

Human Rights and Technology Consultation

02 October 2018

About the Warren Centre for Advanced Engineering

The Warren Centre brings industry, government, and academia together to create thought leadership in engineering, technology, and innovation. We constantly challenge economic, legal, environmental, social, and political paradigms to open possibilities for innovation and technology to build a better future.

The Warren Centre advocates for the importance of science, technology and innovation. Our 30 years' experience of leading the conversation through projects, promotion, and independent advice drives Australian entrepreneurship and economic growth.

This submission is our response to the Human Rights and Technology Issues Paper distributed by the Australian Human Rights Commission. The specific consultation questions addressed are 5, 6 and 7 related to AI-informed decision making.

Executive Summary

The Warren Centre observes that benefits and harms have been produced by artificial intelligence and machine learning technologies as those technologies inform decisions. We believe that very strong public benefits in economic growth and productivity improvements will be realised globally, especially through enhanced health outcomes from AI. It is important that machine learning and artificial intelligence continue to develop. However, attention must be applied to mitigate the negative effects of AI.

We offer the following recommendations:

1. Additional research should be undertaken in the development of "ethical" frameworks for machine learning (ML) and artificial intelligence (AI). In some critical cases, the safety and efficacy of AI algorithms may need to be verified.

The Warren Centre for Advanced Engineering Ltd

School of Information Technologies Building J12
University of Sydney NSW 2006
T: (02) 9351 3752 F: (02) 9351 2012

E: warrenc@sydney.edu.au W: thewarrencentre.org.au
twitter.com/thewarrencentre facebook.com/thewarrencentre



2. Efforts should be undertaken to open multi-disciplinary engagement and discourse on AI. This should include IT developers, engineers, lawyers, ethicists, government, industry, academia and the public.
3. Greater public education and awareness are needed to understand the real benefits, risks and issues related to AI. Good public awareness informs better democracy and support for rational public policy.
4. Ethical AI guidelines should be developed based on Australian values of privacy, fairness and transparency.
5. Australia should establish and lead international dialogues on AI and robotics ethics.
6. Enhanced STEM education along with diversity and inclusivity efforts are needed to promote domestic talent development and to minimise the potential for distorting effects of bias in machine learning and AI algorithms.

1. Current Situation

1.1 AI is advancing rapidly

In recent years, technology advances in sensors, digital commercial transactions, massive data accumulation, machine learning, artificial intelligence and sophisticated robotic actuation have created remarkable innovations and commercialisation demonstrations. There is very strong potential for continued improvement in economic productivity and advancement in human health from these technologies. There is concern for employment displacement, and it is likely that the rapid advancement of technologies will disrupt industries and jobs more quickly than some people can retrain and adapt. The interim technology disruption will create a period of social discontent as employment conditions adjust. Productivity gains will create wealth that is not evenly distributed. First movers in markets are already creating new platforms with radically improved services and product offerings that disrupt established companies. Globalisation in the past three decades has already exposed areas of Australia's economy where international competition effectively displaces local economic activity.

Although there is not perfect agreement in technical communities on the definition of artificial intelligence, we agree with the AHRC's suggested definition of 'narrow AI'. Applying that definition, there are many examples of machine learning and narrow artificial intelligence operating today in Australia.

1.2 AI is imported to Australia

Examples such as Google Maps have strong domestic Australian input into the development of the technology, but many of the AI systems currently operating in

Australia were imported from the United States. Due to the scale of the economy and size of the population, there are distinct advantages for scaling up AI systems in the US and China. While Europe is also a large market with significant technological prowess and market capital, to date, it is American and Chinese firms who have leapt forward with the 'killer apps' that are consumer-facing. Facebook, Amazon, Apple, Google and Uber are building significant platforms for economic activity. Sometimes, these platforms are monopolies or oligopolies. Both the ML/AI-based business methods and the ML/AI algorithms are being tested overseas and imported to Australia. In China, TenCent, Alibaba, Baidu and Didi Chuxing are building similar models to their US counterparts and rapidly achieving great scale of user adoption.

Potential problems that are observed overseas are likely to appear in Australia.

