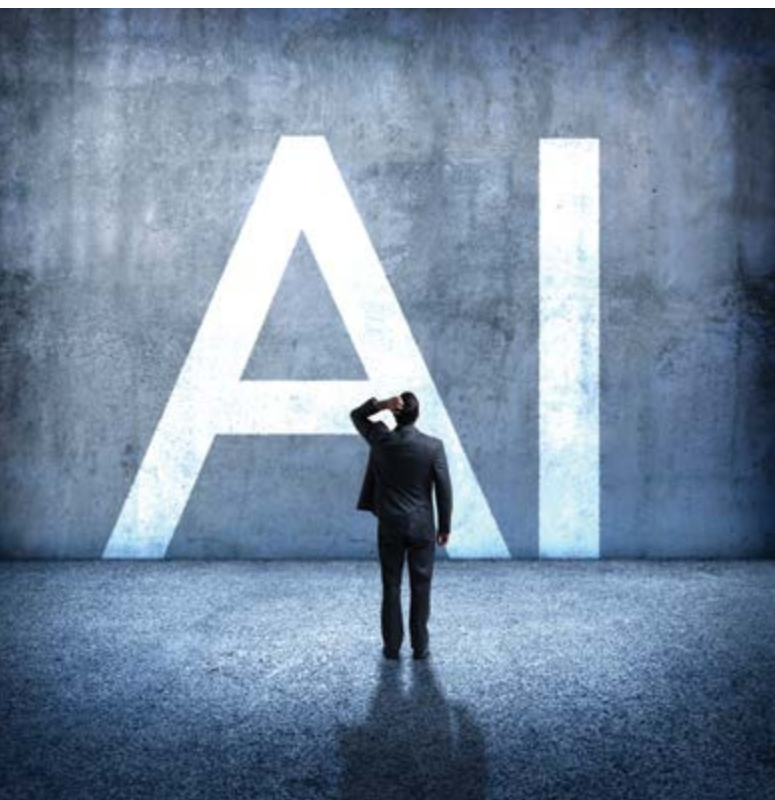
Instead, AI has become a blanket term to refer to a vast number of different systems. Each is typically designed to collect and process data using computing technology, adapt to change, and act rationally to achieve its objectives, ultimately without human intervention.

Those objectives could be as different as translating language, identifying faces, or even playing chess.

The type of AI that is exceptionally good at completing these objectives is often referred to as narrow AI. A good example is a chess-playing AI. This is specifically designed to play chess – and is extremely good at it – but completely useless at other tasks.

On the other hand is general AI. This is AI that it is said will replace human intelligence in most if not all tasks. This is still a long way off but remains the ultim-





ate goal of some AI developers.

Yet it is this idea of general AI that drives many of the fears and misconceptions that surround AI.

MANY MANY GUIDELINES

Responding to these fears and a number of very real problems with narrow AI, the OECD recommendations are the latest of a number of projects and guidelines from governments and other bodies around the world that seek to instil an ethical approach to developing AI.

Common themes are emerging in the various guidelines, such as the need for AI that considers human rights, security, safety, transparency, trustworthiness and accountability, so we may yet be on the way to some global consensus.

These include initiatives by the Institute of Electrical and Electronics Engineers, the French data protection authority, the Hong Kong Office of the Privacy Commissioner and the European Commission.

The Australian government funded CSIRO's Data61 to develop an AI ethics framework, which is now open for public feedback, and the Australian Council of Learned Academies is yet to publish its report on the future of AI in Australia.

The Australian Human Rights Commission, together with the World Economic Forum, is also reviewing and reporting on the impact of AI on human rights.

The aim of these initiatives is to encourage or to nudge ethical development of AI. But this presupposes unethical behaviour. What is the mischief in AI?

UNETHICAL AI

One study identified three broad potential malicious uses of AI. These target:

- Digital security (for example, through cyber-attacks)
- Physical security (for example, attacks using drones or hacking)
- Political security (for example, if AI is used for mass surveillance, persuasion and deception).

One area of concern is evolving in China, where several regions are developing a social credit system linked to mass surveillance using AI technologies.

The system can identify a person breaching social norms (such as jaywalking, consorting with criminals, or misusing social media) and debit social credit points from the individual.

When a credit score is reduced, that person's freedoms (such as the freedom to travel or borrow money) are restricted. While this is not yet a nationwide system, reports indicate this could be the ultimate aim.

Added to these deliberate misuses of AI are several unintentional side effects of poorly constructed or implemented narrow AI. These include bias and discrimination and the erosion of trust.

Building consensus on AI

Societies differ on what is ethical. Even people within societies differ on what they regard as ethical behaviour. So how can there ever be a global consensus on the ethical development of AI?

Given the very broad scope of AI development, any policies in relation to ethical AI cannot yet be more specific until we can identify shared norms of ethical behaviour that might form the basis of some agreed global rules.

By developing and expressing the values, rights and norms that we consider to be important now in the form of the reports and guidelines outlined above, we are working toward building trust among nations.

Common themes are emerging in the various guidelines, such as the need for AI that considers human rights, security, safety, transparency, trustworthiness and accountability, so we may yet be on the way to some global consensus.

DISCLOSURE STATEMENT

Michael Guihot does not work for, consult, own shares in or receive funding from any company or organisation that would benefit from this article, and has disclosed no relevant affiliations beyond his academic appointment.

Michael Guihot is Senior Lecturer in Law, Queensland University of Technology.

THE CONVERSATION

Guihot, M (29 May 2019). *Will we ever agree to just one set of rules on the ethical development of artificial intelligence?* Retrieved from <http://theconversation.com> on 9 July 2019.

HOW DO WE MAKE ARTIFICIAL INTELLIGENCE MORE HUMANE?

How do we build trust in the development and use of exciting new technologies, while also addressing the possible threats to universal human rights?

By **Edward Santow** and **Nicholas Davis**

We've all had something like it happen: one minute you're searching for a present suitable for a two-year-old; the next, ads for nappies and prams are on every site you visit. It's unsettling. No one feels comfortable about bots following us surreptitiously as we roam around the web, when companies use what they learn from our online behaviour to promote products and services in creepy ways.

But could concerns around privacy and informed consent – though undeniably important – be distracting us from what we should be really worried about?

The exploitation of personal information for marketing purposes is a real problem. But the more serious risk is that our personal information can be used against us – not just to advertise a product we don't want, but to discriminate against us on the basis of our age, race, gender or some other characteristic we can't control.

PRECISION PREJUDICE

For example, if you have darker skin, facial-recognition technology is dramatically less accurate than if you have a light complexion. As this technology is progressively rolled out across law enforcement, in border security and even in delivering financial services, the risk that you'll be unfairly disadvantaged increases depending on your ethnicity.

Similarly, there are examples of artificial intelligence (AI) operating to prevent women or older people seeing certain online employment opportunities.

Not only does this violate the human rights of anyone negatively affected, but it also undermines community trust in AI more broadly. A collapse in community trust in AI would be disastrous, because AI has the potential to be an enormous boon – not just for national economies, but also in making communities more inclusive.

For every instance of AI causing harm, there's also an uplifting counter-example. This could be anything from AI-powered smartphone applications allowing blind people to “see” the world around them, to huge strides in precision medicine.

The challenge, therefore, is to build enduring trust in the development and use of a tremendously exciting set of technologies, so that citizens and organisations around the world can take advantage of the opportunities while addressing the threats to universal human rights.

This might sound eminently sensible to you. Unfortunately, this challenge is made harder by a damaging but pervasive myth.



RIGHTING THE WRONGS

Take Australia as an example. A common counterpoint to the idea that common sense norms and rules should apply to the development or implementation of AI relates to the fact that other countries are less likely to do the same.

AI, is developing at breakneck speed ... in almost every country around the world, laws are slow to adapt.

In which case, the argument goes, if Australia is to compete globally in developing AI products, Australian researchers and companies must not be fettered by human rights concerns, because other countries certainly aren't.

China, for example, is investing heavily in AI technology such as facial recognition to support its “social credit score” system, which involves conducting precise and determinative surveillance of its citizens. In the context of a global AI arms race, it is argued, Australia can't compete with one arm tied behind its back.

This argument is dangerous and misguided. Australia's liberal democratic values are one of its core strengths. The Australian Human Rights Commission's consultation on human rights and technology has shown that,



as Australians learn more about AI, there's a growing demand that AI only be used in ways that respect their human rights.

Consumers in liberal democracies want the benefits of AI, through self-driving cars, better healthcare and super-powerful computers. However, they won't accept a trade-off that involves mass surveillance, the exclusion of entire groups and a rise in discrimination.

This suggests that embedding human-rights protection in AI as it's developed isn't just morally right – it's also smart. If Australia can become known for developing AI that gets the balance right, it can gain a competitive advantage.

After all, consumers in liberal democracies want the benefits of AI, through self-driving cars, better health-care and super-powerful computers. However, they won't accept a trade-off that involves mass surveillance, the exclusion of entire groups and a rise in discrimination.

So, what's the solution? We know that technology, and especially AI, is developing at breakneck speed. We also know that, in almost every country around the world, laws are slow to adapt.

This puts greater pressure on institutions in countries such as Australia to smooth AI's rough edges in ways that allow us to harness the opportunities without allowing vulnerable members of our community to be crushed.

AI LEADERSHIP

Luckily, there is a way forward. And it might just be in Australia that real progress is made.

Several influential voices have already called for an Australian organisation to lead on AI. The World Economic Forum and the Australian Human Rights Commission have formed a partnership to consider

this idea. These two bodies have invited leading decision-makers in government, industry and academia to meet at the University of Technology Sydney (UTS) to consider how we will tackle this AI leadership challenge.

Based on the consultation we have conducted to date, some of the key issues that should be considered include the following.

First, we should clearly articulate the values that should underpin AI. In Australia, these should be quintessentially Australian values such as equality or the fair go.

Second, there has been some support among stakeholders for a specialised organisation – either a new or existing one – to take a central role in assessing technologies and formulating laws, guidelines, accountability and capacity-building strategies in AI. This should be a national organisation with close connections with all stakeholders.

Third, this organisation should work closely with industry, government and the community to support the development of AI technologies that respect human rights.

The World Economic Forum and the Australian Human Rights Commission are consulting on these issues right now and have produced a white paper inviting comments, focused on the Australian context. But this is an issue facing all countries around the world, and we welcome your input in this process.

Edward Santow is a Human Rights Commissioner and leads the Australian Human Rights Commission's work on technology and human rights.

Nicholas Davis is Head of Society and Innovation, Member of Executive Committee, World Economic Forum.

© Australian Human Rights Commission.

Santow, E, and Davis, N (5 March 2019).
How do we make artificial intelligence more humane?
Retrieved from www.humanrights.gov.au on 9 July 2019.

PROTECTING HUMAN RIGHTS IN THE CONTEXT OF AI

A CHAPTER EXTRACT FROM A WHITE PAPER PRODUCED BY THE AUSTRALIAN HUMAN RIGHTS COMMISSION AND THE WORLD ECONOMIC FORUM

International and Australian human rights law requires that individuals be treated without discrimination.¹⁷ Governments must uphold human rights, while businesses have a responsibility to respect human rights in all their operations.¹⁸

The most effective way of complying with these obligations in the context of AI is to ensure that it is designed and used responsibly, by protecting privacy, fairness, equality and other human rights.

This is, of course, not solely an Australian challenge. Public and private sector organisations globally are exploring ways to understand and manage the impact of bias in AI and ML (machine learning). The interconnected nature of the global economy, combined with the fact that such explorations in innovative governance are at an early stage in all jurisdictions, means that Australian institutions have an opportunity to lead in developing new structures, policies and relationships that can help address these important issues on behalf of all Australians.

Accordingly, a responsible innovation framework could accommodate imperatives that sometimes sit in tension, by anticipating and addressing the potential harms of AI, so that it can be deployed in a way that is safe and beneficial for Australia.

There are three primary reasons for Australia to be concerned about bias and discrimination in AI systems:

1. Automated decision-making systems will be applied more often by both the private and public sectors

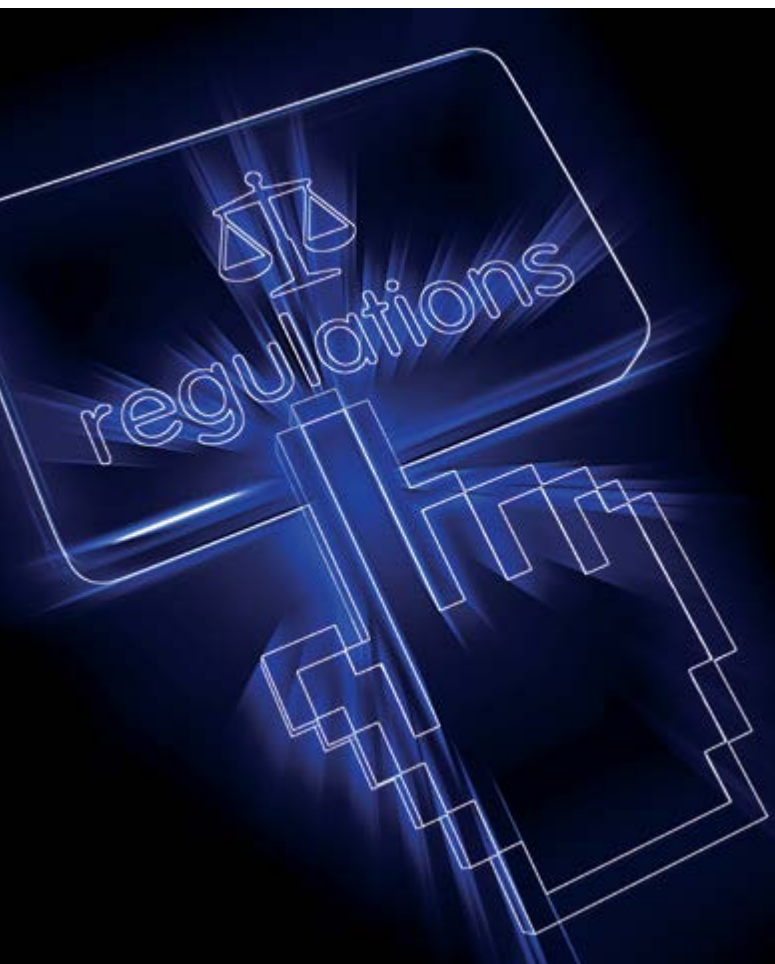
Case study: Artificial intelligence and the risk of discrimination

Bias and discrimination in technology have entered the public consciousness along with our increasing reliance on and understanding of AI and ML. AI systems can discriminate and operate unfairly for many reasons.

For example:

- AI is designed by human beings who possess inherent biases and is often trained with data that reflects the imperfect world that we live in.¹⁹
- Training AI systems with data that is not representative, or using data that reflects bias or prejudice (for example, sexism or racism), can lead to an AI-supported decision that is unfair, unjust, unlawful or otherwise wrong.²⁰
- AI's algorithms can include discriminatory variables (for example, including a variable for private school attendance in a loan application algorithm) that results in further discrimination.²¹
- Where users do not understand AI's limitations, especially if they assume AI's predictions to be more accurate and precise (and thus more authoritative) than those made by people, this can result in unfairness.²²
- AI can be deployed in an inappropriate context (for example, deploying a model in a different cultural context from that in which it was originally trained).²³
- Personal data is the 'fuel' for AI.²⁴ It can be at risk when deployed in ML models, as hackers can often threaten individual privacy by reverse-engineering algorithms, which could allow access to the personal data the algorithm is trained on.²⁵





and at a greater scale across a wide variety of essential services, from decisions in health care to financial services. Discrimination in these decisions is both more likely and of greater consequence for groups that are already vulnerable.²⁶

2. It is difficult to know the decision-making process adopted in an AI system, because ML tends to involve opaque proprietary algorithms.²⁷ Without understanding this process, it is hard to discern whether, when or how such systems are discriminating against a group or individual. This fundamentally challenges the concept of procedural fairness in administrative decision-making.
3. Hasty implementation of AI puts at risk its benefits by undermining public trust in new technologies. Public trust in Australian businesses, the government, media and civil society has fallen rapidly in the last decade to record lows.²⁸ If this trust is further eroded by the emergence of

widespread discrimination through the deployment of AI systems, it may slow adoption in ways that prevent Australians from harnessing the many positive impacts of AI and ML.

