Artificial Intelligence: Governance and Leadership White Paper 2019

Australian Human Rights Commission and World Economic Forum

Response from The University of Melbourne

18 March 2019

Executive Summary

The University of Melbourne welcomes the opportunity to respond to the Australian Human Rights Commission and World Economic Forum's *Artificial intelligence: Governance and Leadership White Paper*. This submission builds on the University's response in 2018 to the Commission's *Human Rights and Technology Issues Paper* and associated policy roundtables hosted at the University.

The University commends the parallel focus on regulatory considerations, in light of the scope, complexity, pace of change and potential social influence of innovative technologies. As per our response to the Commission's human rights paper, the University recommends a multi-faceted and co-regulatory response to protecting human rights and preventing or minimising social harm, coordinated across relevant bodies and underpinned by the principal of 'human rights by design and default'.

An appropriate response to the flow of new technologies is the establishment of a framework that attends to the genuine risks without stifling innovation. Ideally, the framework should promote sustained comparative advantage in ethical design. Existing regulatory arrangements in the domains of privacy and consumer protection, as well as human rights, need to be elaborated and improved to properly accommodate the changing opportunities and risks of novel technologies. The University has therefore recommended mapping of existing regulators to establish the extent of change needed and strengthening of regulatory powers, functions and resources.

When societal damage, or risk of harm, is unambiguously a consequence of innovation, there is a case for regulatory intervention. The overarching objective for Australian law and regulation should be the provision of a regulatory space that promotes and fosters beneficial innovation while protecting against social harm. Enhanced functions of existing regulatory bodies, or in the alternative, a carefully calibrated new RIO with specific and non-duplicative functions, may be a useful supplement to Australia's existing frameworks. Additionally, a specialist advisory body should be established to support responsible regulation and provide high-level advice and engagement on innovative technologies to the public, governments and other regulatory entities.

For more information, please contact Professor Mark Hargreaves, Pro Vice-Chancellor (Research Collaboration & Partnerships) on 03 8344 4447 or m.hargreaves@unimelb.edu.au.

Recommendations

1. **The use of innovative technologies, including AI and ML, should be subject to greater oversight and regulatory responsibility to safeguard social expectations and protections including fairness, accountability, transparency, inclusiveness and non-discrimination in the use of those technologies.**

2. **Existing regulatory bodies, and relevant legal and regulatory frameworks, in Australia should be examined to determine if enhancing their powers and mandates would address regulatory requirements relating to innovative technologies such as AI. The mapping exercise should look at regulatory functions as well as diverse applications of automated decision-making technologies.**

3. **A specialist advisory organisation on AI and innovative technologies should be established to coordinate with other regulatory bodies and provide key supports such as;**

    a. **Providing community education and advice;**

b. **Reviewing and recommending on regulatory needs;**

c. **Providing technological and social analytical capability;**

d. **Conducting self-enabled inquiry into emerging issues;**

e. **Conducting community and cross-sector (including private sector) engagement on AI responsibilities;**

f. **Drawing on international experience/expertise and promoting relevant Australian examples and practices;**

g. **Providing expert guidance to Australian governments and public sector;**

h. **Developing a well-regarded and reliable 'trust' mark or equivalent identifier.**

4. **Inclusivity should be a core aim of a specialist advisory organisation or a RIO, which can be ensured by using a 'universal design' approach to the formation of any new regulatory organisation that treats impact groups as core stakeholders, rather than as exceptions.**

5. **A specialist advisory organisation or RIO should have access to internal and external expert guidance, such as a panel drawn from research and development sectors, to ensure currency of information and breadth of networked expertise.**

1: What should be the main goals of government regulation of AI?

Comments on definition

At the outset, the University raises a definitional concern about the use of 'artificial intelligence (AI) and machine learning (ML)' as the technological focus of this inquiry into regulation. While the White Paper rightly identifies AI as a key driver in current industrial transformation and investment, the fields that are relevant to the themes of this inquiry – and the proposal for a Responsible 'Innovation' Organisation (RIO) – expand beyond AI and ML. A RIO described as such would have relevance to a wide and fast-moving set of innovative technologies, digital or otherwise (e.g. cyber-physical, with combined physical and computational elements), including diverse forms of computing, automation, nanotech, Internet of Things (IoT), advanced materials, biotechnology, and others. Many data protection regimes around the world have preferred a broader or more technology neutral terminology that can keep pace with scientific advancements.[1]