2. Several problems already observed or likely to appear

Problem 2.1: Bias in underlying historical data

Machine learning and artificial intelligence rely on a base of historical data. In numerous examples, it has been demonstrated that empirical data contains bias. Banking and lending data, criminal justice data, and insurance data have been shown to contain bias that harms the public. Bias inherent in historical data may not be easily detected. These problems are highlighted in the Issues Paper.

Problem 2.2: Bias programmed into technology by biased humans

Section 4.2 of the Issues Paper describes sexism and LGBTI cyber abuse in online forums and Section 7 of the Issues Paper describes the need to consider diversity and inclusivity when designing technology. However, programming communities show evidence of strong sexism and cultural bias.¹ Programming in Silicon Valley can be so dominated by young white males that the term "brogramming" is applied by some commentators. Exclusion of females and severe problems of cyber bullying in the industry itself highlight a dysfunctional inclusivity gap. These problems are said to exist also in the Australian tech industry.² The issue of stubborn resistance within the industry to reform was highlighted in 2017 when a Google employee issued an essay against diversity titled, "Google's Ideological Echo Chamber".³

The result of this inclusivity gap affects not just the tech industry workplace but also the

¹ Claire McCullough, Svetlana Chesser and Bart Weathington, "Subtle and Not-So-Subtle Messages of Non-Inclusion", American Society for Engineering Education, 2017 ASEE Zone II Conference, at <http://zone2.asee.org/sessions/program/3/13.pdf>

² <https://www.businessinsider.com.au/sexism-in-australian-tech-culture-joint-statement-2017-7>

³ Lara Williams, "No way up this ladder", *New Scientist*, Volume 235, Issue 3139, 19 August 2017, Pages 22-23.

ubiquitous products of the tech industry. How software is programmed may deeply affect the diverse range of human users who directly use the software and also those who are indirectly “handled” by the software: everyone in modern society.

Problem 2.3: Economic factors affect governance

It is a principle in liberal economies that markets should not be unnecessarily regulated unless there is a clear defect or mischief. The scope of this consultation is human rights, not market operation, but it is worth reviewing some of the economic factors that interact with individual rights. Market failures exist in the rapid expansion of data-rich technology companies. Data feeds machine learning, and machine learning informs artificial intelligence algorithms. The market failures include asymmetrical information, anti-competitive behaviour, predatory pricing and unequal bargaining power. These factors interact with the human rights issues raised and exacerbate problems related to bias and fairness. An AHRC consultation should not ignore factors that would fall before the Australian Competition and Consumer Commission.

Governance problems in the current generation of tech companies and platforms are chronicled in an exposé published in 2017 by Brad Stone.⁴ Two quotes from Stone’s book demonstrate the scale of the governance challenges:

“We are living in an era of robber barons. If you have enough money and can make the right phone call, you can disregard whatever rules are in place and then use that as a way of getting PR. And you can win.”

Both were unleashing changes in communities’ behavior whose full impact on society they couldn’t possibly hope to understand. And each believed that the best tactic was simply to grow, harnessing the political influence of their user base to become too big to regulate.

This year, Apple and Amazon each surpassed US\$1 trillion market capital. Although these companies are comparatively mature, some of the rising technology entrants are attracting billions of dollars of venture capital to fuel the race “from zero to one”, i.e., to achieve a monopoly. The future financial rewards are massive, and the immediate investments are very substantial— multiple billions of dollars of cash. In some cases, the competition is a winner-take-all monopoly. In such a hyperbolic, frenzied environment, there may be little incentive to slow down and consider fairness or unintended harmful consequences. These are fiercely competitive economic races to

⁴ Brad Stone, *The Upstarts: How Uber, Airbnb, and the Killer Companies of the New Silicon Valley Are Changing the World*, 2017.

establish and dominate new markets. In the case of rapid races in development, technologists can feel pressure to cut corners, to take risks and to ignore near misses.⁵ In these cases, the economic and governance factors interact with the human rights issues in this consultation.