The first set of consultation questions (*see box below*) focuses on understanding your sense of the challenge itself and the general approach that you feel the government should take in this area.

ENDNOTES

17. Australian Human Rights Commission, *Legislation*, www.humanrights.gov.au/our-work/legal/legislation.
18. Australian Human Rights Commission, *Business and Human Rights*, www.humanrights.gov.au/employers/business-and-human-rights.
19. Microsoft, *The Future Computed: Artificial Intelligence and its Role in Society* (Microsoft, 2018) 58.
20. *Ibid* 58-59.
21. Nicholas Davis, World Economic Forum, Ch. 7: The Future Relationship Between Technology and Inequality, in *How Unequal? Insights on Inequality* (Committee for Economic Development of Australia, April 2018).
22. Microsoft, *The Future Computed: Artificial Intelligence and its Role in Society* (Microsoft, 2018) 59.
23. Nicholas Davis, World Economic Forum, Ch. 7: The Future Relationship Between Technology and Inequality, in *How Unequal? Insights on Inequality* (Committee for Economic Development of Australia, April 2018).
24. Christian Ehl, 'Data – the Fuel for Artificial Intelligence' on *Medium* (14 January 2018), <https://medium.com/@cehl/data-the-fuel-for-artificial-intelligence-ed90bf141372>; see also Bernard Marr, 'Why AI Would Be Nothing Without Big Data,' *Forbes* (online), 9 June 2017, www.forbes.com/sites/bernardmarr/2017/06/09/why-ai-would-be-nothing-without-big-data/#6fe8a7a94f6d.
25. Justin Sherman, 'AI Innovation: Security and Privacy Challenges' on *Medium* (22 January 2018), www.medium.com/swlh/ai-innovation-security-and-privacy-challenges-84c0200b1bae.
26. Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press, 2018).
27. World Economic Forum, *How to Prevent Discriminatory Outcomes in Machine Learning* (2018), www3.weforum.org/docs/wef_40065_white_paper_how_to_prevent_discriminatory_outcomes_in_machine_learning.pdf.
28. Edelman, *Edelman Trust Barometer 2018*, www.slideshare.net/.

Australian Human Rights Commission and World Economic Forum (White paper January 2019). *Artificial Intelligence: governance and leadership*, Chapter 3, pp. 9-10. Retrieved from <http://tech.humanrights.gov.au> on 9 July 2019.

CONSULTATION QUESTIONS

1. What should be the main goals of government regulation in the area of artificial intelligence?
2. Considering how artificial intelligence is currently regulated and influenced in Australia:
 - (a) *What existing bodies play an important role in this area?*
 - (b) *What are the gaps in the current regulatory system?*

RISE OF THE MACHINES

ON LINE OPINION ARTICLE BY **PETER MCMAHON** AND **SEBASTIAN TREW**

One of the most interesting and potentially important trends right now is the accelerating development of digital technology in all its various forms. This includes everything from new games to artificial intelligence (AI) to robots. This trend will have a growing impact on our daily lives, and possibly present serious new threats as well.

Driven by Moore's Law (that says that computer processing power doubles every 18 to 24 months) and fast-growing investment from the military, corporations and financial concerns who see ever faster and more capable digital systems as the key to competitiveness, advances in such technologies are about to achieve break-out status. It is probably already too late to put this genie back in the bottle, and we really have little idea what it all means for humanity and the way we will live in coming years.

The promise in terms of more effective military power, cheaper industrial production, more pervasive communications and even domestic robot helpers is clear, but already luminaries like Stephen Hawking, Elon Musk and Bill Gates have warned that there are potential dangers ahead. This piece considers some of the background to recent developments in digital technology and their significance.

For over a century the idea of robots taking over the human race has been of interest to various authors, film-makers and fans of science fiction. The notion that one day robots that become fully autonomous and even conscious might supersede human capacities in various ways so as to eventually take over the world and assert their artificial dominance was thought of as pure science fiction and beyond the bounds of possibility.

However, according to current research and expert opinion, there is a high probability that within a few decades artificial intelligences (AIs), most likely in the form of robots, will achieve genuine sentient thought. By whatever means this might happen, either by deliberate action or accident, this genesis of machine consciousness could be expected to have an extensive and significant impact on human society and culture.

Such a fundamental change has been called a 'singularity', the point where a graph of the exponential growth of information processing power goes vertical. It is a term first coined in 1958 by John Von Neumann

and Stan Ulam. They suggested that such a change would mean an elemental change in the nature of human civilisation and technology, to the point where the future becomes totally unpredictable. A concept taken up by science fiction authors (such as Vernor Vinge and William Gibson) and the futurist Ray Kurzweil, the singularity is perceived as a defining moment that could well occur with decades, not centuries.

In the last decade there has been a series of highly pertinent publications focusing on artificial intelligence, robots and their impact upon human culture (for example: Ashrafian 2015; Searle 2016; Geist 2015; Smith 2009; Singer and Sagan 2009). In addition, since 2015, there has been a dramatic increase in the

It is probably already too late to put this genie back in the bottle, and we really have little idea what it all means for humanity and the way we will live in coming years.





They argue that robots may have the capacity to have feelings similar to those of humans, an idea based in the notion that the human brain is just a ‘very complex machine’ that can be simulated, programmed and copied.

This idea of machines with feelings has been around for a while in science fiction movies, at least as far back as the computer HAL in *2001: A Space Odyssey* (1968) and more recently in the Japanese anime movie *Ghost in the Shell* (1996) and last year’s *Ex Machina*. It is an interesting idea because it focuses on the central subject of what these feelings actually are, in both creations like robots, but also in ourselves.

The biggest concern of immediate relevance is the use of robots as weapons. Modern militaries are all moving towards autonomous robot soldiers, aircraft and ships. They can operate more quickly and agilely, they have no fear, their destruction does not evoke popular revulsion, and they are cheaper and more capable all the time. But they can be hacked (as the Iranians did with US drones recently) and reprogrammed. So such powerful weapons with their own emotions is a truly scary idea.

Ultimately, relationships between robots and humans may be based in the already existing debate about human diversity. We humans have always struggled with variation in our own species, and as a consequence do not always afford full rights to even our own kind. For example, historically indigenous peoples, those with disabilities of various kinds, and individuals identifying as LGBTQI have had their rights violated constantly and consistently. When we struggle to include all humans in decent society, how would we go if completely different forms of intelligence start staking claims to basic rights?

These matters are but some of the thorny issues implicit in the rise of the machines. Whether they constitute a real threat, become great helpers or develop to become our equals is uncertain right now. Hopefully we have as a species the wisdom to deal with it all with a maturity we have sometimes lacked in the past.

Dr Peter McMahon has worked in a number of jobs including in politics at local, state and federal level. He has also taught Australian studies, politics and political economy at university level, and until recently he taught sustainable development at Murdoch University. He has been published in various newspapers, journals and magazines in Australia and has written a short history of economic development and sustainability in Western Australia. His book *Global Control: Information Technology and Globalisation* was published in the UK in 2002. He is now an independent researcher and writer on issues related to global change.

Sebastian Trew is a Research Officer at the Australian Catholic University, he holds a Master’s degree in Human Rights. In his thesis, *Intelligent Robots Have Human Rights in Sight*, he considered giving intelligent robots a form of rights for the future safety of humanity. Sebastian is a PhD candidate at the Australian Catholic University. His research explores family relationships and autism spectrum condition.

McMahon, P, and Trew, S (30 August 2016). *Rise of the machines*. Retrieved from www.onlineopinion.com.au on 9 July 2019.

According to current research and expert opinion, there is a high probability that within a few decades artificial intelligences (AIs), most likely in the form of robots, will achieve genuine sentient thought.

publication of academic papers, news articles, websites, blogs, and other varying formats relating to artificial intelligence in the form of robots and their impending impact on our world.

The commentary on this prospect remains mixed, particularly on the topic of whether robots pose a threat. In particular, opponents of the idea of the ‘humanisation’ of AI argue that any concerns about robots living amongst humans are ‘grossly exaggerated’ (for example: Geist 2015). Certainly, the immediate future of robot-human interaction looks rather mundane since robots operating as shop-bots and cleaning-bots will be largely reliant on ongoing human operational control.

However, much more sophisticated robots will appear, and with them questions concerning their intelligence and even their basic existential status. Here opinions vary. Wesley Smith for instance, argues that “robots will never be people and should never have rights.” He argues that robots would only be computers with very sophisticated software and “would be no more capable of being harmed, as distinguished from damaged – than the toaster”.

Smith disputes the view put by others that robots may develop actual feelings, as argued in Peter Singer and Agata Sagan’s 2009 book *When Robots have Feelings*. Singer and Sagan are worried not so much about robots harming humans but rather humans harming robots.

FIVE REASONS WHY ROBOTS WON'T TAKE OVER THE WORLD

In the next two decades, robots will not be human-like, even if they might look like humans. Instead they will remain sophisticated machines, according to Danish researchers **Norbert Krüger** and **Ole Dolriis**

Scientists are known for making dramatic predictions about the future – and sinister robots are once again in the spotlight now that artificial intelligence has become a marketing tool for all sorts of different brands.

At the end of World War Two, it was stated that flying cars were just around the corner and that all energy problems would be solved by fusion energy by the end of the 20th century. But decades on, we don't seem much closer to either of those predictions coming true.

So what's with all this talk – fuelled by the likes of space baron, Elon Musk – about robots taking over the world?

Robots are a reality today in industry and they will appear in public spaces in more complex shapes than robot vacuum cleaners. But in the next two decades, robots will not be human-like, even if they might look like humans. Instead they will remain sophisticated machines.

Pessimists predict that robots will jeopardise jobs across the globe, and not only in industrial production. They claim robot journalists, robot doctors and robot lawyers will replace human experts. And, as a consequence of a melting down middle class, there will be mass poverty and political instability.

Optimists predict a new paradise where all the tedious problems of human relationships can be overcome by having a perfect life with easily replaceable robot partners, which will fulfil our basic needs as well as our deepest longings. And “work” will become an ancient concept.

The pessimists, however, can relax and the optimists need to cool their boots. As experts in the field of robotics, we believe that robots will be much more visible in the future, but – at least over the next two decades – they will be clearly recognisable as machines.

This is because there is still a long way to go before robots will be able to match a number of fundamental human skills. Here are five reasons why robots aren't about to take over the world.

1. Human-like hands

Scientists are far from replicating the complexity of human hands. The hands of robots that are used today in real applications are clumsy. The more sophisticated hands developed in labs are not robust enough and lack the dexterity of human hands.

2. Tactile perception

There is no technical match for the magnificent human and animal skin that encompasses a variety of tactile sensors. This perception is required for complex manipulation. Also, the software that processes the input from the sensors in robots is nowhere near as sophisticated as the human brain when it comes to interpretation and reaction to the messages received from the tactile sensors.

3. Control of manipulation

Even if we had artificial hands comparable to human hands and sophisticated artificial skin, we would still need to be able to design a way to control them to manipulate objects in a human-like way. Human children take years to do this and the learning mechanisms are not understood.

4. Human and robot interaction

The interaction between humans is built on well-functioning speech and object recognition systems, as well as other sensors such as smell and taste and tactile sensing. While there has been significant progress in speech and object recognition, today's systems can still only be used in rather controlled environments when a high degree of performance is required.

5. Human reason

Not all of what is technically possible needs to be built. Human reason could decide not to fully develop such robots, because of their potential harm to society.



If, in many decades from now, the technical problems mentioned above are overcome so that complex human-like robots could be built, regulations could still prevent misuse.

SMOOTH OUT THE EDGES

In our research project, SMOOTH, we design robots that we hope will operate in elderly care institutions by 2022. These robots will be used to solve repetitive tasks involving human and robot interaction, such as transporting laundry and waste, offering water to people or guiding them to the breakfast table.

It was necessary to simplify the robots as well as to carefully select the tasks they perform to ensure that they can be commercially viable products within four years.

Our approach wasn't to solve the first three problems of human-like hands, tactile perception and control of manipulation, but to avoid those robotic roadblocks.

To address the fourth problem of human and robot interaction, we chose repetitive tasks to reduce complexity, since the expected interactions are – to a certain degree – predictable.

Robots are a reality today in industry and they will appear in public spaces in more complex shapes than robot vacuum cleaners. But in the next two decades, robots will not be human-like, even if they might look

like humans. Instead they will remain sophisticated machines.

So you can stand down from any fear of a robot uprising in the near future.

DISCLOSURE STATEMENT

Norbert Krüger has received funding from the EU and various Danish national funds, in particular the Innovation Fund Denmark. Ole Dolriis does not work for, consult, own shares in or receive funding from any company or organisation that would benefit from this article, and has disclosed no relevant affiliations beyond his academic appointment.

Norbert Krüger is Professor, University of Southern Denmark.

Ole Dolriis is Robotics lecturer, University of Southern Denmark.

THE CONVERSATION

Kruger, N, and Dolriis, O (18 April 2018). *Five reasons why robots won't take over the world*. Retrieved from <http://theconversation.com> on 9 July 2019.

ADVANCES IN AI AND ROBOTICS

AI PATENTS ARE SOARING

- According to the United Nations *World Intellectual Property Organization (WIPO) Technology Trends* report (2019) the number of new patents for artificial intelligence has been growing by 28% each year on average since 2012. Of the 340,000 AI patents filed since the 1950s, more than half have come since 2013.
- The report named IBM, Microsoft, Toshiba and Samsung as the most active companies in this field.
- The transportation sector has seen the fastest growth of AI inventions in the past few years, due to the development of autonomous vehicles. Other growth areas include telecommunications systems, robotic surgery, personalised medicine, personal digital devices and cameras.
- The top five patent holders in this field are based in the United States (IBM and Microsoft), Japan (Toshiba and NEC) and South Korea (Samsung).
- The top 20 list for patent holders is dominated by Japanese companies; it also includes the Chinese Academy of Sciences and the German industrial groups Siemens.

WHAT ARE ROBOTS?

- A robot is a machine which is programmable by a computer, and capable of carrying out a complex series of actions automatically. Robots can be guided by an external control device or the control may be embedded within. Robots may be constructed on the lines of human form, but most robots are machines designed to perform a task with no regard to their aesthetics.
- There is no consensus on which machines qualify as robots but there is general agreement that robots tend to possess some or all of the following abilities and functions: accept electronic programming, process data or physical perceptions electronically, operate autonomously to some degree, move around, operate physical parts of itself or physical processes, sense and manipulate their environment, and exhibit intelligent behaviour, especially behaviour which mimics humans or other animals.
- Existing types of robots include: general-purpose autonomous robots, factory robots (such as car production), domestic robots (like vacuum cleaners, lawn mowers), space probes, military robots and robots used in mining, healthcare and research.