In the case that narrower AI capabilities are identified for regulation to avoid social harm in that specific area, an applicable terminology could be 'automated decision-making'. However, too narrow a definition runs a practical risk of regulatory avoidance, in which researchers or companies choose to strategically rename new technologies to remain outside regulation. Additionally, narrowly aimed regulation – such as rules or standards triggered by the use of big data set-based AI – risk being outpaced when AI develops to the point it can be trained with much smaller data sets. As previously

[1] For a broader approach to the complex linkages between technologies and regulatory structures, see the analysis of 'Responsible Research and Innovation': a 'transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)'. Von Schomberg, 2013, p. 63 referenced in:
https://www.frontiersin.org/articles/10.3389/fpls.2018.01884/full

raised in the University's response to the Commission's human rights paper, from a legal perspective the distinction between AI algorithms and other decision-making algorithms is irrelevant, and any regulatory response should be independent of algorithmic implementation details.

Australian needs in a global context

As noted in the White Paper, Australia already lags well behind in terms of Australian companies making sustained investments in AI. There is therefore a risk that heavy-handed regulation will make Australian-led technological advancements less relevant in a world context. At the same time there is growing interest in technologies that create public value.

The University therefore supports the development of a carefully calibrated response which would include an Australian specialist advisory organisation with global reach and connections, that can tap into the significant body of work on innovation, ethics and regulation being undertaken by industry, think tanks, universities, international organisations, and civil society groups globally, and deepen engagement with international partners without duplicating efforts. It is also important that it should reflect uniquely Australian needs and concerns, given the fundamental social protection aspect of its function. For instance, the impacts and interests of technological innovation on Australian indigenous people is not necessarily replicated by overseas algorithmic practices and models of regulation.

Fostering responsibility and public protections

The AI4People report launched in November 2018, *Ethical Framework for Good AI Society*, provides an excellent summary of the opportunities, risks and principles for ethical use of technology. Describing AI as enhanced, or even improved and multiplied, human agency, the report observes:

> The larger the number of people who will enjoy the opportunities and benefits of such a reservoir of smart agency "on tap", the better our societies will be. Responsibility is therefore essential, in view of what sort of AI we develop, how we use it, and whether we share with everyone its advantages and benefits. Obviously, the corresponding risk is the absence of such responsibility.[2]

In developing organisational and/or regulatory responsibility in the Australian context, the goals should accord with social values such as:

- Fairness;
- Accountability;
- Transparency;
- Inclusiveness;
- Non-discrimination.

Goals such as the above would usefully set a baseline for how AI should operate in Australian civil society. These include that AI does not discriminate against citizens, does not bully, coerce or exploit citizens, and does not cause harm to citizens. There is an important power dynamic that gives rise to the necessity of fostering technological responsibility in the public and private sectors; for the most part, citizens are most likely to be affected by automated decision-making but have little ability to influence the content and application of AI.

[2] http://www.eismd.eu/wp-content/uploads/2019/02/Ethical-Framework-for-a-Good-AI-Society.pdf page 8.

Promoting innovation and market protections

It is important that in regulating to achieve these goals, innovation is not stifled. Alongside protection of the public, regulation should facilitate and ensure the appropriate functioning of a competitive marketplace for new technologies. A well-designed regulatory structure could set up a lasting framework for ways in which innovative breakthroughs can be taken to market. Equally, it should be remembered that innovation can be fostered and encouraged within a regulatory regime that encourages best practice, rather than allowing a proverbial 'race to the bottom' in the AI context.

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?

Domestically and internationally, institutions and companies are looking closely at the expansive social implications, including altered pressures on regulatory arrangements, brought on by the advancement of AI and innovative technologies. In recent years the OECD has been investigating the digital transformation of economies and societies.[3] UNESCO is leading a program on technological transformations with a focus on the Sustainable Development Goals and human impacts.[4] The Institute of Electrical and Electronics Engineers (IEEE), one of the world's largest industry standards bodies, has convened numerous global standards projects on AI and in 2018 launched an Ethics Certification Program for Autonomous and Intelligent Systems.

While these projects are looking at distinct aspects of AI, regulation (and its limitations) is emerging as a theme. Jurisdictions outside Australia are moving to regulate aspects of AI and serve as a useful comparison. For instance in mid-late 2018 the California legislature became the first US state to pass bills regulating rapidly-developing technologies including the IoT, AI, and chatbots. Europe already has in place a comprehensive General Data Protection Regulation (GDPR), which include provisions for data protection by design and the right of data subjects 'not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her' (subject to certain limited exceptions).