Problem 2.4: Asymmetry in know-how, hardware and data

In the US bricks-and-mortar retail sector, an atmosphere of extreme competition has spawned a disturbing use of electronic surveillance techniques in shopping malls and street level stores. UPenn Professor Joseph Turow catalogues some of the strange scenarios in *The Aisles Have Eyes: How Retailers Track Your Shopping, Strip Your Privacy, and Define Your Power*.⁶ Security camera technologies originally deployed to document and discourage shoplifting are being repurposed with facial recognition technology to monitor shoppers' purchasing habits and to accumulate massive amounts of data on consumers. Combined with the 'digital vapour trail' of personal devices, retailers can employ sophisticated tracking techniques using Bluetooth, Wi-Fi, and GPS along with the retailers' own electronic payment systems to collect massive amounts of data. Information from various sources can be synthesised in ways that are not understood by average people. Data collection across multiple devices and platforms is integrated across consumer-owned electronics and store-owned fixed hardware. Privacy policies that are obscure to buyers provide insufficient notice to persons who do not understand the broad and deep level of their personal interaction with data collection, machine learning and artificial intelligence.

Problem 2.5: Experimenting with people's private data

One example of the mischief in the area of data acquisition, machine learning and AI is the 2014 case of OKCupid.⁷ The online dating app gathered information from users to match couples for dates. In an online blog, OKCupid co-founder Christian Rudder proclaimed, "We experiment on humans!"⁸ In some cases, OKCupid deliberately matched the worst fitted partners to test if the normal dating match algorithm worked. Shallow answers ensued when the company was challenged about whether the practice was ethical. Subsequent third-party publication and analysis of

⁵ Armstrong, S. & Bostrom, N. & Shulman, C. (2013): "Racing to the precipice: a model of artificial intelligence development", Technical Report #2013-1, Future of Humanity Institute, Oxford University: pp. 1-8.

⁶ Joseph Turow, *The Aisles Have Eyes: How Retailers Track Your Shopping, Strip Your Privacy, and Define Your Power*, 2017 at <https://yalebooks.yale.edu/book/9780300212198/aisles-have-eyes>

⁷ Jeff Bercovici, *OkCupid's Christian Rudder On Human Experiments And Getting Ugly People Dates*, 14 Sep 2014, Forbes online at <https://www.forbes.com/sites/jeffbercovici/2014/09/09/okcupids-christian-rudder-on-human-experiments-and-getting-ugly-people-dates/#7ca259a920da>

⁸ Alex Hern, "OKCupid experiments are standard 'scientific methods', says founder", The Guardian, 04Aug2014, online at <https://www.theguardian.com/technology/2014/aug/04/okcupid-dating-ethics-facebook-experiments>.

OKCupid data by academics in Europe has occurred.⁹ The incidents prompt questions about whether users of the site should be afforded informed consent before their private data is used for experimental purposes.

Problem 2.6: Exploiting cognitive bias

Society must value human health and human wellbeing to sustain itself. However, in free societies, adults have the right to exercise free will, make their own decisions and pursue their own personal happiness.

Corporations advertise and market to consumers. Some industries such as tobacco, alcohol and gambling harm consumers. Adult choice in these product segments is confounded by addictive behaviours. There is an appearance of freedom and personal choice, but in some cases the marketers exploit human nature to align product placement towards cognitive bias and addictive weaknesses. The social effects have economic consequences. Financial gain moves from the consumer to the marketer, but harms and financial losses are borne by the unwitting consumer or socialised across the community through health insurance costs. In the cases of tobacco, alcohol and gambling, governments have stepped in to regulate or to apply offsetting taxes. In recent years, some jurisdictions have implemented regulations or taxes on sugar citing recent research of addiction harm to public health.^{10,11}

Online platforms, sensors, data acquisition systems, machine learning and artificial intelligence systems have great power to understand users. Vulnerabilities could be exploited for commercial gain. Intellectual, emotional, compulsive or addiction weaknesses could be targeted. The outcomes could be economic or human rights outcomes detrimental to vulnerable people. Emerging technologies such as virtual reality and augmented reality will have much greater power than printed paper media, radio, television or today's internet. Technologies can be developed that "gamify" user activities and exploit gambling types of addictive behaviours.

Economic models of the past do not explain how artificial intelligence might be used to learn and exploit cognitive bias. The behavioural science in this area is relatively

⁹ Michael Zimmer, "OKCupid Study Reveals the Perils of Big-Data Science", 14May2016, Wired, online at <https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/>.

