

Submission to Australian Human Rights Commission and World Economic Forum Artificial Intelligence: governance and leadership

White paper

Firstly, thanks for focussing attention on the important issues raised in your White Paper. While I am not convinced of the merits of establishing a Responsible Innovation Organisation these are obviously important matters we must address, and it is important that all elements of our society and economy collectively own and engage with this responsibility.

1. What should be the main goals of government regulation in the area of artificial intelligence?

This question begs the question of regulation, presupposing that it is required. It also frames the issue of regulation with reference to the frame of a particular technology sector (AI). This is a problematic approach: if it is mirrored for every different area of technology development that may raise regulatory issues (e.g. nanotechnology, biotechnology, gene hacking, biohacking, Internet of Things etc.) then the complexity of the regulatory environment will be rapidly compounded as new fields emerge. A sector focussed approach may also fail to adequately deal with how the different technologies and tools are used in conjunction or issues that arise at their intersection. A proliferation of detailed approaches is likely to be confusing for a variety of stakeholders, including the general public and industry, and may not assist intelligent and sensitive uptake of these new tools.¹

I submit that systems of regulation and law are more readily understood and exercise greater influence where they are simply expressed and provide over-arching guidance for behaviour. Consider for example the simplicity of the prohibitions on misleading and deceptive conduct in the Australian Consumer Law, which are readily understood and have become an important and powerful tool for regulators and commercial actors to ensure appropriate marketplace conduct. Note also the distillation by Commissioner Hayne of very simple principles of guidance in the course of preparing the recent Financial Services Royal Commission report.

Building on these ideas, rather than identifying a new technology, which opens up a new range of possibilities, and looking to develop a discrete regulatory approach around that, I believe that it is likely to be more productive to consider gaps in our broader legal system as they relate to general rights and obligations. Driving the approach from the perspective of higher level characteristics of desirable and undesirable behaviours. It is still useful to consider the cross-impact of regulation and new technologies, but I don't believe we can productively design all regulation through the lens of particular technologies – this will just produce a bewildering kaleidoscope that is more likely to mesmerise than assist everyday practice. Instead we should adopt a principles based approach, in line with recognised regulatory theory.²

¹ I note that UTS advocated a broader Technology Assessment Office but that the focus of the Whitepaper appears to be more focussed on AI per se. There would naturally be other complexities in trying to create any one body to oversee the full gamut of technology developments - and I am not advocating that - but simply highlighting that this other viewpoint also illustrates that there are many fields that raise such issues.

² J Black, *Principles Based Regulation: Risks, Challenges and Opportunities* (2007) London School of Economics and Political Science

Naturally there is a breadth of areas that new technologies such as AI will touch on, including but by no means limited to privacy, security, human rights perspectives, labour regulation, contracting (including via intelligent agents), consumer data rights and even fundamental constitutional issues.³

It is also important to identify areas of existing regulatory frameworks that are insufficiently “technology neutral” and therefore run the risk of becoming outdated by being too particularly tuned to past technologies. Overreacting to current technological change now raises the risk of creating more of such issues in the future.

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?

See comments in response to question 1.

Setting aside regulation, many bodies will exert some degree of “influence” over AI in Australia, and it is perhaps this “influence” that is more significant in its reach than any regulatory systems. Clearly multinational businesses, research organisations, think-tanks and foreign legal systems exert an enormous influence on these issues as Australia is necessarily only ever going to contribute a small fraction of the world’s broader technological developments, business and governance/ethical systems.⁴ Local counterparts in education (at all levels), business, government and broader civil society obviously play a role.

Going beyond that there will be many areas where current regulatory systems arguably fail to respond appropriately to the advent of AI/machine learning.

To provide one very small example, consider the issue of copyright as it relates to machine generated works. For many decades this is an area that various law reform bodies and reports have pointed to as being a gap in Australia’s copyright system.⁵ In contra distinction other countries dealt with part of this problem some time ago.⁶

³ See e.g. Nicolas Suzor, *Lawless: the secret rules that govern our digital lives* (Cambridge Uni Press, 2019) - pre print available at <https://osf.io/preprints/socarxiv/ack26>

⁴ One obvious example of the international reach of regulatory systems is the EU GDPR and its influence on businesses with links to EU personal information

⁵ see e.g. Copyright Law Review Committee, *Computer Software Protection* (Canberra: Office of Legal Information and Publishing, Attorney-General’s Department, 1995)

⁶ e.g. the deeming provisions in the UK’s Copyright Design and Patents Act – though the extent to which those “patches” will sufficiently address works fully generated by AI is debatable – see for example – Dr Paul Lambert “Computer Generated Works and Copyright: Selfies, Traps, Robots, AI and Machine Learning” *European Intellectual Property Review* (n°1 - vol. 39 2017) available at <https://osf.io/preprints/lawarxiv/np2jd/>