In Australia, the existing regulatory bodies that play a role, to various degrees, in this area include:

- The Australian Competition and Consumer Commission (ACCC);
- The Office of the Australian Information Commissioner;
- The Australian Securities and Investments Commission;
- Office of the National Data Commissioner;
- Specific industries regulators, such as the Actuaries Institute and Australian Communications and Media Authority (ACMA);[5]
- The Australian Human Rights Commission.

The legal framework for protecting human rights in the context of new technology is dispersed across a range of Acts and Instruments, including the Privacy Act, anti-discrimination laws, and the Australian Consumer Law. Policy platforms such as the National Innovation and Science Agenda, supported by government offices like Data61 and the Office of the Chief Scientist, would also have relevance.

[3] http://www.oecd.org/going-digital/ai/

[4] https://en.unesco.org/artificial-intelligence

[5] ACMA have a significant role in relation to the NBN as well as technology companies operating in the media-communications space (e.g. live-streaming on Facebook).

Enhance regulatory powers in existing bodies

The University notes the comments in the White Paper excerpted from contributors to the Commission's earlier human rights paper, including the ACCC's recommendations, highlighting the need for greater regulatory oversight of digital platforms. However, the University does not, at this stage, recommend the establishment of an entirely new regulatory body to absorb all functions necessary to meet this need in the particular context of AI.

As a first step, the University recommends due consideration be given to whether enhancing and expanding the functions and powers of existing regulatory bodies could meet the new regulatory needs relating to innovative technologies. Many of the regulatory bodies listed above are already working collaboratively on challenges relevant to their sectoral remit or overlapping their remits. Establishing a new national RIO without considering workable alternatives in the present regulatory constellation could lead to duplication and impediments to innovation, which would be counterproductive to the intent of the reforms. With carefully calibrated enhancements (including in Australian law), and appropriate resourcing to cover the new areas of responsibility, existing regulators could efficiently and effectively step in to collaboratively cover regulatory gap.

Along these lines, the AI Now Institute's report in December 2018 recommended that governments should regulate AI by expanding the powers of sector-specific agencies to oversee, audit, and monitor these technologies by domain. Looking at the US, the report argued that 'a national AI safety body or general AI standards and certification model will struggle to meet the sectoral expertise requirements needed for nuanced regulation. We need a sector-specific approach that does not prioritize the technology but focuses on its application within a given domain'.[6] The University submits that this observation is relevant to present Australian circumstances.

The University's previous submission to the Commission's human rights inquiry outlined numerous ways in which Australia's existing legal framework could be updated to manage the challenges and maximise the opportunities created by advances in technology. It also flagged the useful comparative framework of the European Union's GDPR, which is significantly more advanced than the equivalent regulations in Australia and may be used as a guide to updating Australia's legal and regulatory arrangements.

Role for a specialist advisory organisation

While not a national RIO in the form sketched out by the White Paper, the University suggests there is scope and need for establishing a new specialist entity with deep technical expertise, so far as it does not replicate or overlap the remit of existing regulatory bodies. A specialist organisation in this vein could lead a coordinated model with other regulators, as well as provide the following supports:

- A platform for community education and advice;
- A mandate to review, monitor and recommend on responsible regulation;
    - This should include identifying what powers regulators need over time, supporting them to build specialist skills and looking ahead to help them anticipate emerging issues for their sectors;
- Technical/technological capability, to assist other regulators where the workings of innovation technologies are opaque or are not well understood;
- Social science and diverse disciplinary capability, to contextualise and humanise the social impact of innovative technologies;

[6] https://ainowinstitute.org/AI_Now_2018_Report.pdf page 4.

- Ability to conduct self-enabled inquiries into emerging issues;
- Mandate to facilitate community and cross-sector engagement;
- Provision of expert guidance to Australian, State and Territory governments, to assist with internal public service use and design of innovative technologies, and to assist with the standards required by governments in procurement and partnership;
- Strong links to draw on international experience/expertise and promote relevant Australian examples and practices;
- Potentially, ability to conduct *in camera* inquiries where a public issue has arisen in regard to commercially held algorithmic function; and
- Potentially, development of a well-regarded and reliable trust mark or verification to appropriately signal public confidence in the fairness and safety of a product or process impacted by innovative technology.