¹⁰ Barry Popkin and Corinna Hawkes, "Sweetening of the global diet, particularly beverages: patterns, trends, and policy responses", *The Lancet – Diabetes and Endocrinology*, Vol 4, Issue 2, pp174-186, 01Feb2016, online at [https://www.thelancet.com/journals/landia/article/PIIS2213-8587\(15\)00419-2/fulltext](https://www.thelancet.com/journals/landia/article/PIIS2213-8587(15)00419-2/fulltext)

¹¹ Franc Sassi et al., "Equity impacts of price policies to promote healthy behaviours", *The Lancet – Taskforce on NCDs and Economics*, Volume 391, Issue 10134, pp 2059-2070, 19May2018, online at [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(18\)30531-2/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)30531-2/fulltext).

new.¹² Consumers do not always act rationally in their own best interest. Purchasing behaviours might be exploited with clever “choice architecture”, and emerging technologies might amplify or exaggerate human weaknesses.

Political-economic descriptions might be stuck in an old-fashioned Adam Smith era. The modern concept of Moore’s Law and the exponential growth of transistors and computing power drives the rate of technology deployment. Sometimes that deployment comes in bursts. For example, one generation of a smart phone integrates GPS location, wifi-enabled maps and cloud-based computation to enable taxi substitute ride share apps very quickly. Economic disruption to an industry occurs very rapidly. This is the definition of the “killer app” that goes viral.

A circumstance could be imagined whereby a rapid-growth, AI-informed marketing decision maker deploys opposite a vulnerable population of consumers. The (legal) opioid epidemic has caught pharmaceutical regulators off guard by its speed. A digital epidemic that affects consumer economics or that exploits the privacy, security, safety or fair treatment of users through exploiting cognitive bias is an easily foreseeable future scenario. In some respects, this is already occurring. The issues outlined so far would align. Rapid growth companies could be “too big to regulate”, too fast to regulate or too wealthy to regulate (Issue #3). They might possess asymmetrical technical hardware, technical know-how and data in a contest versus average people (Issue #4), and they might experiment on their users (Issue #5) and exploit addictive weaknesses or cognitive biases.

3. Critical gaps in law / regulatory systems

3.1 International markets, but lack of international norms

There is currently a high volume of technology creation in the US (especially Silicon Valley) and in China. The US exports a large volume of digital technology globally, and much of Australia’s technology import volume originates from the US. The EU creates less digital technology but has undertaken more considered regulation on privacy and consumer protection, reflecting European sensibilities about individuals’ rights in commerce. There are vast differences in US, EU and Chinese technology regulations. Recent developments in US net neutrality,¹³ reinforcement of the Great Firewall of China,¹⁴ and EU action against Facebook’s trans-Atlantic communication of personal data¹⁵ are all examples of differences in domestic regulatory regimes. There are no international standards on data privacy, and the emerging fields of machine

¹² Daniel Kahneman. Thinking, Fast and Slow. Farrar, Straus & Giroux. 2011

¹³ See for example, <https://thewarrencentre.org.au/tag/net-neutrality/>

¹⁴ See for example, <https://thewarrencentre.org.au/whatsapp-china-great-firewall-reinforced/>

¹⁵ See for example, <https://thewarrencentre.org.au/facebook-awaits-data-protection-ruling/>

learning and artificial intelligence are areas where community expectations of fairness and ethical programming could increase challenges due to the opacity of AI algorithms. In the absence of international standards and local regulations (or appropriately effective and balanced local enforcement), de facto standards of commercial behaviour may be imported to Australia and normalised before local governments can react. Certainly Facebook and Twitter have re-set Australian norms about interpersonal communication. In the past two years, there are deep concerns in the US and UK (at least) about the effects of foreign actors on domestic elections.¹⁶ Antisocial digital media ‘bots’ have spouted fake news and stirred discontent that destabilises governments.

3.2 Misalignment to Australian community standards

It is likely that data acquisition, machine learning and artificial intelligence systems that are imported would not align to Australians’ sense of fairness.