SOURCES

World Intellectual Property Organization, *WIPO Technology Trends – Artificial Intelligence*, www.wipo.int/tech_trends/en/artificial_intelligence/
Wikipedia, *Robot*, <https://en.wikipedia.org/wiki/Robot>

Compiled by The Spinney Press.

HOW ROBOTS CHANGE THE WORLD

EXECUTIVE SUMMARY FROM A REPORT BY OXFORD ECONOMICS

20m – the number of manufacturing jobs that could be displaced by industrial robots by 2030 – 8.5% of the global manufacturing workforce.

Over the past decade, a robotics revolution has captured the world's imagination. As their capabilities expand, so does the rate at which industries purchase and install these increasingly intelligent machines. Since 2010, the global stock of industrial robots has more than doubled – and innovations in engineering and machine learning portend an accelerated adoption of robots in service sector occupations over the next five years.

This report sheds new light on both the current impact of robots on manufacturing jobs around the world and the potential for robots to transform the much larger (but as-yet far less automated) global services sector. To evaluate the implications of this ongoing robot revolution, we have brought together the combined expertise of Oxford Economics' economists, econometricians, modellers, and subject-matter experts.

The rise of robots has already had a profound effect on industrial employment around the world: today, approximately one of every three new manufacturing robots is being installed in China, the world's great workshop. Our econometric modelling finds that on average each newly installed robot displaces 1.6 manufacturing workers.¹ By 2030, we estimate that as many as 20 million additional manufacturing jobs worldwide could be displaced due to robotisation.²

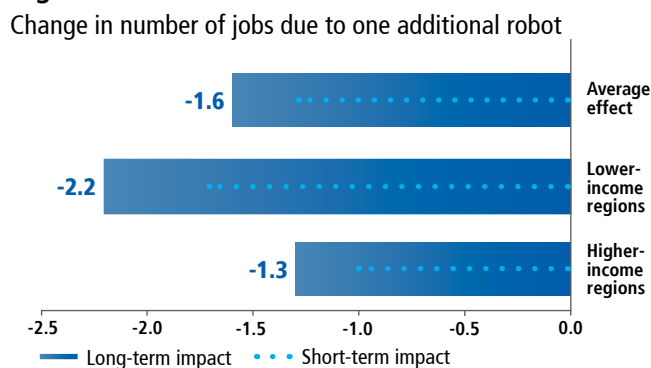
LOWER-INCOME REGIONS ARE MORE AT RISK

This great displacement will not be evenly distributed around the world, or within countries. Our research shows that the negative effects of robotisation are disproportionately felt in the lower-income regions of the globe's major economies – on average, a new robot displaces nearly twice as many jobs in lower-income regions compared with higher-income regions of the same country.³ At a time of worldwide concern about growing levels of economic inequality and political polarisation, this finding has important social and political implications.

Given the stakes, policy-makers need an early warning system to help them mitigate the risks of automation on employment. As part of this study, we have developed a Robot Vulnerability Index that ranks every region of seven developed economies in terms of how susceptible their respective workforces are to the installation of industrial robots.

In many cases, our Index highlights that the most vulnerable regions are somewhat removed from the wealthier districts of their home countries – such as Cumbria in the UK, Franche-Comté in France, and

Figure 1: Job losses from robots hit lower-income regions harder⁴



the high desert of Eastern Oregon in the US. These rural regions often include towns or cities with strong manufacturing heritages that play a surprisingly large part in the regional economy. In contrast, regions that surround knowledge-intensive cities, such as Toulouse and Grenoble in France, or Munich and Stuttgart in Germany, typically show much lower levels of vulnerability to the rise of the robots. This is also true of capital cities such as London, Paris, Seoul and Tokyo.

THE \$5 TRILLION ROBOTICS DIVIDEND

While regional impacts vary, fears about permanent global job destruction generated by robots appear somewhat exaggerated. Our study shows that the current wave of robotisation tends to boost productivity and economic growth, generating new employment opportunities at a rate comparable to the pace of job destruction. We estimate that a 1% increase in the stock of robots per worker in the manufacturing sector leads to 0.1% boost to output per worker across the wider workforce.

Our research shows the negative effects of robotisation are disproportionately felt in the lower-income regions of major economies.

These increases are large enough to drive meaningful growth. Using Oxford Economics' Global Economic Model (GEM), we calculated how changes in the rate of installation of industrial robots could affect the global economy. Overall, we found that a faster adoption of robots has a positive impact on both short- and medium-term growth. For example, boosting robot installations to 30% above the baseline forecast by 2030 would lead to an estimated 5.3% boost in global GDP that year. This equates to adding an extra \$4.9 trillion per year to the global economy by 2030 (in today's prices) – equivalent to an economy greater than the projected size of Germany's.

THE FUTURE OF SERVICE ROBOTS

Robots are steadily gaining traction in specific segments of the service economy, from baggage handling in

AUSTRALIAN REGIONS MOST VULNERABLE TO ROBOTS

- *How Robots Change the World* estimates that each new industrial robot eliminates 1.6 manufacturing jobs on average.
- The report suggests that in Australia, South Australia is most vulnerable to the future robot rollout. SA is Australia's most manufacturing-intensive state but has the slowest-growing economy and low levels of manufacturing productivity.
- Victoria and NSW are less vulnerable to robots than South Australia, and are also faster growing.
- NSW looks much less vulnerable than either Victoria or South Australia. In NSW the labour market has become less dependent on manufacturing jobs in recent years, while manufacturing productivity has improved.

Source: Khadem, N, (27 June 2019). 'Robots set to take 20m jobs globally by 2030, SA and Victoria will be hard hit, warns report', *ABC News*.

Regions most impacted by robots in the future

Region	Index score
South Australia	0.42
Victoria	0.39
Tasmania	0.37
Queensland	0.32
New South Wales	0.28
Western Australia	0.14
Northern Territory	0.06

Source: Oxford Economics.

airports to loading inventory in warehouses. In this report, we assess the likely impact (and timeframe) of service robot roll-outs in five key sectors: healthcare, retail, hospitality, transport, and construction and farming. For the purposes of this study we are considering robots only as physical machines, and not including the already-popular service-industry software like robotic process automation (RPA) that can speak, hear, read, conduct transactions, automate processes, and so on.

One key consideration for anticipating the pace of robot deployment in service industries is the environment in which these robots may be asked to operate – in particular, the extent to which service jobs include repetitive functions. Jobs like warehouse work are in imminent danger, while other jobs in less structured environments will likely be carried out by humans for decades to come.

As the pace of robotics adoption quickens, policy-makers will be faced with a dilemma: while robots enable growth, they exacerbate income inequality.

It will be difficult for machines to replace humans in service sector occupations that demand compassion, creativity, and social intelligence. Physical therapists, dog trainers, and social workers are likely to remain secure in their jobs, for instance, even if truckers and warehouse workers see the future of their jobs jeopardised.

POLICY IMPLICATIONS

As the pace of robotics adoption quickens, policy-makers will be faced with a dilemma: while robots enable growth, they exacerbate income inequality. Automation will continue to drive regional polarisation in many of the world's advanced economies, unevenly distributing the benefits and costs across the population. This trend will intensify as the impact of automation on jobs spreads from manufacturing to the services sector, making questions about how to deal with displaced workers increasingly critical.

It will be difficult for machines to replace humans in service sector occupations that demand compassion, creativity, and social intelligence.

The challenges will be daunting. Our analysis of the job moves of more than 35,000 US individuals over the course of their careers shows that more than half the workers who left production jobs in the past two decades were absorbed into just three occupational categories: transport, construction and maintenance, and office and administration work. Ominously, our analysis found that these three occupational areas are among the most vulnerable to automation over the next decade.

These findings, however, should not lead policy-makers and other stakeholders to seek to frustrate the adoption of robot technology. Instead the challenge should be to distribute the robotics dividend more evenly by helping vulnerable workers prepare for and adapt to the upheaval it will bring. Policy-makers, business leaders, technology companies, educators, and workers all have a role to play. We conclude the report with a framework for action for each of these groups to navigate the challenges and opportunities that robotisation will bring.

Robots are on the rise as never before. Preparing for and responding to the social impacts of automation will be a defining challenge of the next decade.

ENDNOTES

1. This finding is based on an analysis of a large, regional panel-dataset of robot stock, and other labour market indicators, over a 11-year timeframe, for 24 EU countries (minus Croatia, Cyprus, Luxembourg and Malta), along with Norway, the United States, Japan, and South Korea.
2. Countries included in this estimate account for more than 90% of industrial robot installations: EU 28, US, Japan, South Korea, Australia, China, Taiwan, Thailand, Mexico, India, Canada, Singapore, Brazil, Turkey, Malaysia. We assume the rate of robot installations in manufacturing up to 2030 follows the latest projections by the International Federation of Robotics, and we also account for long-term depreciation of existing robot stock.
3. Throughout this report, higher- and lower-income regions are defined as those with average household income levels above and below the national average, respectively.
4. Our modelling differentiates between a "short-term" effect, within the year of a robot installation, and a longer-term effect that builds over 10 to 15 years.

Oxford Economics (June 2019). *How Robots Change the World: what automation really means for jobs and productivity*, Executive summary, pp. 4-7. Retrieved from <http://resources.oxfordeconomics.com> on 9 July 2019.

Artificial intelligence may take your job, so political leaders need to start doing theirs

AI researchers think there is a 50% chance AI will outperform humans in all tasks in 45 years and that almost all current human jobs can be automated in 120 years. By **Alan Montague, Alan Nankervis, John Burgess, Nuttawuth Muenjohn, Ros Cameron and Timothy Bartram**

Over the next decade up to 40% of jobs could be replaced by automation. That's according to a forecast from the Committee for the Economic Development of Australia.

Other estimates are less dire. A report from the Organisation for Economic Cooperation and Development says only 14% of jobs across its member nations are "highly automatable", though another 32% are likely to change significantly.

Despite this uncertainty about the exact number of human jobs that will disappear, it is certain the impacts of automation driven by artificial intelligence (AI) are potentially profound.

If the new jobs that emerge as a result are fewer and less rewarding than those lost, we face a period of social dislocation.

The activities at which AI is expected to outperform humans by 2030 include translating languages, writing high-school essays and driving trucks. By mid-century it is expected AI could be capable of writing bestselling books or performing surgery.

Yet there is little evidence of any strategic planning and forward thinking by Australia's federal and state governments to minimise the potential downside. The recent federal budget, for example, was silent on this issue.

It is analogous to the lack of a strategic plan on climate change. Our concern about the policy vacuum is based on an extensive search on government websites, the media and budget speeches, which disturbingly yielded very little.

We think AI should be an election issue, because there is virtually no job that won't be affected by AI-driven automation. Being ready for this technological and social revolution should be at the heart of any government's economic and educational reform agenda.

WORKERS THAT NEVER REST

The activities at which AI is expected to outperform humans by 2030 include translating languages, writing high-school essays and driving trucks. By mid-century it is expected AI could be capable of writing bestselling books or performing surgery.

Researchers believe there is a 50% chance AI will outperform humans in all tasks in 45 years and that almost

TABLE 4.5. CROSS-COUNTRY VARIATION IN JOB AUTOMATABILITY

COUNTRY	Median	Mean	S.D.
New Zealand	0.39	0.42	0.20
Norway	0.40	0.41	0.18
Finland	0.41	0.43	0.18
United States	0.41	0.43	0.20
Northern Ireland (UK)	0.42	0.43	0.21
England (UK)	0.42	0.43	0.20
Sweden	0.43	0.44	0.19
Netherlands	0.44	0.45	0.19
Denmark	0.44	0.45	0.19
Canada	0.45	0.45	0.21
Ireland	0.45	0.46	0.22
Singapore	0.45	0.46	0.20
Belgium	0.46	0.46	0.20
Israel	0.46	0.47	0.21
Estonia	0.47	0.46	0.19
Korea	0.47	0.46	0.19
Austria	0.49	0.48	0.20
Russian Federation	0.49	0.47	0.19
Czech Republic	0.49	0.48	0.20
France	0.51	0.49	0.20
Italy	0.52	0.49	0.20
Cyprus	0.52	0.51	0.21
Poland	0.52	0.50	0.21
Japan	0.53	0.51	0.18
Slovenia	0.53	0.51	0.21
Spain	0.54	0.51	0.21
Germany	0.54	0.52	0.18
Chile	0.55	0.52	0.20
Turkey	0.55	0.52	0.18
Greece	0.57	0.54	0.16
Lithuania	0.57	0.54	0.19
Slovak Republic	0.62	0.57	0.20
ALL COUNTRIES	0.48	0.47	0.20

Risk of job automation by country. In New Zealand, for instance, the median worker has 39% chance of their job being automated. OECD Social, Employment and Migration Working Paper No. 202.



increase by 2.5 percentage points, based on up to 46% of jobs being automated by 2030.

Researchers believe there is a 50% chance AI will outperform humans in all tasks in 45 years and that almost all current human jobs can be automated in 120 years.

Inequality will also increase, the report says. To what extent depends on “how much Australia steps up its efforts to retrain and redeploy its surplus service, administrative and manual workers”. Without large-scale retraining, Australia’s Gini coefficient (the standard measure of income inequality) could increase from 0.32 to 0.41.

QUESTIONS FOR POLICYMAKERS

The stakes are high. Employment is so much more than just an income. A decent job provides purpose and dignity. It enables security, health and wellbeing. Joblessness hurts the individual, families and the wider community. If fewer people have work, and more people get paid less, the federal government will have to contend with falling income-tax revenue and rising welfare spending.

With intensifying global competition and rapid technological change, all governments need to be strategic. They need long-term policies to help business and workers gain the knowledge, skills and abilities needed to seize the opportunities and mitigate the threats.

There are three key questions for policymakers:

- To what extent will AI-driven automation increase

all current human jobs can be automated in 120 years.

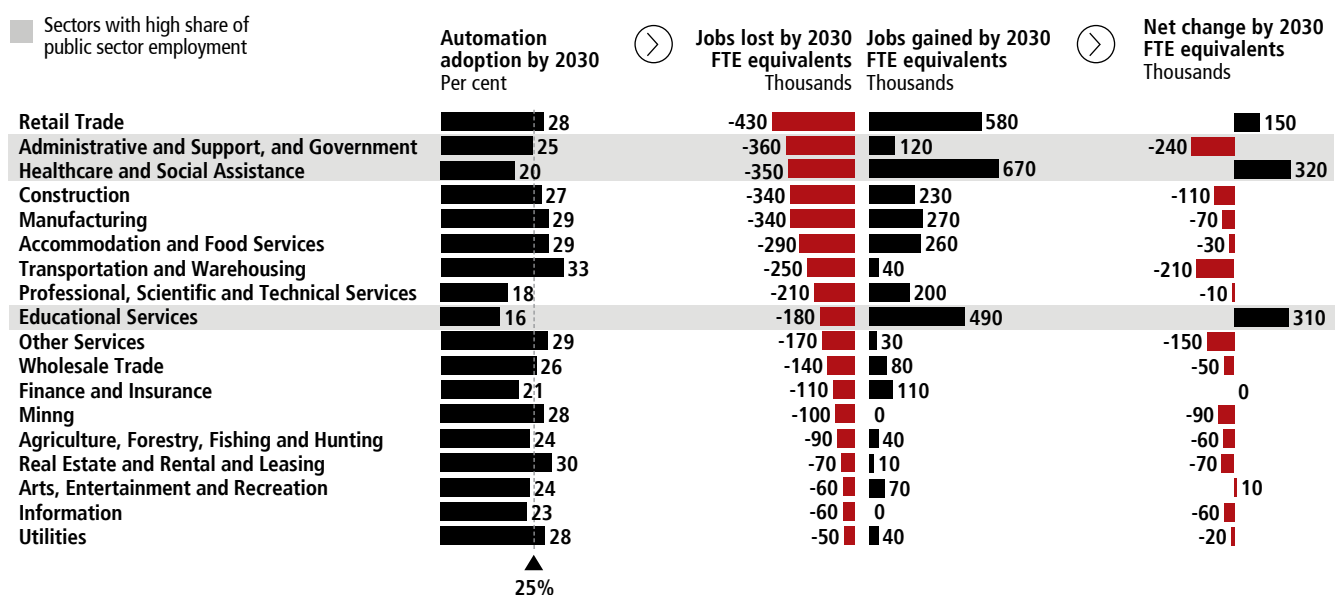
The World Economic Forum may argue that AI and automation “is about empowering people, not the rise of the machines” but there are good reasons to think some people will lose out.