To consider another example, there is the issue of potential legal personality rights for AI. Beyond the trivial “media stunt” example of the Saudi recognition of the robot Sophia as a legal person, this is an issue that has gathered much fictional speculation and commentary over the years. It is worth noting that, dependent on how such recognition was configured, it may not only have the negative implications discussed in various articles.⁷ Rather it could be viewed as a tool to impose general systems of regulation on AI and machine systems directly – similarly to the grant of legal personality rights to other artificial constructs such as companies.⁸ This is worthy of much more detailed consideration than this brief submission can do justice to, and is a topic that has been the subject of much debate and many proposals elsewhere.⁹ It may also have implications for narrower issues such as that of the copyright example given above, to the extent that an AI with legal personality might then be up for consideration as qualifying as an “author”.

These examples are provided not in any sense as an attempt to provide an exhaustive overview of areas of regulatory deficit/problems but as a small indication of some of the wide ranging, complex, long standing and global debates about many of these matters. Attempting to map and regulate these issues in one hit within our national boundary is likely to be an overly ambitious and fraught exercise.

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

My personal belief is that there is significant value in existing organisations (in business, research, education and training, civil society) considering and exploring the issues raised in the White Paper. However I don’t see a prima facie case for sufficient value being created by an RIO focussed discretely on the areas of AI/machine learning. Instead it may be possible to use existing organisations, grant schemes and other mechanisms to encourage further collaborative debate and link with similar efforts internationally.

A non-exhaustive list of existing organisations with some potential remit or activity in fields related to the topics raised by the White Paper include:

- Businesses developing and deploying AI and Consulting/Advisory service providers
- Government research agencies –including but not limited to CSIRO (Data61 and its partners such as the Gradient Institute)¹⁰
- Federal Government – from a purchasing, development and cross departmental policy development perspective
- Innovation Science Australia and its supporting office inside the Department

⁷ See e.g. Could an artificial intelligence be considered a person under the law?

<https://theconversation.com/could-an-artificial-intelligence-be-considered-a-person-under-the-law-102865>

⁸ See the following article for discussion as to how existing US company laws may already be used to give “effective” legal personhood to AI systems - Shawn Bayern, The Implications of Modern Business Entity Law for the Regulation of Autonomous Systems 19 STAN.TECH. L.REV. 93 (2015)

⁹ eg European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) at <http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P8-TA-2017-0051&language=EN&ring=A8-2017-0005#BKMD-12> (not taken up by the EC – see e.g. Thomas Burri, “The EU is right to refuse legal personality for Artificial Intelligence” at <https://www.euractiv.com/section/digital/opinion/the-eu-is-right-to-refuse-legal-personality-for-artificial-intelligence/>

¹⁰ <https://gradientinstitute.org/>

- Funding programs – as the AHRC will be aware in the 2018–2019 Australian budget, the government announced a 4 year investment to support the development of AI in Australia including a Technology Roadmap, a Standards Framework, and a national AI Ethics Framework (work on this being undertaken by Data61).
- Beyond core grant funding for the development of AI technologies¹¹ this may and should include funding for investigation of ethical and regulatory issues. At this stage examples of such funding seem fewer than those directed at AI enabled projects and there is room for more focus on ethical and regulatory issues - even if these are parts of other projects. One of the few existing examples of funding into AI ethics related matters is a project under the Supporting Responses to Commonwealth Science Council Priorities.¹²
- Learned Academies including Australian Academy of Technology and Engineering, Australian Academy of Science, Australian Academy of the Humanities
- Professional bodies such as the Australian Computer Society
- Law reform bodies including the Australian Law Reform Commission

The above are only a handful of the types of organisations within Australia that are relevant. Internationally there is of course a much larger body of work, and this will necessarily be much more influential than local efforts. Beyond the most directly influential actors (businesses developing and deploying AI and governments supporting/regulating activity) there are a wide array of think-tanks representing interested individuals, not for profits, research institutions and industry associations as AHRC/WEF will be aware. A few examples of many include:

- The Future of Life Institute¹³
- Institute for the Future (including tools it has created such as the Ethical OS Toolkit¹⁴ - and specific Labs within it such as the Digital Intelligence Lab¹⁵)
- Markkula Center for Applied Ethics¹⁶
- TechUK – e.g. its recent report *Digital ethics in 2019*¹⁷ **Note:** further to earlier discussion on the dangers of a narrow sectoral approach this report very deliberately broadened its focus beyond AI: “*techUK believes the digital ethics debate should widen from simply focusing on AI and look to identify issues that could arise from the long-term adoption and deployment of other advanced digital technologies whether used independently or in contributions with each other*”
- The Ada Lovelace Institute¹⁸ (under the Nuffield Foundation banner)