Australia’s community standards of ethics and fairness are distinct.¹⁷ Australia’s population has diverse multi-ethnic and multi-cultural roots. There are strong cultural ties to Britain and Europe, but proximity to the Asia-Pacific region as well as unique indigenous cultural influences define Australian values. Australia’s history also created a unique sensibility of what is fair and right.

Australians share a common self-identity that embraces diverse origins but still values community engagement.¹⁸ Fair dinkum and the fair go are alive and well. Gender, ethnic and sexual orientation diversities are respected.

3.3 Stewardship and scholarship needed

Machine learning and artificial intelligence offer massive opportunities for improved economic productivity and advancement in fields such as health care, but there is a strong need for trust, accountability, transparency, security and privacy.

The UK’s Royal Society recently recommended that the strong benefits and countervailing challenges posed by machine learning and AI require “careful stewardship”.¹⁹ In its review, the UK Government Office for Science issued key

¹⁶ See for example, <https://thewarrencentre.org.au/russian-rampage-first-the-us-and-now-the-uk/>

¹⁷ Stephen Johnston, Helen McGregor and Elizabeth Taylor, “Practice-focused ethics in Australian engineering education”, *European Journal of Engineering Education*, 2000, Vol 25, No 4, 315-324.

¹⁸ Mark McCrindle, “21st Century Australia: Understanding the Shifts” [online], *Culturescope*, Vol. 93, 2010: 4-6.

¹⁹ *Machine learning: the power and promise of computers that learn by example*, Royal Society, 2017 at <https://royalsociety.org/topics-policy/projects/machine-learning/>.

recommendations on education, governance and research funding.²⁰ There should be understanding of how machines make decisions so that external bodies can hold computer algorithms accountable if harmful decisions are made by AI.

In America, legal researcher Matthew Scherer says that a “regulatory vacuum” exists and that there is presently a serious shortage of legal scholarship on AI.²¹ Citing examples such as the outcomes of autonomous driving accidents and adverse outcomes from medical AI, Scherer expresses deep concern that governments lack competencies to address the speed of new technology deployment and the lack of visibility of how AI is actually making its decisions. He states:

The increasing ubiquity of AI makes it all but certain that AI systems will generate many public risks. Those risks may prove difficult for the legal system to address, because AI presents challenges not raised by the public risks of the twentieth century.

Ultimately, Scherer recommends that expert government agencies are needed with powers delegated from legislatures.

A different American legal researcher argues that a new breed of expert government agency similar to pharmaceutical or medical device regulators (the US FDA or Australian TGA) is needed, a so-called “FDA for Algorithms”, to regulate artificial intelligence to prove that certain high-risk systems are “safe and effective for intended use”.²²

3.4 Need for STEM education, diversity and inclusivity

For many years, the Warren Centre has advocated for strong STEM education to prepare students for the future.²³ We currently operate a program called Inclusion² focussing on encouraging greater gender diversity and inclusivity for the STEM professions and the entrepreneurship community. It is obvious that Australia’s efforts to improve diversity and inclusivity must accelerate.

3.5 Specific questions in the consultation

²⁰ *Artificial intelligence: opportunities and implications for the future of decision making*, Government Office for Science, 2016, at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf

²¹ Matthew U Scherer, “Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies and Strategies”, *Harvard Journal of Law and Technology*, 29:353, Spring 2016, p 353.

²² Andrew Tutt, “An FDA for Algorithms”, *Administrative Law Review*, Winter 2017, 69:83.

²³ See for example, “Vision for a STEM Nation”, 2015 at <https://thewarrencentre.org.au/wp-content/uploads/2014/11/Warren-Centre-Submission-wc2980-14.pdf>

(Question 6a) Specifically addressing the overarching objectives of regulation, we recommend:

1. Provide benefits to Australian society: wealth, health, wellbeing
2. Encourage innovation and market experimentation
3. Encourage self-regulation through the growth of professionalisation and performance-based outcomes
4. Implement light touch regulation where mischief or market failure occurs.
5. Avoid heavy regulation except where significant harms occur. Adopt evidence-based policy making.