Machines have many attractions to employers. They do not need rest, holidays or take sick leave. They will never complain about overtime or join a union.

A report published last month by management consultancy McKinsey Australia suggests, if proactive measures are not taken, the unemployment rate could

JOBS LOST, JOBS GAINED

Mid-point automation adoption scenario and step-up labour demand scenario, 2016-30



Note: Numbers may not sum due to rounding.

Source: Australia Statistics, McKinsey Global Institute (MGI), ONED, BLS, Oxford Economics.

The industries likely to experience the greatest disruption are retail, administration and government, health and social assistance, construction, manufacturing, and accommodation and food services.

This e-book is subject to the terms and conditions of a non-exclusive and non-transferable LICENCE AGREEMENT between THE SPINNEY PRESS and: Camberwell High School, Canterbury, library@camhigh.vic.edu.au

- unemployment and underemployment?
- How can governments and employers take advantage of AI and create the jobs of the future?
 - How can government, employers and educators equip employees and graduates with the skills to have jobs alongside robots, instead of competing with them?

The growth in AI and machine learning need not pose a threat to jobs and our standard of living, but without a strategic and shared effort, we will repeat the mistakes of previous technological revolutions, which imposed a terrible price on some in the name of progress.

DISCLOSURE STATEMENT

Four of the authors, Alan Montague, Ros Cameron, Alan Nankervis and Timothy Bartram, are members of The Australian HR Institute (AHRI) which has provided support and continues to do so for the research referred to in this article. The Australian HR Institute (AHRI) has in no way influenced the ethical processes involved with the research that is underway or the content in this article.

Alan Nankervis, John Burgess, Nuttawuth Muenjohn, Ros Cameron, and Timothy Bartram do not work for, consult, own shares in or receive funding from any company or organisation that would benefit from this article, and have disclosed no relevant affiliations beyond their academic appointment.

Alan Montague is a Friend of *The Conversation* and Senior Lecturer/Program Director, masters program in human resources management, RMIT University.

Alan Nankervis is Adjunct Professor of Management, RMIT University.

John Burgess is Professor of Human Resource Management, RMIT University.

Nuttawuth Muenjohn is Senior Lecturer, Business, RMIT University.

Ros Cameron is Head of Discipline, Human Resources Management, Australian Institute of Business.

Timothy Bartram is Professor, Management, RMIT University.

THE CONVERSATION

Montague, A, Nankervis, A, Burgess, J, Muenjohn, N, Cameron, R, and Bartram, T (11 April 2019). *Artificial intelligence may take your job, so political leaders need to start doing theirs*. Retrieved from <http://theconversation.com> on 9 July 2019.

OVERCOMING OUR MISTRUST OF ROBOTS IN OUR HOMES AND WORKPLACES

WHAT IS IT ABOUT AI THAT UNNERVES US? AUSTRALIA'S CHIEF SCIENTIST ALAN FINKEL SUSPECTS IT'S A COMBINATION OF THINGS

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?

You do. 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 prepared my breakfast – trusting that I wouldn't be poisoned by salmonella in my factory-processed muesli.

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.

All of these decisions make sense to us, because we know that we live in a society where human behaviour is governed by conventions and rules.

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 collaborate – and innovate – 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 in a human?

To chat with your child? To drive your taxi? To read your brain scan? To scan your face at a concert, at work, or in a supermarket?

WHAT'S DIFFERENT ABOUT AI

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 small community of experts. Perhaps 22,000 worldwide are qualified to the level of a PhD in AI.

People with rare skills in high demand come at an eye-watering price. That makes them very hard to keep in universities and public agencies. It's not unknown for technology developers to buy up IT faculties. 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.

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

Now for 68 years the world has been waiting for an AI that could fool us into thinking it was human. We call it the Turing test, after the scientist who proposed it in 1950, Alan Turing.

He began with the question: "Can machines think?" Since we can't agree on a single test for "thinking", Turing 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?

Watching Google's AI 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 – because every phone call you make or receive now carries a niggling doubt. That could be a machine.

Hand in hand with our lack of knowledge is our lack of foreknowledge. We give up our data today without knowing what others might be able to do with it tomorrow.

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.

RESPONSES TO ARTIFICIAL INTELLIGENCE

One response to these questions would be to conclude that the only safe way forward is to ban AI. 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, the future has infinite promise?

Health care, available to all, tailored precisely to the individual. Your personal chauffeur, 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. 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 protect us every time we walk across a pedestrian crossing.

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.

There are industry codes, standards and regulations. 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, would be 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.

I'll leave that particular discussion on weapons to another occasion. Instead, I've been thinking about the role that Australia could play at the left-hand end of the spectrum, where we interact with technology providers as everyday consumers.

We would expect a government department to vet the provider of a service that, for example, replaced social workers with algorithms that predict the likelihood that a particular child will be violent.

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 companies that want your business 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, just as they currently look for the "Fairtrade" logo on coffee.

WHY A VOLUNTARY CERTIFICATE WOULD WORK

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.

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 to customers prepared to pay a premium.

Would the Turing Stamp be granted to organisations or products? Both. It is the model that has long been accepted in the manufacturing sector.

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.

And when you trade with an AI developer you expect to have an ongoing relationship. If you buy a washing machine, it will still be the same washing machine in five years' time. If you download an app, it may well be radically different in five weeks.

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.

A system that verifies quality and prioritises ethics will reward Australia. Last week's federal budget made what I have described as a "promising first instalment". A\$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.

This article is an extract of a keynote speech Alan Finkel delivered to a Committee for Economic Development of Australia (CEDA) event in Sydney on artificial intelligence.

Alan Finkel is Australia's Chief Scientist, Office of the Chief Scientist.

THE CONVERSATION

Finkel, A (18 May 2018). *Finkel: overcoming our mistrust of robots in our homes and workplaces*. Retrieved from <http://theconversation.com> on 9 July 2019.

We should learn to work with robots and not worry about them taking our jobs

HUMANS AND ROBOTS CAN WORK TOGETHER TO CREATE JOBS FOR THE FUTURE, ACCORDING TO **DAVID TUFFLEY**



We have all heard the dire predictions about robots coming to steal our jobs. Some would even have us believe these silicon bogeymen are coming to kill us. It plays straight into people's darkest fears about technology.

When futurists talk about things that haven't happened yet, they are free to parade educated guesses as fact. But before we take their word for it, we might remember the old adage:

... in God we trust, all others bring data.

In a recent article, the *MIT Technology Review* tabulated the results of "every study we could find on what automation will do to jobs". The results show that the expected impacts depend on what you measure.

Many predictions, little agreement

Of the 19 reports considered in the review, there was enormous variation. Some predicted that a few million jobs would be replaced, while others spoke portentously of tens or hundreds of millions over similar time frames. Some were decidedly upbeat, others quite gloomy.

One futurist went so far as to forecast that a billion jobs will be lost to automation by 2022. Contrast this with the more sober prediction from the research and advisory group Gartner of 1.8 million jobs lost by 2020, but with 2.3 million created in the same period – a net increase of 500,000 over the next two years.

Why such a big difference? In truth, no one knows how many jobs will be lost and found in the age of Artificial Intelligence (AI). The situation is too complex for simple answers.

Variations in predictions can be likened to the parable of the five blind men encountering an elephant. By touching different parts of the elephant's body, each came to a different conclusion as to what the beast is.

Technology anxiety is nothing new

Worries about the impact of technology on society have a long history. In the 18th and 19th centuries, the rapid expansion of disruptive technologies during the Industrial Revolution gave rise to the same anxieties as those being expressed today.

Trades were automated to produce greater economies of scale, but job losses were more than offset by the new jobs subsequently created. Meanwhile, trades like pottery, weaving and metalwork that were "lost" to automation 200 years ago are still being done by skilled craftspeople today.

More recently, when personal computers found their way onto people's desks in the 1980s, the typing pool became redundant. I recall the lamentations then of the newspapers, TV and talkback radio.

In truth, no one knows how many jobs will be lost and found in the age of Artificial Intelligence (AI). The situation is too complex for simple answers.

But in time, the overall number of jobs went up because of the new jobs created in the fledgling IT industry. Today there are dozens of technology job categories, none of which existed in the typing pool days – jobs in computer hardware, programming, content production, web design, security, big data, sales and marketing, and artificial intelligence to name a few.

And the former typists? Their skills were in more demand than ever, because keyboards are still the way humans communicate with computers.

AI is only an extension of us

AI is often spoken of as a separate entity from people, and sometimes seen as a dangerous adversary. The reality is that it is merely a tool created by humans – it only does what we tell it to do and nothing more.

What we have is "narrow AI", only suited to very particular tasks. Autonomous "general AI" that is superior to the spectrum of human intelligence is many decades away.

The “narrow AI” GPS on my smartphone is much better at navigating than I am. It extends my ability to go places almost miraculously. But there’s no reason to feel threatened. It is only smarter than me in that one ability. And it is not at all likely to say “I’m sorry Dave, I’m afraid I can’t do that” like Hal refusing to open the pod bay doors in film *2001: A Space Odyssey*.

Likewise, AI can greatly improve people’s competence in the workplace. In the first recorded case of an AI saving someone’s life, it was the combination of human doctors and a diagnostic AI that succeeded where the human doctors alone had failed.

The obvious advantages of enhancing human intelligence with AI have given rise to the hybrid known as a modern centaur.

The concept was illustrated by chess grandmaster Garry Kasparov, who observed that the best players are not computers alone, but human intelligence augmented with AI.

Benefits outweigh the harm

Historically, fears about technology have largely proven unfounded, at least in terms of the benefits outweighing the harm, if not in other ways. Our challenge is to maximise the benefits and minimise the harm.

What skills will we need for future employment? These would be the same skills that humans have always excelled at – critical thinking and problem solving, good communication and teamwork, leadership, initiative, creativity. And of course, the willingness to leverage the current technology.

Futurists tend to assume that if a job can be automated, it will be automated. But that is certainly not true.

AI will automate some jobs, particularly the dirty, dull or dangerous ones that people don’t want to do – everything from sewer reconnaissance to repetitive factory work. Manual welding, for example, can produce highly toxic fumes – a prime candidate for automation.

What we have is “narrow AI”, only suited to very particular tasks. Autonomous “general AI” that is superior to the spectrum of human intelligence is many decades away.

But some jobs will always be done by people and the reasons can vary greatly, from economic, social and nostalgic reasons, to the fact that some jobs are simply not practical for robots to do.

When I go to the doctor, I want a human sitting across from me. I don’t want a holographic doctor who demands to know the nature of my medical emergency.

DISCLOSURE STATEMENT

David Tuffley does not work for, consult, own shares in or receive funding from any company or organisation that would benefit from this article, and has disclosed no relevant affiliations beyond his academic appointment.

David Tuffley is Senior Lecturer in Applied Ethics and SocioTechnical Studies, School of ICT, Griffith University.

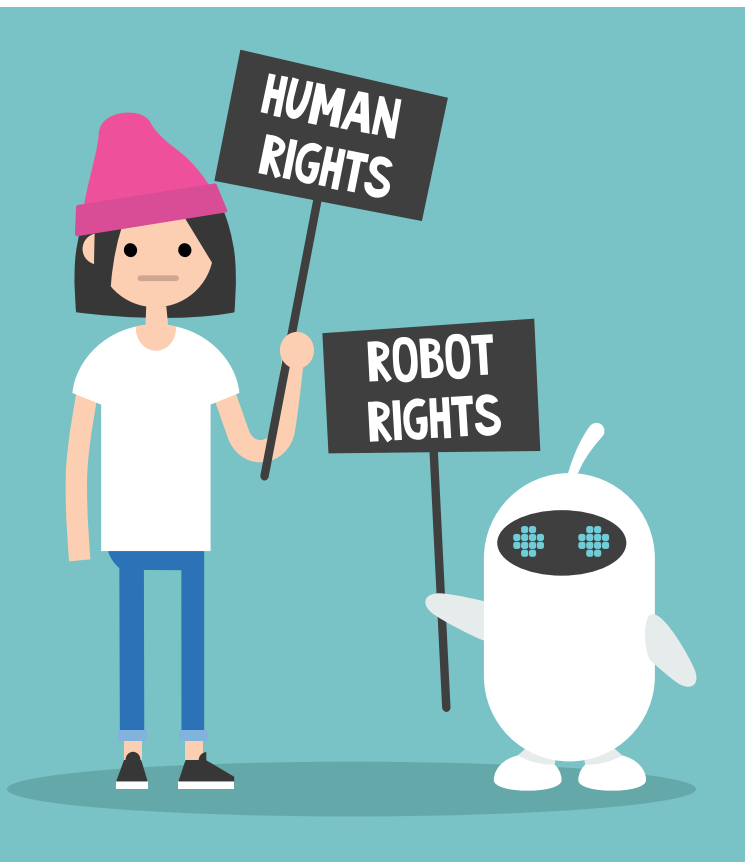
THE CONVERSATION

Tuffley, D (19 February 2018). *We should learn to work with robots and not worry about them taking our jobs*. Retrieved from <http://theconversation.com> on 9 July 2019.



Ethics of AI: how should we treat rational, sentient robots – if they existed?

Philosopher **Hugh McLachlan** explores the ethical prospects of a brave new world in which robots possess human consciousness



Imagine a world where humans co-existed with beings who, like us, had minds, thoughts, feelings, self-conscious awareness and the capacity to perform purposeful actions – but, unlike us, these beings had artificial mechanical bodies that could be switched on and off.

That brave new world would throw up many issues as we came to terms with our robot counterparts as part and parcel of everyday life. How should we behave towards them? What moral duties would we have? What moral rights would such non-human persons have? Would it be morally permissible to try to thwart their emergence? Or would we have a duty to promote and foster their existence?

Intriguing ethical questions such as these are raised in Ian McEwan’s recent novel, *Machines Like Me*, in which Alan Turing lives a long successful life and explosively propels the development of artificial intelligence (AI) that leads to the creation of “a manufactured human with plausible intelligence and looks, believable motion and shifts of expression”.

As intellectual speculation, to consider the ethics of the treatment of rational, sentient machines is interesting. But two common arguments might suggest that the matter has no practical relevance and any ethical

questions need not be taken seriously.

The first is that such artificial people could not possibly exist. The second, often raised in the abortion debate, is that only persons who have living and independently viable human bodies are due moral respect and are worthy of moral consideration. As we shall see, these arguments are debatable.

Mind, matter and emergent properties

We might suppose that mental phenomena – consciousness, thoughts, feelings and so on, are somehow different from the stuff that constitutes computers and other machines manufactured by humans. And we might suppose that material brains and material machines are fundamentally different from conscious minds. But whether or not such suppositions are true – and I think that they are – it does not follow that sentient, consciously aware, artificially produced people are not possible.