¹¹ e.g. Cooperative Research Centres grants as referenced at <https://www.abc.net.au/radio/programs/am/chief-scientist-calls-for-regulation-on-artificial-intelligence/9774302> and many examples of ARC funding for such projects e.g. “Driving innovation in next-gen AI medical technologies” <https://www.arc.gov.au/news-publications/media/media-releases/driving-innovation-next-gen-ai-medical-technologies>

¹² <https://rms.arc.gov.au/RMS/Report/Download/Report/d6b15b2b-3a50-4021-8e6f-6c7ef1cba553/0>

¹³ <https://futureoflife.org/>

¹⁴ <https://ethicalos.org/>

¹⁵ <http://www.iftf.org/partner-with-iftf/research-labs/digital-intelligence-lab/>

¹⁶ <https://www.scu.edu/ethics/focus-areas/technology-ethics/>

¹⁷

https://www.techuk.org/index.php?option=com_techuksecurity&task=security.download&file=techUK_Digital_Ethics_2019_paper_14834.pdf

¹⁸ <https://www.adalovelaceinstitute.org/>

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

My submission is made on a personal not an organisational basis. Notwithstanding my personal opinion it is conceivable that, dependent on how it was framed, an RIO could add value to Flinders University's teaching, research and external engagement efforts in particular as they relate to AI and machine learning. However it is not the only means to that end.

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

This is again begging the question of the need for such an organisation. However to make a deeper point, if something like a new RIO is desired, or alternatively if it is desired to create and explication of current issues to inform and guide the work of existing bodies to help improve co-ordination of Australia's approach to such matters, then I submit that the issues created by the widespread advent/adoption of AI/machine learning need greater exploration first before framing the business case, in accordance with business model design principles. In saying this I appreciate that the White Paper is in part attempting to do exactly that. However I want to reinforce that if this initiative is to be advanced further then it would benefit substantially from additional time spent on framing the issues first rather than trying to jump to quickly to a solution.

6. If Australia had a Responsible Innovation Organisation:

- (a) What should be its overarching vision and core aims?**
- (b) What powers and functions should it have?**
- (c) How should it be structured?**
- (d) What internal and external expertise should it have at its disposal?**
- (e) How should it interact with other bodies with similar responsibilities?**
- (f) How should its activities be resourced? Would it be jointly funded by government and industry? How would its independence be secured?**
- (g) How should it be evaluated and monitored? How should it report its activities?**

See response in relation to question 5 above – but made more strongly in relation to these points, which are begging the question too much – in particular as they relate to powers, functions, structures and methods of operation – these are the tools that can only be properly designed once there had been greater exploration of the issues. In my opinion trying to deal with some of these matters in the first stage of consultation is jumping the gun.

Having said that there are obviously some useful ideas listed under the functions section of the white paper including discussion in particular in relation to the role of education, best practice models and ethical guidelines. Simply collating and distilling the essence of the products of other organisations working on these matters globally would be a useful task for someone/some collaboration to undertake – even if this was not performed by a RIO per se.

Overall we would certainly benefit from the type of cross disciplinary collaborative work envisaged by the Gradient Institute:

“In research, technologists have to work in a true scientific expedition alongside experts in areas such as ethics, law, policy and behavioural sciences. Their collective

job is to clarify the complex web of causes and effects going from design choices for AI systems all the way up to the wellbeing of individuals and society.

In application, they have to work with businesses, governments and non-governmental organisations to deploy those AI systems for the benefit of the people served by these organisations. They also have to put their boots on the ground and engage consistently over time with government to build the trust required to help shape policy and regulation informed by the latest research findings.

In education, they must work with experts in other fields to support a holistic training of the next generation of ethical machine learning scientists and engineers, as well as actively engage in public forums to promote an informed and reasoned dialogue about AI.”¹⁹

All concerned citizens, not just technologists, need to engage in that debate and help shape the formal and informal checks and balances that will operate to help guide the evolution of our brave new world. And as Nicolas Suzor points out, this may well include new methods to “effectively marshal social pressure in day to day governance where the legal system is too cumbersome” that will help keep unbridled commercial interests in check, while at the same time resisting “the efforts of governments around the world to introduce new restrictions that unjustifiably limit our freedoms or threaten the conditions for autonomy and innovation”.²⁰ We cannot delegate that responsibility to any one institution.

¹⁹ <https://gradientinstitute.org/blog/1/>

²⁰ Nicolas Suzor, *Lawless: the secret rules that govern our digital lives* (Cambridge Uni Press, 2019) – conclusion.