(Question 6b) Regarding principles to achieve objectives, the DeepMind principles are worthy of attention (though not intended in a regulatory framework):²⁴

1. Social benefit
2. Rigorous and evidence-based
3. Transparent and open
4. Diverse and interdisciplinary
5. Collaborative and inclusive

(Question 7) The paper asks if there is a role for organisations that promote responsible innovation, self-regulatory, co-regulatory and ‘regulations by design’ approaches. The Warren Centre for Advanced Engineering has undertaken a number of long-range projects to redefine professional and industry approaches or alternatives to regulations. In the early 1990s the Warren Centre convened reform of fire safety engineering to adopt performance-based solutions. We are re-visiting that work now in the context of professionalising fire safety engineers. In the early 2000s, the Warren Centre undertook a long-term project to explore professional performance and balance risk versus innovation. The outcome of that project was a voluntary Protocol for Performance which records a standard of care for professional engineers. The Protocol for Performance sits alongside the Engineers Australia definitions of ethics and competency. Based on these experiences, the Warren Centre believes there is significant scope for organisations to promote such approaches.

In our experience, it is a necessary step to engage the technical experts working in this space and create a cohesive group of professionals. The Hippocratic Oath or the Johnson & Johnson Credo are examples of concise aspirational targets.

Other examples exist in engineering during the development of new technologies. At the beginning of the Industrial Revolution, engineers created codes to govern steam

²⁴ DeepMind is a subsidiary of Alphabet. See <https://deepmind.com/applied/deepmind-ethics-society/principles/> .

engines, pressure vessels and pressure relief valves. The American Society of Mechanical Engineering (ASME) created standards from 1880 to 1920 to address boiler explosions that resulted in fatalities in the US. From water boilers and steam systems, safety technology expanded into the new fields of manufacturing, storing and transporting petroleum. A whole new field of process safety developed with strong contributions from the IChemE, the American Institute of Chemical Engineers, the UK Health and Safety Executive, the US Chemical Safety Board and similar international organisations. Safety advocates like the UK's Trevor Kletz and Australia's Rolf Prince pioneered scholarship in safety technology. These heroes of the profession were role models and standard bearers who established new paradigms for professional performance in safety engineering.

New digital technologies could benefit from organisations that promote responsible innovation, self-regulatory, co-regulatory and 'regulations by design' approaches. Presently, we are aware of some groups led by major corporations or centred around academics. To be successful, such groups need new ethics/safety technology research that is separated from industry and commercial bias, but closely informed by real world industry knowledge. They need high professional and technical skill along with a sense of professional identity, status of a profession and commitment to a sense of shared values.

Conclusion

To maximise the positive economic effects of digital technologies, we recommend the following. Additional research should be undertaken in the development of "ethical" machine learning and artificial intelligence. Moving beyond academia and technology developers, efforts should be undertaken to open multi-disciplinary engagement and discourse on AI. This should include IT developers, engineers, lawyers, ethicists, government, industry, academia and the public. Advanced technologies including data acquisition, machine learning and artificial intelligence are critical to the economic growth of Australia. Proper regulatory settings can support economic growth. Better public awareness can inform better democracy and better public policy. At a time of economic disruption, some views may inhibit innovation and forfeit domestic economic growth opportunities to foreign countries.

Ethical AI guidelines should be developed based on Australia's own unique sensibilities and values of privacy, fairness and transparency. The nation should not delegate this to foreign companies or foreign governments. No company in California should dominate this discussion without an Australian voice. It should not be delegated to a hardware supplier in Shenzhen. Australia should establish and lead international dialogues on AI and robotics ethics and seek to influence

international norms to reflect Australia's own ethic of fairness. Australia has a high capacity to influence emerging international law. This is an area of high importance internationally and domestically. It has high consequences economically and to the rights of citizens around the world. It is worthy of national effort.

Enhanced STEM education along with diversity and inclusivity efforts are needed to promote domestic talent development and to minimise the potential for the distorting effects of bias in AI and machine learning algorithms.

Contacts

Ashley Brinson

Executive Director, The Warren Centre for Advanced Engineering

ashley.brinson@sydney.edu.au, (02) 9351 3752

About the Warren Centre for Advanced Engineering

The Warren Centre constantly challenges the economic, legal, environmental, social and political issues raised by innovation. We collaborate with industry, government and academia to achieve globally significant outcomes.

<https://thewarrencentre.org.au/>