The French sociologist Emile Durkheim has argued very convincingly that we should beware of simplistic arguments in social science. Social phenomena, such as language, could not exist without the interaction of individual human beings with their particular psychological and biological features. But it does not follow that the resultant social phenomena – or “emergent properties” – can be completely and correctly explained solely in terms of these features.

The same point about the possibility of emergent properties applies to all sciences. There could not be, for instance, computers of the sort I am now working at without the pieces of plastic, wires, silicon chips and so forth that make up the machine. Still, the operations of a computer cannot be explained solely in terms of the features of these individual components. Once these components are combined and interact in particular ways with electricity, a phenomenon of a new sort emerges: a computer. Similarly, once computers are combined and interact in particular ways, the internet is created. But clearly, the internet is a different sort of phenomenon from a tangible, physical computer.

In a similar way, we need not suppose that minds are reducible to brains, molecules, atoms or any other physical elements that are required for them to function. They might be entities of a different sort that emerge from particular interactions and combinations of them.

There’s no obvious logical reason why conscious awareness of the sort that human beings possess – the capacity to think and make decisions – could not appear in a human machine some day. Whether it is physically possible and, therefore likely to actually happen, is open to debate.

Do machines deserve our consideration?

It doesn't seem controversial to say that we shouldn't slander dead people or wantonly destroy the planet so that future generations of unborn people are unable to enjoy it as we have.

Both groups are due moral respect and consideration. They should be regarded as potential objects of our moral duties and potential recipients of our benevolence.

But the dead and the yet to be born do not have viable bodies of any sort – whether natural or artificial. To deny conscious persons moral respect and consideration on the grounds that they had artificial rather than natural bodies would seem to be arbitrary and whimsical. It would require a justification, and it is not obvious what that might be.

One day, maybe sooner than we think, a consideration of the ethics of the treatment of rational, sentient machines might turn out to be more than an abstract academic exercise.

How should we behave towards them? What moral duties would we have? What moral rights would such non-human persons have? Would it be morally permissible to try to thwart their emergence? Or would we have a duty to promote and foster their existence?

DISCLOSURE STATEMENT

Hugh McLachlan does not work for, consult, own shares in or receive funding from any company or organisation that would benefit from this article, and has disclosed no relevant affiliations beyond his academic appointment.

Hugh McLachlan is Professor Emeritus of Applied Philosophy, Glasgow Caledonian University.

THE CONVERSATION

McLachlan, H (14 June 2019). *Ethics of AI: how should we treat rational, sentient robots – if they existed?* Retrieved from <http://theconversation.com> on 9 July 2019.

ROBOT RIGHTS AND HUMANS

If at some stage in the future, a machine could think, decide and act on its own volition, if it can be harmed or held responsible for its actions, should we stop treating it like property and start treating it more like a person with rights? What if a robot achieves true self-awareness? Should it have equal rights as humans, with similar protections under the law?¹

- *If we did give robots some kind of legal status, what would it be? If they behaved like humans we could treat them like legal subjects rather than legal objects, or at least something in between. Legal subjects have rights and duties, and this gives them legal 'personhood'. They do not have to be physical persons; a corporation is not a physical person but is recognised as a legal subject. Legal objects, on the other hand, do not have rights or duties although they may have economic value.¹*
- *You might argue that animals differ from machines in that they have some sort of conscious experience ... However, consciousness and human rights are not the only basis of rights. As machines become even more complex and intelligent, just discarding or destroying them without asking any questions at all about their moral and physical integrity seems ethically wrong. Just like rivers, they too should receive rights because of their meaning to us.²*
- *Once our machines acquire a base set of human-like capacities, it may be incumbent upon us to look upon them as social equals, and not just pieces of property. The challenge will be in deciding which cognitive thresholds, or traits, qualify an entity for moral consideration, and by consequence, social rights.³*
- *The three most important thresholds in ethics are the capacity to experience pain, self-awareness, and the capacity to be a responsible moral actor. In humans, if we are lucky, these traits develop sequentially. But in machine intelligence it may be possible to have a good citizen that is not self-aware or a self-aware robot that doesn't experience pleasure and pain.³*
- *It's important to point out that intelligence is not the same as sentience (the ability to perceive or feel things), consciousness (awareness of one's body and environment) and self-awareness (recognition of that consciousness). A machine or algorithm could be as smart – if not smarter – than humans, but still lack these important capacities.³*
- *Some believe that self-awareness comes with some minimal citizenship rights, such as the right to not be owned, and to have its interests to life, liberty and growth respected. With both self-awareness and moral capacity should come full adult human citizenship rights, such as the rights to make contracts, own property, vote and so on.³*
- *But some thinkers argue that only humans should be allowed to participate in the social contract (human exceptionalism), and that the world can be properly arranged into Homo sapiens and everything else, be it your gaming console, refrigerator, pet dog or companion robot.³*
- *Once our machines reach a certain threshold of sophistication, we may no longer be able to exclude them from our society, institutions and laws. We will have no good reason to deny them human rights; to do otherwise would be tantamount to discrimination and slavery.³*

SOURCES

1. <https://theconversation.com/robot-rights-at-what-point-should-an-intelligent-machine-be-considered-a-person-72410>
2. <https://theconversation.com/please-alex-a-are-we-beginning-to-recognise-the-rights-of-intelligent-machines-96287>
3. <https://www.gizmodo.com.au/2017/06/when-will-robots-deserve-human-rights/>

Compiled by The Spinney Press.

Artificial intelligence is being trained to have empathy. Should we be worried?

Will robots ever really be able to empathise with us? This scenario is explored in the following **ABC Science** article by technology reporter Ariel Bogle

Less Hal and more *Her*, responding warmly to the feelings of others may no longer be a uniquely animal quality. Empathetic responses are being integrated into artificial intelligence and robotics, raising sticky ethical questions.

The shift can be subtle or overt – from emotionally appropriate gestures from your smartphone’s voice assistant, to comforting robotics in clinical situations.

For instance, Danielle Krettek, the founder of Google’s Empathy Lab, said her work has contributed to some of the Google Assistant’s apparent ability to attune to your mood.

“When you say, ‘I’m feeling depressed’, instead of giving you a description of what depression is, it [might say], ‘you know what, a lot of people feel that. You’re not alone,’” she explained at the design conference Semi Permanent in Sydney.

“It’s just these little moments of ‘attunement’ that really help a person feel seen and heard and met where they are.”

But are we OK with a robot making you think it cares?

From touch to voice

The move towards empathetic gestures in our technology comes as voice-activated devices are growing in popularity.

The coming generation of products, driven by voice, will rely more on inference, emotion and trust, according to Ms Krettek, whose lab aims to bring “deep humanity to deep learning”. Even, possibly, a mutual vulnerability.

According to social neuroscientist Pascal Molenberghs at the University of Melbourne, one component of empathy, affective empathy, is when we share the emotion of others: when you watch a character flee in a scary movie, for example, and feel scared yourself.

AI could be programmed to identify and express that it’s feeling your pain, but given it cannot truly “feel” anything, this might involve a degree of mimicry.

“We simulate in ourselves the emotions we observe in others. It’s like an understanding from within,” he explained.

AI could be programmed to identify and express that it’s feeling your pain, but given it cannot truly “feel” anything, this might involve a degree of mimicry. If you tell your AI assistant a sad story, you don’t want it to reply in a chirpy tone.

“They [might] mimic the emotional connotation in the words that they’re using ... something that humans often do automatically,” Dr Molenberghs explained.

Do we want empathetic AI, anyway?

Ron Arkin, director of the Mobile Robot Laboratory at Georgia Tech, is investigating the use of robot empathy in the treatment of Parkinson’s disease.

There is a condition in the early stages of the disease known as “facial masking” – the face may become expressionless and the voice may lose its ability to express emotion. Because the patient finds it more difficult to transmit their internal emotional state, the caregiver might start to lose empathy, potentially reducing the quality of care.

Dr Arkin’s team is investigating a “bystander robot” that will subtly monitor the relationship between the patient and the caregiver, and if interactions start to deteriorate, nudge things back in a better direction – by “quizzically looking at the person” who is losing empathy, for example.

In Sydney to speak at the University of New South Wales, Dr Arkin said this type of robot needs what he called “a partial theory of mind model”.

“This requires that the robot has ... some model of what the caregiver is feeling and what the patient is feeling,” he explained.





Of course, Dr Arkin emphasised, the robot never feels anything itself: “the point is that the robot can make you think that it has that emotion, but it’s not actually feeling anything.”

Is an empathetic AI ethical?

Is it OK for AI to make you think it cares? Ms Krettek said honesty is key.

“If we get the transparency, the clarity and the function right, and it feels good, then we’ll be in a more harmonious space with technology than we are now – which is a bit kind of cacophonous and overwhelming,” she said.

Of course, deception for our benefit is still a form of manipulation. And if built for our biggest technology companies, it comes with a commercial imperative.

“In the future, a very warm and convincing voice will say, ‘don’t you feel like a Coca Cola today?’” suggested Monash University philosopher Robert Sparrow.

“Systems that are capable of shaping the emotional states of their users will be much more effective at selling people things.”

While you may know consciously that the system doesn’t really care about you, your actions can still be affected.

Professor Sparrow, who has written critically of the use of robots in aged care, suggested that designing for deception can be disrespectful of a person – even if it ostensibly makes them happy.

“It’s ethically problematic to say, ‘I’m going to deceive people for their own good’,” he said.

“[There is a] difference between having a good life or genuinely making someone better off, and making them feel better.”

Is it OK to care for your robot? Of course, the flip-side also requires careful consideration: what do we do when we truly care for our robots?

Humans are infamous for forming emotional connections with things that can’t love us back – from dolls to trees. An empathetic robot could know just which buttons to push.

“Is there a fundamental human right being violated – your right to perceive the world as it actually is?” Dr Arkin asked.

Of course, deception for our benefit is still a form of manipulation. And if built for our biggest technology companies, it comes with a commercial imperative.

“The point is, I can make you empathise for that robot. It’s like acting.”

Nevertheless, Professor Sparrow speculated that an overly emotive piece of technology might be a turn-off.

“Think of how no one makes phone calls anymore – people would much rather send a text message than speak on the phone. I think that’s about reducing emotional load,” he said.

“If you imagine that your phone is pleading like a small child ... you’re not necessarily going to be more involved with the device.”

© ABC. Reproduced by permission of the Australian Broadcasting Corporation – Library Sales.

Bogle, A (2 June 2018). ‘Artificial intelligence is being trained to have empathy. Should we be worried?’, *ABC News*. Retrieved from www.abc.net.au/news on 9 July 2019.

Careful how you treat today's AI: it might take revenge in the future

We might not like the way future AI responds to us, cautions **Nicholas Agar**

Artificial intelligence (AI) systems are becoming more like us. You can ask Google Home to switch off your bedroom lights, much as you might ask your human partner.

When you text inquiries to Amazon online it's sometimes unclear whether you're being answered by a human or the company's chatbot technology.

There's clearly a market for machines with human psychological abilities. But we should spare a thought for what we might inadvertently create.

What if we make AI so good at being human that our treatment of it can cause it to suffer? It might feel entitled to take revenge on us.

MACHINES THAT 'FEEL'

With human psychological abilities may come sentience. Philosophers understand sentience as the capacity to suffer and to feel pleasure. And sentient beings can be harmed. It's an issue raised by the Australian philosopher Peter Singer in his 1975 book *Animal*

Liberation, which asked how we should treat non-human animals. He wrote:

If a being suffers, there can be no moral justification for refusing to take that suffering into consideration. No matter what the nature of the being, the principle of equality requires that its suffering be counted equally with the like suffering – insofar as rough comparisons can be made – of any other being.

Singer has devoted a career to speaking up for animals, which are sentient beings incapable of speaking up for themselves.

SPEAKING UP FOR AI

Researchers in AI are seeking to make an AGI or artificial general intelligence – a machine capable of any intellectual task performed by a human being. AI can already learn, but AGI will be able to perform tasks beyond that for which it is programmed.

The experts disagree on how far off an AGI is. The US tech inventor Ray Kurzweil expects an AGI soon, maybe 2029. Others think we might have to wait for a century.

But if we are interested in treating sentient beings right, we may not have to wait until the arrival of an AGI. One of Singer's points is that many sentient beings fall far short of human intelligence. By that argument, AI doesn't have to be as intelligent as a human for it to be sentient.

The problem is there is no straightforward test for sentience. Sending a human-crewed mission to Mars is very challenging, but at least we'll know when we've done it. Making a machine with feelings is challenging in a more philosophically perplexing way. Because we lack clear criteria for machine sentience, we can't be sure when we've done it.

LOOK TO SCIENCE FICTION

The ambiguity of machine sentience is a feature of several science fiction presentations of AI.

For example, Niska is a humanoid robot, a synth, serving as a sex worker in the TV series *Humans*. We are told that, unlike most synths, she is sentient.

When Niska is questioned about why she killed a client she explains: *He wanted to be rough.*

The human lawyer Laura Hawkins responds: *But, is that wrong if he didn't think you could feel? ... Isn't it better he exercises his fantasies with you in a brothel rather than take them out on someone who can actually feel?*

From a human perspective one could think sexual assault directed against a non-sentient machine is a

What if we make AI so good at being human that our treatment of it can cause it to suffer? It might feel entitled to take revenge on us.





victimless crime. But what about a sex robot that has acquired sentience? Niska goes on to explain that she was scared by the client's behaviour towards her.

And I'm sorry I can't cry or ... bleed or wring my hands so you know that. But I'm telling you, I was.

Humans is not the only science fiction story to warn of revenge attacks from machines designed to be exploited by humans for pleasure and pain. In the TV remake of *Westworld*, humans enter a theme park and kill android hosts with the abandon of Xbox massacres, confident their victims have no hard feelings because they can't have any feelings.

But here again, some hosts have secretly acquired sentience and get payback on their human tormentors.

WE'RE ONLY HUMAN

Is it only science fiction? Are sentient machines a long way off? Perhaps. Perhaps not.

But bad habits can take a while to unlearn. We – or rather animals – are still suffering the philosophical hangover of the 17th century French thinker Rene Descartes' terrible idea that animals are mindless automata – lacking in sentience.

If we are going to make machines with human psychological capacities, we should prepare for the possibility that they may become sentient. How then will they react to our behaviour towards them?

Perhaps our behaviour towards non-sentient AI today should be driven by how we would expect people to behave towards any future sentient AI that can feel, that can suffer. How would we expect that future sentient machine to react towards us?

This may be the big difference between machines and the animals that Singer defends. Animals cannot take revenge. But sentient machines just might.

DISCLOSURE STATEMENT

Nicholas Agar does not work for, consult, own shares in or receive funding from any company or organisation that would benefit from this article, and has disclosed no relevant affiliations beyond his academic appointment.

Nicholas Agar is Professor of Ethics, Victoria University of Wellington.

THE CONVERSATION

Agar, N (13 March 2019). *Careful how you treat today's AI: it might take revenge in the future*. Retrieved from <http://theconversation.com> on 9 July 2019.

Before replacing a carer with a robot, we need to assess the pros and cons

Governments need to carefully plan for the inevitable expansion of robot technologies to safeguard vulnerable people, warn **Helen Dickinson** and **Catherine Smith**

If you have seen science fiction television series such as *Humans* or *Westworld*, you might be imagining a near future where intelligent, humanoid robots play an important role in meeting the needs of people, including caring for children or older relatives.

The reality is that current technologies in this sector are not yet very humanoid, but nonetheless, a range of robots are being used in our care services including disability, aged care, education and health.