Map existing regulation to maintain coherence

Before introducing a new specialist entity or RIO in the terms of the White Paper, it will be important to carefully map the scope and function of existing regulatory regimes and bodies. This will be an opportunity to ensure a good fit between what is already in place and any new initiatives, to ensure coherence across the sector and avoid duplication.

As raised earlier, AI more narrowly defined is a decision-making capability in software that is application and scenario specific. The mapping exercise should therefore also look at application areas, e.g. advertising and direct marketing; job automation; robotics and others, to identify if there is an existing regulatory regime that provides appropriate coverage.

3: Would there be significant economic and/or social value for Australia in establishing a Responsible Innovation Organisation?

Improved regulation of innovative technologies would add social value for Australia by improving the general public's trust and familiarity with AI and related technologies. There is currently a lack of general education about technology and its regulation, which contributes to limited public support for its uses. There is also a perception in some parts of the population that the Australian Government is not well-equipped to work with cutting-edge digital technologies; a specialist advisory organisation could raise the public perception of the government's data handling capabilities.

There could be significant social and economic value in Australia taking a leading role in this area, rather than catching up with, or adapting, the approach taken elsewhere globally. It is worth observing that private and public institutions are moving swiftly in the area of AI skills and social impacts in other countries. For instance, in an effort to keep pace with China and the United States, the United Kingdom (UK) Government has recently announced plans to fund over 1,000 places for Masters and PhD students to train in AI technologies at the cost of approximately £115 million. Businesses, including Google DeepMind, BAE Systems and Cisco, have also pledged to help fund 200 new AI Masters courses at UK universities.[7] A carefully calibrated regulatory response in the immediate term, with appropriate review and revision, designed to enhance Australian capabilities could enable Australia to stay relevant in a global context and gain social and economic benefits of doing so.

[7] Sam Shead, 'U.K. Government To Fund AI University Courses With £115m', *Forbes*, 22 February 2019: https://www.forbes.com/sites/samshead/2019/02/20/uk-government-to-fund-ai-university-courses-with-115m/#5bfce7c4430d1/4

4: Under what circumstances would a Responsible Innovation Organisation add value to your organisation directly?

A specialist regulatory organisation could provide an additional and valuable channel for the University's research and educational expertise to flow out beyond the University to influence and inform public policy development. This is discussed in more detail below at (6d). A strong regulatory framework designed to support innovative breakthroughs being taken to market could also assist the University to manage risks that adversely impact our core businesses of teaching and research, while facilitating the productive expression of university-generated IP.

A high-functioning RIO or specialist advisory entity could, in turn, assist the university sector by providing informed forecasts on how teaching and public education services can be adapted or deployed to reflect and/or help people manage emerging technologies, trends and issues.

5: How should the business case for a Responsible Innovation Organisation be measured?

The University is silent on this point.

6: Features of a Responsible Innovation Organisation

This section of the White Paper sought feedback on the potential powers, functions, aims, connections, resourcing and evaluation of a RIO.

Inclusivity as a core aim (6a)

Inclusivity should be a core aim of any regulatory intervention. In a recent article published by Nature, an observation was made in relation to in the use of AI in healthcare in the UK:

> As with the advent of the car, many serious implications will be emergent, and the harshest effects borne by communities with the least powerful voices. We need to move our gaze from individuals to systems to communities, and back again. We must bring together diverse expertise, including workers and citizens, to develop a framework that health systems can use to anticipate and address issues. This framework needs an explicit mandate to consider and anticipate the social consequences of AI – and to keep watch over its effects.[8]

The White Paper noted the importance that the RIO's approach and governance be inclusive, with 'special attention given to those who are particularly affected by new technologies and most susceptible to the threats associated with them'. The University recommends a universal design approach could effectively inform the design of an inclusive RIO, by placing the perspectives of all people as central to the design process rather than as exceptional, and at the same time maintaining reasonable opportunities for human intervention and oversight of the design processes.

The universal design approach ensures that people with disability, children and young people, older people, and people from CALD backgrounds are contemplated as core stakeholders, rather than being viewed as special or vulnerable groups whose perspectives and needs are an add-on to the RIO's core model or business.[9] While some Federal, State/Territory, local governments are starting to identify

[8] Melanie Smallman, 'Policies designed for Drugs won't work for AI', *Nature*, Vol 567, 7 March 2019, p7.