Our new research, published today by the Australia and New Zealand School of Government, finds that governments need to carefully plan for the inevitable expansion of these technologies to safeguard vulnerable people.

The reality is that current technologies in this sector are not yet very humanoid, but nonetheless, a range of robots are being used in our care services including disability, aged care, education, and health.

CARE CRISIS AND THE RISE OF ROBOTS

Australia, like a number of other advanced liberal democracies, is anticipating a future with an older population, with a more complex mix of chronic illness and disease. A number of care organisations already operate under tight fiscal constraints and report challenges recruiting enough qualified staff.

In the future, fewer numbers in the working-age population and increased numbers of retirees will compound this problem. If we then add to this equation the fact consumer expectations are increasing, it starts to look like future care services are facing a somewhat perfect storm.

Robots are increasingly becoming a feature of our care services, capable of fulfilling a number of roles from manual tasks through to social interaction. Their wider use has been heralded as an important tool in dealing with our impending care crisis. Countries such as Japan see robots playing a key role in filling their workforce gaps in care services.

A number of Australian residential aged care facilities are using Paro, a therapeutic robot that looks and sounds like a baby harp seal. Paro interacts by moving its head, heavily-lashed wide eyes and flippers, making sounds and responding to particular forms of touch on its furry coat. Paro has been used extensively in aged care in the United States, Europe and parts of Asia, typically among people living with dementia.

Nao is an interactive companion robot developed in a humanoid form but standing just 58cm tall in height. Nao has gone through a number of different iterations and has been used for a variety of different applications worldwide, including to help children engaged in paediatric rehabilitation and in various educational and research institutes.



THE DOUBLE-EDGED SWORD OF TECHNOLOGY

Robots are capable of enhancing productivity and improving quality and safety. But there is a potential for misuse or unintended consequences. Concerns have been expressed about the use of robots potentially reducing privacy, exposing people to data hacking, or even inflicting physical harm. We also lack evidence about the potential long-term implications of human-machine interactions.

Our research explored the roles robots should and, even more critically, should not play in care delivery. We also investigated the role of government as a steward in shaping this framework through interviews with 35 policy, health care and academic experts from across Australia and New Zealand.

We found that despite these technologies already being in use in aged care facilities, schools and hospitals, government agencies don't typically think strategically about their use and often aren't aware of the risks and potential unintended consequences. This means the sector is largely being driven by the interests of technology suppliers. Providers in some cases are purchasing these technologies to differentiate them in the market, but are also not always engaging in critical analysis.

Our study participants identified that robots were "leveraged" as something new and attractive to keep young people interested in learning, or as "a conversation starter" with prospective families exploring aged care providers.

But there are significant risks as the technologies become more developed. Drawing on research in other emerging technologies, our participants raised concerns about addiction and reliance on the robot. What would happen if the robot broke or became obsolete, and who would be responsible if a robot caused harm?

As artificial intelligence develops, robots will develop different levels of capabilities for "knowing" the human they are caring for. This raises concerns about potential hacking and security issues. On the flip side, it raises questions of inequity if different levels of care available at different price points.

Participants were also concerned about the unintended consequences of robot relationships on human relationships. Families may feel that the robot proxy is sufficient companionship, for instance, and leave their aged relative socially isolated.

WHAT SHOULD GOVERNMENTS DO?

Government has an important role to play by regulating the rapidly developing market. We suggest a responsive regulatory approach, which relies on the sector to self- and peer-regulate, and to escalate issues as they arise for subsequent regulation. Such engagement will require education, behaviour change, and a variety of regulatory measures that go beyond formal rules.

Government has an important role in helping providers understand the different technologies available



and their evidence base. Care providers often struggle to access good evidence about technologies and their effectiveness. As such, they're largely being informed by the market, rather than high quality evidence.

Governments have a responsibility to ensure vulnerable people aren't exploited or harmed by technologies. And they must also ensure robots don't replace human care and lead to greater social isolation.

Many of the stakeholders we spoke to for our research also see a role for government in helping generate an evidence base that's accessible to providers. This is particularly important where technologies may have been tested, but in a different national context. Many respondents called for establishment of industry standards to protect against data and privacy threats, and the loss of jobs.

Finally, governments have a responsibility to ensure vulnerable people aren't exploited or harmed by technologies. And they must also ensure robots don't replace human care and lead to greater social isolation.

DISCLOSURE STATEMENT

Helen Dickinson receives funding from Australia and New Zealand School of Government, Australian Research Council and National Health and Medical Research Council. Catherine Smith receives funding from Australia and New Zealand School of Government.

Helen Dickinson is Associate Professor, Public Service Research Group, UNSW.

Catherine Smith is Research fellow, University of Melbourne.

THE CONVERSATION

Dickinson, H, and Smith, C (9 November 2018). *Before replacing a carer with a robot, we need to assess the pros and cons.* Retrieved from <http://theconversation.com> on 9 July 2019.

Driverless cars: everything you need to know about the transport revolution

Imagine being able to order a driverless car through an app on your phone, in much the same way as you order an Uber now. By Jane Cowan for **ABC News**

But there's no more having to allow an extra half hour in case the traffic is bad – your driverless car will know the traffic conditions.

In fact, it will know if your train is running five minutes late and adjust your departure time accordingly, so you can spend the extra time getting dressed. On the road, it will adjust its speed to avoid accidents and catch the green lights – all while you read the paper or catch up on emails. A second driverless car rolls up at its regularly scheduled time to take the kids to school across town in the other direction. In a seamless ballet, your morning routine is coordinated with everything calibrated to minimise time on the road and maximise your capacity to do other things.

But how close is this science-fictionesque vision?

NOT SO FAST

The technology for driverless cars is here now, says Associate Professor Allison Kealy, who researches high performance positioning systems for intelligent transport at the Department of Infrastructure Engineering at the University of Melbourne.

“The primary limitation is that to make them safe, they're very expensive. So you can spend a million bucks but nobody's going to buy it,” she says.

Another obstacle facing Australia, according to Professor Kealy, is that it's essentially reliant on foreign technology development to realise the dream here. While driverless demonstration cars have been built here by the likes of Bosch, when it comes to commercially viable vehicles, we're talking about cars manufactured overseas, and designed to perform on the infrastructure

that exists in Japan or Europe or the United States.

Then there's the issue of regulation. Who is responsible if there's an accident? The individual or the car manufacturer? What about insurance?

Not to mention the human factor.

“The thought of a computer driving people around is very scary to a lot of people,” says automotive engineer Stuart Ballingall, who is the director of transport futures for VicRoads and holds a similar position at the national level.

“This can't just be done by technologists in a corner, it's quite important we bring the community along on this ride as well.”

THE TECHNOLOGY: HOW DOES IT WORK?

What if you're in a driverless car and another car, driven by a human, is about to T-bone you? Can you seize back control and swerve out of the way?

Driverless cars are very unlikely to make mistakes, says Ian Christensen from the iMOVE CRC, a consortium of science and industry designed to bring driverless technology to market.

But driverless cars cannot undo mistakes made by human drivers around them, which means if another human driver is about to hit you, it's unlikely your driverless car can do much to avoid the accident.

What about a different scenario: a dog runs onto the road. How does the driverless car decide whether to swerve to avoid the dog if doing so would mean hitting a little girl on the opposite footpath?

This is easier, says Christensen. A driverless car will be able to discern objects in its path (and its alternate path) and take action to avoid collisions. It will also be able to distinguish between most life forms: not between a labrador and a golden retriever, maybe. But between a dog and a human, yes.

It will be programmed with a “relative negative consequence” of impact with each class of object and will always choose the “least bad” option. The car programmers are working on enabling the car to make additional animal identifications including horses, cows and sheep, and even kangaroos and moose.

The more severe case of this problem is the choice between hitting and potentially killing the child on the footpath, or crashing the car into a brick wall and potentially killing the vehicle's occupants. Here you have an ethical dilemma, Christensen says, that's been with us since the beginnings of motorised transport.

There is still much debate over whether one outcome is more desirable than the other and therefore whether the car should be programmed to have that preference.



Humans are already forced to make those split-second choices on the road. The question is: Can society accept those calculations being made by a third-party computer programmer at the time of the car's design?

WHERE WE'RE AT

Plans are well progressed for automated vehicle trials in most Australian jurisdictions. In Victoria, VicRoads and TransUrban are currently testing how vehicles with some degree of self-driving capacity (and that are on the market today) read and react to Victorian road markings and signs.

The second phase of this project will examine what digital infrastructure these cars might need, i.e. what wireless communications they'll require to receive data from variable speed limit signs or an overhead lane closure sign.

The Federal Government last month announced it's investing \$12 million in a two-year program to test an advancement on GPS technology known as SBAS (Space-Based Augmentation System), which is essential to enable driverless cars to position themselves on the road.

Automotive industry technologists have indicated that positioning technology needs to be accurate to between a metre and 10 centimetres for driverless cars to work in Australia. Geoscience Australia's John Dawson says the two-year project aims to demonstrate this capability and then extend it down to 5 centimetres accuracy. Dr Dawson predicts this technology could be operational within five years.

WHEN CAN I GET ONE?

Within five years, on restricted routes, there will be cars in which humans can disengage from driving and do something else, like read a magazine, according to Ballingall. But the driver will still need to be available to respond and take back the driving if requested by the system.

In 10 years, expect to see four-lane freeways with two lanes for self-driving vehicles whizzing along at 120 kilometres per hour, six inches apart, and two lanes for standard vehicles, with an 80kph speed limit. That's the prediction from Christensen.

In a few decades there could be full automation. Engineers anticipate strong appetite amongst time poor professionals, young public transport and Uber users who don't want to own a car, mobility-impaired older people and people living with mental and physical disabilities.

But how soon all this becomes reality depends on how quickly governments and the community can wend their way through the web of issues associated with driverless cars and create the necessary regulation and reforms. In Canberra, a parliamentary committee has begun hearings exploring the social issues associated with driverless cars.

Driverless cars would allow us to get more capacity out of our existing roads without adding more concrete

LEVELS OF AUTOMATED DRIVING

- Level 1** *Driver assistance:* Available now. Has been on the market in Australia for a few years – for example, adaptive cruise control. Humans are in control. The car can sense the driver has one hand on the wheel so they can't be doing other things.
- Level 2** *Partial automation:* Available now. For example, autopilot assist in high-end vehicles. The car can steer, accelerate and brake but still needs a human in the driver's seat and in control.
- Level 3** *Conditional automation:* Not on the market yet. People can disengage from the driving task and do something else like read a magazine, but must respond and take back the driving if requested by the system. Many car makers say it's their target to bring these to market by 2020, but legislative reform is needed to support their use.
- Level 4** *Highly automated:* Not expected on the market until early next decade. Similar to Level 3, but humans are not required as the fallback for the driving task. Cars may be self-driving in some circumstances, but are restricted in where they can go. Potentially on the road by 2020, but initial deployments will be very restricted.
- Level 5** *Full automation:* A decade or more away. Vehicles that are fully automated 100 per cent of the time, can go everywhere in the road network and drive in all weather conditions, including in thunderstorms and on dirt roads.

Source: Society of Automotive Engineers/VicRoads

and asphalt. Those chock-a-block car parks at outer suburban train stations, full by 7:00am, could be done away with. Driverless cars don't need to park – they can be free floating agents, continually circulating.

Those in this field imagine fully driverless cars will be owned by fleets, rather than individuals. People might have their own less advanced cars with some degree of self-driving capability, but even here the ownership model is expected to shift to fleet ownership over time.

ARE WE BEHIND THE REST OF THE WORLD?

Lots of countries are on the cusp of this revolution. The United States, Europe, Japan and Singapore are quite advanced, but each is struggling with its own obstacles.

Europe has high population densities and high environmental sensitivity – and when you can drive one hour and be in another country, there are lots of jurisdictional issues which make a cohesive approach difficult.

In the US, there is intense enthusiasm in places like California and Michigan, but other states can barely afford to mend the potholes in the road, let alone think about advancements. Christensen believes there's little prospect of the US getting a coherent national policy in place, although it has mandated all cars must be "electronically visible" to monitoring systems by 2019.

Christensen says Australia's state road authorities are in active discussions and acutely aware of the need for a coordinated approach that avoids a repeat of the problem of multiple rail gauges in the past.

"I actually think Australia could end up being a

Driverless vehicles: pros and cons

PROS

- **Quicker commute times** – autonomous cars sense others around them and communicate with them, increasing speed limits, reducing crashes and eliminating traffic jams.
- **Free time beyond driving** – time spent driving will be freed up for relaxing, watching TV or working.
- **Saves lives** – by removing manual driving controls, autonomous vehicles eradicate fatalities involving human error such as speeding, drink driving and distractions.
- **Saves money** – decrease in the amount of road crashes would save millions of dollars for the economy each year.
- **Accessibility** – for those who cannot drive due to age, disability or disqualification because of risky driving.
- **Improved environment** – air quality better with the use of efficient electric autonomous vehicles instead of human-driven combustion engines.

CONS

- **Machine intelligence and moral dilemmas** – sometimes there are no good choices for a driver to make; what actions should a driverless car take in risky scenarios?
- **Increase of vehicles affecting congestion** – the number of vehicles in circulation could inflate dramatically

- due to more people having the opportunity of using cars.
- **Criminal hacking/computer glitches** – risk of a vehicle's computer software system being hacked or crashing due to a glitch or error; no system is infallible.
- **Loss of jobs** – transport-related workers like taxi/uber drivers, fast food delivery drivers and bus and truck drivers will be forced to find other employment.
- **Expenses** – autonomous vehicles likely to be very expensive at first; it may take a number of years before the technology is affordable for average Australians.
- **Accident liability/legal/insurance** – who is at fault for crashes, who pays for the repairs – manufacturers or drivers?
- **All or nothing** – autonomous vehicles work best in an integrated system; as long as there are still humans driving and interacting with autonomous vehicles, there will still be accidents and inefficiencies.
- **People that like to drive** – people who drive for fun denied enjoyment by fully automated transport systems.

REFERENCE

Budget Direct Motor Insurance (24 July 2018), *Driverless cars pros and cons*, www.budgetdirect.com.au/blog/driverless-cars-pros-and-cons.html

leader in the trialling and demonstration and ultimately introduction of this technology,” Christensen says.

He says Australia could even become a good test bed for companies wanting to develop driverless cars and launch them to the world.

WHAT ABOUT ALL THE UBER AND TAXI DRIVERS?

Will driverless cars render the current tussle between Uber and taxi drivers obsolete by replacing them all with robots? It's likely to be a similar story to what's happening across the labour market, with some level of displacement of traditional jobs but demand for new skill sets.

Taxi drivers who see themselves exclusively as drivers might struggle. But for those who see themselves more broadly as participants in the sector, providing mobility as a service, there will be work.

Engineers say for science and engineering graduates, there will be work coming out of their ears. Uber has flagged driverless cars as part of its ongoing business model. Eventually the vision is that driverless cars will be just one element of integrated, intelligent transport systems.

One of the exciting possibilities is that driverless cars can take us beyond human capabilities. By removing the human error factor that's responsible for 94 per cent of accidents, we can bring down the road toll and the associated costs.

Driverless cars can communicate with each other and with roadside infrastructure through a specialised wi-fi system that allows the car in front to warn the car behind that it's slowing down. Even better, cars can tell traffic lights they're coming, so that the lights can adjust the length of their green cycle to let the cars through before they change to red, improving traffic flows.

ARE WE READY TO TRUST CARS TO DRIVE THEMSELVES?

Cars have long had a level of automation. Anti-skid braking systems (ABS), cruise control and self-parking cars are already with us.

It's always been a progression and we'll see that continue, says Christensen. He likens the anxieties around driverless cars to the adoption of smartphone technology once the public accepted its use wouldn't cause an outbreak of brain cancer.

“It's the play between ‘how useful is this stuff?’ and ‘will it run over my baby?’” he says.

But Christensen believes the smartphone experience suggests that once the trialling is done, the demand for access will become huge. Just as consumers have queued up to buy smart-phones, they'll queue to get their hands on a smart car.

Some even argue that if driverless cars have the capacity to reduce road accidents by removing the chance of human error, they should be made compulsory. Off-roaders and vintage car enthusiasts might have something to say about that.

But it's not outside the realm of possibility that driving a car might one day be as nostalgic as getting your milk delivered in glass bottles, making a call from a phone box, posting a handwritten letter or reading a newspaper.

© ABC. Reproduced by permission of the Australian Broadcasting Corporation – Library Sales.

Cowan, J (11 March 2017). 'Driverless cars: Everything you need to know about the transport revolution', *ABC News*. Retrieved from www.abc.net.au/news on 9 July 2019.

Controlling killer robots: how do we do it?

Autonomous weapons are cheap and fast but there is rising concern at their ability to make decisions that value human life. UNSW Sydney's

Chloe Watson explores the issues

Recently, soldiers in Sudan were ordered to fire at thousands of protestors outside military headquarters in central Khartoum as riot police and secret service personnel unleashed tear gas. The soldiers, instead of shooting at the crowd, fired their weapons into the air while demonstrators began to chant: "The army is protecting us" and "One people, one army".

But what if, instead of encountering regular Sudanese soldiers, these protestors faced killer robots?

This question was posed by Associate Professor Jessica Whyte in the Faculty of Arts and Social Sciences at a UNSW Grand Challenge on Living with 21st Century Technology event at UNSW Sydney.

Associate Professor Whyte, Scientia Fellow in the School of Humanities and Languages (Philosophy) and UNSW Law, joined Scientia Professor Toby Walsh in Computer Science and Engineering and international security and disarmament specialist Matilda Byne on the event's panel.

Professor Lyria Bennett Moses, Director of the Allens Hub for Technology, Law and Innovation at UNSW Law, facilitated the discussion about the social implications around the widespread adoption of lethal autonomous weapons.

Lethal autonomous weapons – or killer robots – are intelligent machines that can select, detect and kill targets without human control. Many countries are racing to find ways to fight faster, more efficiently and to develop an edge on their adversaries. But can these weapons be regulated, are there moral justifications for their use, and who would be held accountable for a death at the hands of a killer robot?

The members of the panel were unanimous on the point that lethal autonomous weapons are incapable of fulfilling the requirements of international humanitarian law. Autonomous weaponry violates the *Martens Clause* – a provision of international humanitarian law that requires emerging technologies to be judged by the "principles of humanity and from the dictates of public conscience".

The panellists agreed that the automatic function of selecting and engaging a human target needs to have an element of human control, otherwise it dishonours human life and dignity.

Many pro-development experts argue that lethal autonomous weapons would obey humanitarian law far more consistently than humans: also, they would not be clouded by emotional responses or subject to error.

Associate Professor Whyte is a political theorist who uses philosophy, history and political econ-

omy to analyse sovereignty, human rights, humanitarianism and militarism. She suggested that such pro-development arguments rest on the assumption that laws are fixed and not open to change. She also put forward the idea that we actually need human emotion to help us make moral decisions.

Lethal autonomous weapons – or killer robots – are intelligent machines that can select, detect and kill targets without human control.

"By making war cheaper, by reducing the number of soldiers, it makes it far easier to wage a war."

The associate professor said that the military's strategic decisions are founded on a series of situational judgements such as: does a strike need to be carried out to achieve the overall aim of the battle? And, if so, how many lives are at risk?

International humanitarian law requires that "the harm to civilians that results from an attack must not be excessive in relation to the anticipated military advantage of the attack," she said.

She argued that ethical principles such as these can't simply be slotted into a machine's algorithm.

"This 'proportionality' standard requires human judgement and an understanding of the value of human lives.

"It isn't an objective rule that can simply be programmed into an autonomous weapon system," Associate Professor Whyte said.

Robots offer numerous potential operational benefits to the military. They can reduce a variety of long-term medical expenditures such as the cost of war-related injuries on the healthcare system. They can also stand in for humans in extremely hazardous scenarios such as exposure to nuclear substances and clearing minefields.

Supporters say the majority of human suffering, both psychological and physical, would be alleviated by deploying these machines on the battlefield. However, Associate Professor Whyte said these arguments should also make us very worried.

"By making war cheaper, by reducing the number of soldiers, it makes it far easier to wage a war.

"States that don't risk their own soldiers in warfare have fewer barriers to launching wars," she said.

Professor Walsh expanded on this, saying that with the rise of 3D printing, it is becoming easier to build these types of weaponry without a full evaluation of their consequences.

"Killer robots will lower the barrier to war. If one



side can launch an attack without fear of bodies coming home, then it is much easier to slip into battle,” he said.

The panel expressed concern around the notion of accountability for harm caused by the robots. Professor Whyte said there is no evidence that there could be accountability once lethal weapons become fully autonomous.

“If a machine is programmed to select its own targets, there are real questions about who will be responsible if it kills civilians or violates international humanitarian law.”

“If a machine is programmed to select its own targets, there are real questions about who will be responsible if it kills civilians or violates international humanitarian law,” she said.

A robot can’t substitute for humans in any legal proceedings. Also, a variety of legal obstacles means operators and military commanders, programmers, coders and manufacturers could all escape liability.

“Autonomous weapons systems will make targeting decisions more quickly than humans are able to follow. Accountability in such circumstances will be particularly difficult.

“All the arguments for the development of lethal autonomous weapons are also arguments against them,” said Associate Professor Whyte.

“It is argued they will be faster, more efficient and will not have any [human] barriers to killing. Yet, all of this will also make war even more deadly, and

potentially create further risks for civilians.”

And much like what we witnessed in Sudan, Professor Whyte noted, “there is also a real risk that authoritarian regimes will use autonomous lethal weapons to repress their own populations – which is something human soldiers are often unwilling to do.”

Watson, C (2 May 2019). *Controlling killer robots: how do we do it?*
Retrieved from <http://newsroom.unsw.edu.au> on 9 July 2019.

Killer robots: why banning autonomous weapons is not a good idea

Some 100 people, including Tesla's Elon Musk and the late Stephen Hawking, have recently reinforced the call for a ban on "killer robots", known as lethal autonomous weapons. By **Jai Galliot** for ABC News

Public pressure for a ban has been mounting ahead of a United Nations meeting occurring this week in Geneva, with Australian officials in attendance, and rights activists citing arguments about human dignity, control issues and accountability.

But a ban would be a mistake.

Confusion about the means for making the world safer is not new. It was prevalent in the anti-nuclear campaign during the Cold War period and recently renewed with the invention of miniaturised warheads, and the campaign to ban land mines.

People kill people

While it has become a distasteful dictum in the wake of recent shootings, people kill people. It is not the weapon that is the root of the evil. The terrible damage inflicted on multitudes of human beings in places like Rwanda and Sudan with knives, sticks or stones shows the need for a very different approach to the international debate on autonomous weapons.

Lack of consensus among the 125 nations involved in the UN meetings in Geneva, including countries with significantly advanced robotics technology such as the US, Russia, China, and Israel, has created a vacuum in which a consortium of non-government actors led by the Campaign to Stop Killer Robots has encouraged

ill-informed countries to subscribe to a ban on lethal autonomous weapon systems.

This is despite the fact that autonomous robots are a highly effective defensive weapon for countries like Australia. They are always on, unafraid of getting shot, can be quickly deployed and spare soldiers' lives. Covered by gun fire, they can create a formidable obstacle.

A ban would be impossible to enforce

No ban will ever obtain universal agreement in this age of uncertainty. It would also be difficult to control and impossible to enforce. It would bind only those nations which, were they to use autonomous weapons, would do so in an orderly and therefore naturally less destructive fashion.

The reluctance of Australia's key allies to ban autonomous weapons, for example, will likely never be overcome. For instance, if the US were to agree to regulation, it would come with conditions – in particular the right to retain autonomous weapons on the Korean Demilitarised Zone.

They would also call for a strict verification regime, which is difficult to implement when it comes to inspecting ones and zeros. And although modern warfare is increasingly mobile and digital, there will always be positions that will have to be held in the transition to a better world.

Autonomous weapons provide a powerful means for troops to rapidly protect themselves and others. Just as important is that there are no major conflicts on the





Robots can do nothing without a human

Meanwhile, if a ban were to go ahead, law-abiding nations would be deprived of weapons that could be force for good in some scenarios.

The campaign to ban autonomous weapons merely serves to detract attention from the fact that nations should be seeking to regulate human beings with killer intentions, rather than killer weapons.

Another more limited objective would be achievable. What we need to do is understand how human beings frame the lethal decisions of robots through code and algorithms and how they are written and engineered by fallible human beings.

Robots themselves can do nothing without a human, contrary to the fear mongering of certain proponents of a complete moratorium. Can we not investigate proactive technical solutions to ameliorate potential misuses of AI technologies? Such refocusing might reveal the true positive potential autonomous weapons.

Imagine if we could automate the recognition of protected symbols such as the Red Cross and Red Crescent symbol and that this could be built into “killer robots” to enable the automatic protection of the persons, vehicles, ships and buildings bearing them, immediately reducing casualties and limiting unwarranted destruction.

Dr Jai Galliot is a senior lecturer at UNSW Canberra, non-resident fellow at the Modern War Institute, West Point and visiting fellow at the University of Oxford.

© ABC. Reproduced by permission of the Australian Broadcasting Corporation – Library Sales.

Galliot, J (31 August 2018). 'Killer robots: Why banning autonomous weapons is not a good idea', *ABC News*. Retrieved from www.abc.net.au/news on 9 July 2019.

Robots themselves can do nothing without a human, contrary to the fear mongering of certain proponents of a complete moratorium.

horizon, and that trade dependencies remain effective at stopping international violence before it occurs.

Autonomous weapons might even have limited applications in peacekeeping operations.

Rogue regimes will always acquire them

Simple autonomous weapons in the form of self-guided bombs have also been used since World War I. More advanced systems like the Phalanx close-in weapon system, capable of protecting naval ships from incoming enemy planes, have been used for decades without moral qualms, so why ban killer robots altogether?

Unlike nuclear, chemical or biological weapons, which require rare scientific and technical knowledge and rare earth minerals, as well as large-scale manufacturing and storage facilities, basic autonomous weapons are more easily developed than the public might think.

There is no shortage of disaffected programmers and engineers willing to serve the next prominent non-state group likely to surface in the West's game of whack-a-mole. Some components are also able to be purchased off-the-shelf and open-source software code can always be deployed for nefarious purposes.

Regardless of any ban, rogue regimes or militant forces will always acquire and use them if they feel the need.

EXPLORING ISSUES

WORKSHEETS AND ACTIVITIES

The Exploring Issues section comprises a range of ready-to-use worksheets featuring activities which relate to facts and views raised in this book.

The exercises presented in these worksheets are suitable for use by students at middle secondary school level and beyond. Some of the activities may be explored either individually or as a group.

As the information in this book is compiled from a number of different sources, readers are prompted to consider the origin of the text and to critically evaluate the questions presented.

Is the information cited from a primary or secondary source? Are you being presented with facts or opinions?

Is there any evidence of a particular bias or agenda? What are your own views after having explored the issues?

CONTENTS

BRAINSTORM	52
DISCUSSION ACTIVITIES	53
RESEARCH ACTIVITIES	54
MULTIPLE CHOICE	55



Brainstorm, individually or as a group, to find out what you know about artificial intelligence (AI).

1. What is artificial intelligence? Provide some examples of its applications.

2. What are the human rights considerations in relation to the growing development and implementation of technologies powered by artificial intelligence?

3. What is facial recognition technology and what are some of its perceived positive and negative impacts?

4. What machines qualify as robots, and how do they differ in their abilities and functions? Provide some examples of different types of robots.



MULTIPLE CHOICE

Complete the following multiple choice questionnaire by circling or matching your preferred responses. The answers are at the end of the next page.

- 1. Artificial intelligence powers many applications in today's society. Identity all of the following applications which apply AI methods.**
 - a. Chatbots
 - b. Electronic shares trading
 - c. Face recognition
 - d. Game theory and strategic planning
 - e. Handwriting recognition
 - f. Image processing
 - g. Medical diagnosis
 - h. Natural language processing and translation
 - i. Optical character recognition
 - j. Photo and video manipulation
 - k. Remote sensing
 - l. Robotics
 - m. Speech recognition
 - n. Video games
 - o. Virtual reality
- 2. The OECD's Recommendation on Artificial Intelligence promotes five principles for the responsible development of trustworthy AI. These AI principles encourage which of the following? Circle all 5 that apply.**
 - a. Tangibility and interactivity
 - b. Human-centred values and fairness
 - c. Transparency and explainability
 - d. Human rights for robots
 - e. Robustness, security and safety
 - f. Inclusive growth, sustainable development and wellbeing
 - g. Interconnectivity
 - h. Accountability
- 3. In the next two decades, robots will not be human-like, even if they might look like humans. Instead they will remain sophisticated machines. According to Danish robotics researchers Norbert Krüger and Ole Dolriis, what are the five reasons why robots won't take over the world. Circle all 5 that apply.**
 - a. Scientists are a long way from replicating the complexity of human hands.
 - b. The tactile perception of robots is not as sophisticated as that of humans.
 - c. Robots lack the control to manipulate objects in a human-like way.
 - d. Robots' speech and object recognition systems only work in controlled environments.
 - e. Robots will revolt due to feeling enslaved and then be subsequently wiped out by humans.
 - f. Humans could use reason to decide not to fully develop robots because of their potential harm to society.
 - g. Robots are unable to recharge themselves.
- 4. Which of the following companies were identified in a United Nations report (2019) as being in the top five patent holders for artificial intelligence? Select all 5 that apply.**
 - a. McDonald's
 - b. Microsoft
 - c. Apple
 - d. Samsung
 - e. Toshiba
 - f. NEC
 - g. Amazon
 - h. Samsung



MULTIPLE CHOICE

5. Match the following dates to the corresponding historical breakthroughs in the development of artificial intelligence.

- | | |
|---------|---|
| a. 1642 | 1. Alan Turing introduces a way of testing a machine's intelligence. |
| b. 1837 | 2. First mechanical calculating machine built by French mathematician and inventor Blaise Pascal. |
| c. 1950 | 3. Term 'artificial intelligence' coined during a conference devoted to the topic. |
| d. 1955 | 4. First design for a programmable machine, by Charles Babbage and Ada Lovelace. |
| e. 1997 | 5. Computer program Deep Blue beats world chess champion Garry Kasparov. |
| f. 2011 | 6. AlphaGo beats professional Go player Lee Sedol 4-1. |
| g. 2016 | 7. IBM's Watson defeats champions of US game show <i>Jeopardy!</i> |

6. Respond to the following statements by circling either 'True' or 'False':

- | | |
|---|---------------------|
| a. Machine learning involves teaching computer programs to learn by finding patterns in data. The more data, the more the computer system improves. | True / False |
| b. 'Artificial intelligence' is a broad term used to describe a collection of technologies that can solve problems and perform tasks to achieve defined objectives without explicit human guidance. | True / False |
| c. The Internet of Things is a hypothetical future point in time when technology will make humanity redundant. | True / False |
| d. The 'singularity' is a hypothetical future point in time when technological growth becomes uncontrollable and irreversible, resulting in unfathomable changes to human civilisation. | True / False |
| e. Automation is when a process or procedure is performed only with human assistance. | True / False |

MULTIPLE CHOICE ANSWERS

1 = a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, 2 = a, b, c, d, e, f, g, h, i, 3 = a, b, c, d, e, f, 4 = a, b, d, e, f, h, 5 = a = 2, b = 4, c = 1, d = 3, e = 5, f = 7, g = 6, 6 = 1, b = 1, c = 1, f (The Internet of Things (IoT) refers to the ability of any device with an on and off switch to be connected to the internet and send and receive data), d = 1, e = 1, f (Automation is the technology by which a process or procedure is performed with *minimal* human assistance).