[9] See https://www.wired.com/2017/03/voice-is-the-next-big-platform-unless-you-have-an-accent/

and encourage the application of universal design principles, particularly in relation to the built environment, the most prominent current example is the National Disability Strategy 2010-2020.[10]

As outlined in the University's response to the human rights Issues Paper, universal design principles are appropriate to the development of AI to ensure systematic bias or prejudice is not, even inadvertently, built into datasets or processes. Universal design should also be reflected in the design of regulatory frameworks, as stated by the UN Convention on the Rights of Persons with Disabilities (preamble para (o) and Article 33).

Powers and functions (6b)

Whether established in an existing regulator or in a new RIO, there must be enhanced supports created for people to seek redress from technology-driven harms by ensuring people's complaints are fairly handled and remedies enforced. As mentioned in other parts of this submission, an innovative technologies regulator could also create value for AI applications in Australia by developing a well-regarded and reliable trust mark or verification.

Structure (6c)

If a new RIO as sketched out by the White Paper is ultimately recommended by the Commission, a useful starting model is the ACCC. The ACCC is an effective regulator as it achieves a broad cross-sector reach, conducts stakeholder consultations, balances regulatory oversight with specialist expertise, and has a suite of powers allowing effective enforcement of the regulatory regime it administers. The consumer protection regime allows for both the ACCC and individual consumers to bring actions to protect their rights. Notably, the Australian Government fully funds the ACCC. A regulatory body that is co-funded by industry would need to be carefully structured to ensure its independence and avoid perception of regulatory capture.

Internal and external expertise (6d)

The University strongly recommends a specialist advisory organisation or a RIO should have inbuilt and external assets of computing technical expertise, including people from outside the Sydney/Canberra/Melbourne axis. As stated earlier in this response, it is also crucial to have social science skills that will help technologists and bureaucrats understand the impact of innovative technologies in the community.

The Australian university sector has a key role to play in the trajectory and impact of new technologies as a key nexus point of education, research and development. Publicly funded research is an enabler of innovation in the broader economy and is increasingly global and multi-disciplinary. Universities and partners in industry and elsewhere, being familiar with the rate of technological change and disciplinary complexity, offer a wealth of expertise to respond to these challenges as they arise.

The University recommends an External Advisory Group be established as part of the RIO to draw in expertise from relevant sectors, including universities. This would be consistent with the approach taken for other regulatory bodies. For instance, ASIC takes guidance from several advisory panels including an External Advisory Panel, which channels senior level advice from the financial services industry and other sectors.

[10] See https://www.humanrights.gov.au/sites/default/files/NDS%20PDF.pdf p 30.

References

https://www.frontiersin.org/articles/10.3389/fpls.2018.01884/full

http://www.eismd.eu/wp-content/uploads/2019/02/Ethical-Framework-for-a-Good-AI-Society.pdf

http://www.oecd.org/going-digital/ai/

https://en.unesco.org/artificial-intelligence

https://ainowinstitute.org/AI_Now_2018_Report.pdf

https://www.forbes.com/sites/samshead/2019/02/20/uk-government-to-fund-ai-university-courses-with-115m/#5bfce7c4430d1/4

https://www.wired.com/2017/03/voice-is-the-next-big-platform-unless-you-have-an-accent/

https://www.humanrights.gov.au/sites/default/files/NDS%20PDF.pdf

List of contributors from The University of Melbourne

Dr Greg Adamson, Associate Professor and Enterprise Fellow is Cyber Security, Melbourne School of Engineering

Mr Doron Ben-Meir, Vice-Principal (Enterprise)

Professor Bruce Bonyhady, Executive Chair and Director, Melbourne Disability Institute

Professor Mark Hargreaves, Pro Vice-Chancellor (Research Collaboration & Partnerships)

Dr Yvette Maker, Senior Research Associate, Melbourne Social Equity Institute

Professor Scott McQuire, School of Culture and Communication

Professor Bernadette McSherry, Foundation Director, Melbourne Social Equity Institute

Associate Professor Tim Miller, School of Computing and Information Systems

Associate Professor Jeannie Paterson, Melbourne Law School

Professor Megan Richardson, Melbourne Law School

Associate Professor Ben Rubinstein, Senior Lecturer, School of Computing and Information Systems

Dr Rajeev Samarage, Senior Data Scientist, Melbourne Institute of Applied Economics and Social Research

Professor Liz Sonenberg, Pro Vice-Chancellor (Digital & Data)

Associate Professor Mark Taylor, Deputy Director, Centre for Health, Law and Emerging Technologies (HeLEX @Melbourne)

Professor Tony Wirth, School of Computing and Information Systems

Ms Kelly Farrow, Policy and Government Relations Adviser, Chancellery