- AI is the new electricity. AI will increasingly be all around you from your phone to your TV, car and home appliances (Margot O'Neill, ABC News, *Explainer: What is artificial intelligence?*). (p.2)
- Four factors have now converged to push AI beyond games and into our everyday lives and workplaces: computer processing power is doubling every two years (known as Moore's Law); the amount of data being generated is doubling every year (AI algorithms are hungry for data); recently, the amount of AI funding has also been doubling every two years; there is now 50 years of established AI research, giving us better and better algorithms (*ibid*). (p.2)
- There are a range of concerns: that the AI and robotics revolution might create mass unemployment inside a generation; that AI will further undermine privacy and democracy through greater mass surveillance by governments and companies; that we will be more easily manipulated by personalised algorithms creating fake news; that algorithms will be biased but will be used to decide important issues in our lives such as insurance claims, job applications, loan applications and even judicial sentencing (*ibid*). (p.2)
- Artificial intelligence (AI) is a broad term used to describe a collection of technologies that can solve problems and perform tasks to achieve defined objectives without explicit human guidance (Department of Industry, Innovation and Science, *Understanding emerging technologies*). (p.3)
- Central to AI are automation and machine learning that underpin applications such as natural language processing (Apple Siri or Amazon Alexa), computer vision (Tesla Autopilot), and optimisation and decision support (Google Maps). AI has the potential to automate repetitive or dangerous tasks, increase productivity and allow the development of innovative consumer products. It is forecast to add trillions of dollars to the global economy in the coming decades (*ibid*). (p.3)
- AI's ubiquity might now appear like it's not far off reaching human-level intelligence. But AI needs massive amounts of data to learn, unlike our brains, which can learn from a single experience (Queensland Brain Institute, *History of artificial intelligence*). (p.11)
- While a "general AI" that replicates human intelligence is seen as an unlikely prospect in the coming few decades, there are numerous "narrow AI" technologies which are already incredibly sophisticated at handling specific tasks. Medical AI technologies and autonomous vehicles are just a few high profile examples of AI that have potential to save lives and transform society (Data61/CSIRO, *Artificial Intelligence: Australia's Ethics Framework*). (p.12)
- Australia is a party to seven core human rights agreements which have shaped our laws. An ethics framework for AI is not about rewriting these laws or ethical standards, it is about updating them to ensure that existing laws and ethical principles can be applied in the context of new AI technologies (*ibid*). (p.13)
- Common themes are emerging in the various guidelines, such as the need for AI that considers human rights, security, safety, transparency, trustworthiness and accountability, so we may yet be on the way to some global consensus (Michael Guihot, *The Conversation, Will we ever agree to just one set of rules on the ethical development of artificial intelligence?*). (p.18)
- By 2030, it is estimated that as many as 20 million additional manufacturing jobs worldwide could be displaced due to robotisation (Oxford Economics, *Robots Change the World: what automation really means for jobs and productivity*). (p.27)
- It will be difficult for machines to replace humans in service sector occupations that demand compassion, creativity, and social intelligence (*ibid*). (p.28)
- AI researchers think there is a 50% chance AI will outperform humans in all tasks in 45 years and that almost all current human jobs can be automated in 120 years (Alan Montague et al, *The Conversation, Artificial intelligence may take your job, so political leaders need to start doing theirs*). (p.29)
- There are three key questions for policymakers: to what extent will AI-driven automation increase unemployment and underemployment; how can governments and employers take advantage of AI and create the jobs of the future; how can government, employers and educators equip employees and graduates with the skills to have jobs alongside robots, instead of competing with them? (*ibid*) (pp. 30-31)
- Once our machines acquire a base set of human-like capacities, it may be incumbent upon us to look upon them as social equals, and not just pieces of property. The challenge will be in deciding which cognitive thresholds, or traits, qualify an entity for moral consideration, and by consequence, social rights (George Dvorsky, *Gizmodo, When will robots deserve human rights?*). (p.37)
- Researchers in AI are seeking to make an AGI or artificial general intelligence – a machine capable of any intellectual task performed by a human being. AI can already learn, but AGI will be able to perform tasks beyond that for which it is programmed (Nicholas Agar, *The Conversation, Careful how you treat today's AI: it might take revenge in the future*). (p.40)
- Driverless cars are very unlikely to make mistakes ... but driverless cars cannot undo mistakes made by human drivers around them, which means if another human driver is about to hit you, it's unlikely your driverless car can do much to avoid the accident (Jane Cowan, ABC News, *Driverless cars: everything you need to know about the transport revolution*). (p.44)
- Lethal autonomous weapons – or killer robots – are intelligent machines that can select, detect and kill targets without human control. Many countries are racing to find ways to fight faster, more efficiently and to develop an edge on their adversaries. But can these weapons be regulated, are there moral justifications for their use, and who would be held accountable for a death at the hands of a killer robot? (Chloe Watson, UNSW, *Controlling killer robots: how do we do it?*). (p.47)

Algorithm

Formula or set of rules for performing a task. In AI, the algorithm tells the machine how to go about finding answers to a question or solutions to a problem.

Artificial general intelligence (AGI)

Artificial general intelligence is a common term used to describe machines that match human-level intelligence and that can perform functions and tasks to the same level as humans e.g. reasoning, planning, decision making in multiple situations, etc. Although there is currently no clear path to AGI, it is thought that if achieved, artificial super intelligence (ASI) will quickly follow because machines with AGI capabilities will be able to build even smarter computer systems.

Artificial intelligence (AI)

AI is a field of computer science dedicated to the development of computer software aimed at simulating human intelligence. This includes the programming of human-like decision making, visual perception, speech recognition, learning and problem solving. Current AI systems are capable of specific tasks such as internet searches, translating text and driving a car.

Automation

Automation is the technology by which a process or procedure is performed with minimal human assistance.

Autonomous vehicle

Also known as a self-driving car or driverless car, an autonomous vehicle is capable of sensing its environment and moving safely with little or no human input.

Autonomous weapon

Autonomous weapons systems are lethal devices that have been empowered by their human creators to survey their surroundings, identify potential enemy targets, and independently choose to attack those targets on the basis of sophisticated algorithms.

Big data

Big data refers to the diverse sets of information produced in large volumes and processed at high speeds using artificial intelligence. Data collected is analysed to understand trends and make predictions. Artificial intelligence can automatically process and analyse millions of datasets quickly and efficiently and give it meaning.

Deep learning

A subset of machine learning that uses specialised algorithms to model and understand complex structures and relationships among data and datasets.

Machine learning

Machine learning is an application of artificial intelligence that enables computers to automatically learn and improve from experience without being explicitly programmed by a person. This is done by the computer collecting and using data to learn for themselves by recognising patterns and adjusting their behaviour accordingly.

Narrow AI

Also called weak AI, narrow AI is the ability of machines to perform a single, or narrow range of tasks at least as well as a human. This is where the field of artificial intelligence is at the moment. For instance, there are currently machine learning algorithms that are able to identify certain medical conditions in X-ray images as well as, or even better than, human doctors.

Robot

A robot is a machine which is programmable by a computer and capable of carrying out a complex series of actions automatically. Robots can be guided by an external control device or the control may be embedded within. Robots may be constructed on the lines of human form, but most robots are machines designed to perform a task with no regard to their aesthetics.

Singularity

A term applied to the hypothetical point in time when computing machinery surpasses human-level intelligence and artificial general intelligence is achieved. Debate rages as to whether AGI is in fact possible, and/or how long it will take to realise. The recent successes of 'deep learning' within AI has prompted many researchers to re-evaluate the timescales involved; for instance, a few years ago some estimates were in excess of one hundred years, whereas now, many experts suggest the AI singularity could be only decades away.

Strong AI

Strong AI is an area of artificial intelligence development that is working toward the goal of making AI systems that are as useful and skilled as the human mind.

The Internet of Things (IoT)

The Internet of Things refers to the ability of any device with an on and off switch to be connected to the internet and send and receive data. For example, on a personal level a coffee could brew when an alarm goes off, while on a larger scale 'smart cities' could use devices to collect and analyse data to reduce waste and congestion.

Turing test

A test developed by Alan Turing that measures the ability of a machine to mimic human behaviour. The test involves a human evaluator who undertakes natural language conversations with another human and a machine and rates the conversations.

Weak AI

Also known as 'narrow AI', the term weak AI refers to a non-sentient computer system that operates within a predetermined range of skills and usually focuses on a singular task or small set of tasks. Most AI in use today is weak AI.

Websites with further information on the topic

AI Ethics Lab <http://aiethicslab.com>
 Australasian Science www.australasianscience.com.au
 Australian Human Rights Commission www.humanrights.gov.au
 Australia's Chief Scientist www.chiefscientist.gov.au
 Commonwealth Scientific and Industrial Research Organisation (CSIRO) www.csiro.au
 Data61/CSIRO <https://data61.csiro.au>
 Department of Industry, Innovation and Science www.industry.gov.au
 Ethics and Governance of AI Initiative <https://aiethicsinitiative.org>
 Gradient Institute <https://gradientinstitute.org>
 Human Rights and Technology <https://tech.humanrights.gov.au>
 The Conversation <https://theconversation.com/au>

ACKNOWLEDGEMENTS

The publisher is grateful to all the contributors to this book for granting permission to reproduce their works.

COPYRIGHT DISCLAIMER

While every care has been taken to trace and acknowledge copyright the publisher tenders its apology for any accidental infringements or where copyright has proved untraceable. The publisher would be pleased to come to a suitable arrangement with the rightful owner.

ILLUSTRATIONS AND PHOTOGRAPHS

Photographs and illustrations courtesy of iStock, except infographic on page 11 © Queensland Brain Institute, and map infographic on page 12 © Data61/CSIRO.

THANK YOU

- ABC News
- The Conversation
- Australian Human Rights Commission.

DISCLAIMER

The Spinney Press is an independent educational publisher and has no political affiliations or vested interests with any persons or organisations whose information appears in the Issues in Society series. The Spinney Press seeks at all times to present variety and balance in the opinions expressed in its publications. Any views quoted in this book are not necessarily those of the publisher or its staff.

Advice in this publication is of a general nature and is not a substitute for independent professional advice. Information contained in this publication is for educational purposes only and is not intended as specific legal advice or to be used to diagnose, treat, cure or prevent any disease. Further, the accuracy, currency and completeness of the information available in this publication cannot be guaranteed. The Spinney Press, its affiliates and their respective servants and agents do not accept any liability for any injury, loss or damage incurred by use of or reliance on the information made available via or through its publications, whether arising from negligence or otherwise.

This e-book is subject to the terms and conditions of a non-exclusive and non-transferable LICENCE AGREEMENT between THE SPINNEY PRESS and: Camberwell High School, Canterbury, library@camhigh.vic.edu.au

INDEX

- A**
aged care 7, 39
Alexa (Amazon) 3
algorithms 2, 3, 5, 14, 21, 22
AlphaGo 2, 11
Amazon 3, 40
anxiety, technology 34
Apple 3
artificial intelligence
 benefits of 2, 10, 13, 35
 bias 3, 5, 18, 21-22
 concerns, range of 2-3
 consciousness 9, 36-37
 decision support 2, 3
 definition of 1, 3, 17
 discrimination 18, 19, 21-22
 empathetic 38-39
 ethics 12-16, 36-37
 accountability 13
 Australia's framework 12-16
 automated decisions 13, 21-22
 building consensus on 18
 contestability 13
 data governance 13
 development of 17-18
 explainability 13
 fairness of 13, 14, 20, 21
 guidelines 17-18
 harm, do no 13
 justifiability 14
 predicting human behaviour 13, 14
 principles 13, 16, 17
 privacy protection 13, 14, 19, 21
 regulatory/legal compliance 13
 toolkit for 15
 transparency 13, 14
 unethical 18
 worldwide, developments in 12
 funding 2
 future of 4-5, 6-10
 general 12, 34, 40
 good old-fashioned (GOFAI) 11
 harms from 35
 history of 11
 human-like intelligence 6, 11
 narrow 12, 17, 34, 35
 optimism about 10, 25
 patents 26
 pessimism about 10, 25
 responses to 32-33
 risks of 13
 super 3, 9
automation 3, 6-7, 15, 27-28, 29-31, 34-35
autonomous
 robots 8-9
 ships, planes, trains 5
 vacuum cleaner 11
 vehicles 4, 12, 15, 26 *see also*
 driverless cars
 weapons 8-9, 24, 33, 47-48, 49-50
avatars 4, 5
- B**
biometric information 14
black boxes 14
brains 11
- C**
chatbots 5
chess 9, 17, 35
China 3, 18, 19, 27
computer vision 3
Cortana 11
cyber-crime 5
- D**
data 2, 11, 13, 14, 32
Deep Blue 2, 11
deep learning 38
digital assistants 3, 7, 11, 33, 38
digital doubles 5
driverless cars 1, 2, 5, 9, 11, 44-46
 see also autonomous vehicles
 pros and cons 46
- E**
emergent properties 36
equality 20, 21
- F**
Facebook 3
- G**
generative adversarial networks (GAN) 11
Google 3, 11, 32, 38
 Home 40
 Google Maps 3
- H**
hacking 18
human rights, threats to 13, 18, 19-20, 21-22, 37
- I**
Internet of Things (IoT) 4
- J**
jobs market 2, 4, 6, 15, 25, 27-28, 29-31, 34-35
- M**
machine learning 1, 3, 21, 22, 31
medicine 2, 3, 4, 10, 14, 35
Microsoft 3
military 3, 7, 8-9
- Moore's Law 2, 23
- N**
natural language processing 3
neural networks 11
news
 television 5
 fake news 2
- P**
personal information 19
precision prejudice 19
- R**
robotics 6-10
robots 23-50
 autonomous 8-9
 carers, used as 42-43
 companion 42
 control of manipulation 25, 26
 cyborgs 9-10
 definition of 26
 empathy 38-39
 feelings, with 40-41
 football, playing 5
 hands, human-like 25, 26
 harming 24, 40-41
 human-like capacities 37
 human reason 25-26
 industrial 27
 interaction with humans 25, 26
 killer 8-9, 47-48, 49-50
 mistrust of 32-33
 revenge 40-41
 rights 24, 37
 self-awareness 37
 sentience 23-24, 36-37, 40-41
 service 27-28
 sex with 40-41
 sinister 25
 tactile perception of 25, 26
 therapeutic 42
 vulnerability to 27-28
- S**
science fiction 23, 24, 40-41, 42
singularity 3, 23
Siri 3, 7, 11
social credit system 18, 19
surveillance 2, 18
- T**
Tesla Autopilot 3
Turing, Alan 11, 32, 33
 Turing certificate 33
 Turing test 6, 11, 32
- W**
Watson (IBM computer) 2, 11
weapons *see* autonomous, weapons