University of New South Wales Law Research Series

RESPONSE TO 'ARTIFICIAL INTELLIGENCE: GOVERNANCE AND LEADERSHIP' WHITE PAPER CONSULTATION

NAYAN BHATHELA, LYRIA BENNETT MOSES, MONIKA ZALNIERIUTE, ROGER CLARKE, KAYLEEN MANWARING, KATHY BOWREY, FELICITY BELL AND MATTHEW KEARNES

[2019] *UNSWLRS* 15

UNSW Law
UNSW Sydney NSW 2052 Australia

E: unswlrs@unsw.edu.au
W: http://www.law.unsw.edu.au/research/faculty-publications
AustLII: http://www.austlii.edu.au/au/journals/UNSWLRS/
SSRN: http://www.ssrn.com/link/UNSW-LEG.html

13 March 2019

Response to 'Artificial Intelligence: governance and leadership' White Paper Consultation

About Us

The Allens Hub for Technology, Law and Innovation ('the Allens Hub') is an independent community of scholars based at UNSW Sydney. As a partnership between Allens and UNSW Law, the Allens Hub adds breadth and depth to research on the diverse interactions among technological change, law, and legal practice. The partnership enriches academic and policy debates and drives considered reform of law and practice through engagement with the legal profession, the judiciary, industry, government and the broader community. More information about the Allens Hub can be found at http://www.allenshub.unsw.edu.au/.

Introduction

This document seeks to respond to the questions raised by the Australian Human Rights Commission and World Economic Forum in their 'Artificial Intelligence: governance and leadership' White Paper consultation. It draws upon some of the research conducted at the Allens Hub to make suggestions about if and how systems involving some component of artificial intelligence or automation should be regulated. It also seeks to create further clarity regarding the potential structure and functions of the proposed Responsible Innovation Organisation. We note that our research does not relate to all the questions asked in the White Paper consultation, and so we only set out answers in relation to those matters where our research may be relevant. We are otherwise grateful for the opportunity to present our views and hope this submission will help the Australian Human Rights Commission and World Economic Forum in their consideration.

Primary author: Nayan Bhathela (student intern)
Participants/Researchers: Lyria Bennett Moses, Monika Zalnieriute, Roger Clarke, Kayleen Manwaring, Kathy Bowrey, Felicity Bell and Matthew Kearnes.

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

We believe the Australian Human Rights Commission ('AHRC') and World Economic Forum's ('WEF') question about 'government regulation' in the area of artificial intelligence ('AI') approaches important issues through a distorted lens.

Vague Scope of Regulation

Firstly, we suggest that the targeting of *AI* for regulation introduces a range of difficulties in defining the scope of what is actually the subject of regulation because the term 'artificial intelligence' has a vague meaning. For example, the Oxford Dictionary, defines AI as '[t]he capacity of computers or other machines to exhibit or simulate intelligent behaviour'.[1] The Macquarie Dictionary, on the other hand, defines AI as 'the ability of a computer or other device or application to function as if possessing human intelligence'.[2] Both of these definitions use vague language which makes it difficult to establish the scope of what is captured by the term.
In particular, we suggest that it is unclear how the AHRC and WEF's use of the term differs from *automation*, which is, for example, defined as 'the science of applying automatic control to industrial processes' in the Macquarie Dictionary.[3] Although the terms 'artificial intelligence' and 'automation in decision-making' are not mutually exclusive, they are clearly not the same, but the AHRC and WEF's reference to 'automated systems' and 'automation' alongside their use of the terms AI and 'machine learning' in the White Paper seem to conflate their meaning.[4] It is therefore unclear what the subject of the proposed regulation would be and the context in which such regulation would occur, which creates difficulties in being able to define the main goals of the proposed regulation. For instance, it is unclear if it would apply to a university assessment involving the development of a neural network for learning purposes.

Therefore, we believe that government regulation should be clear about the kinds of AI and automated technologies that are being targeted, and the context in which such regulation should occur. As some of our work has suggested, a more appropriate definition may be what Roger Clark has described as 'complimentary intelligence', which would include systems that '(1) do things well that humans do badly or cannot do at all; and (2) function as elements within systems that include both humans and artefacts, with effective, efficient and adaptable interfacing among them all'.[5] Alternatively, if we take applications such as the Centrelink 'robodebt' system in government administration, COMPAS in the criminal justice system and personalised pricing in commerce as examples of systems that should be regulated, then a term such as 'automation in decision-making in contexts with potential for adverse impact on an individual's human's rights' may also be a more accurate descriptor. These amendments would help to clarify the exact target of the regulation in question and would not encompass forms of AI or automation which ought fairly to be free of government oversight.

---

[1] *Oxford English Dictionary* (online at 27 February 2019) 'artificial intelligence'.
[2] *Macquarie Dictionary* (online at 27 February 2019) 'artificial intelligence'.
[3] *Macquarie Dictionary* (online at 27 February 2019) 'automation'.
[4] Australian Human Rights Commission and World Economic Forum, *Artificial Intelligence: governance and leadership* (White Paper, January 2019) 10, 12-3.
[5] Roger Clarke, 'Regulatory Alternatives for AI', *Roger Clarke's Web-Site* (Web Page, 9 February 2019) <http://www.rogerclarke.com/EC/RAI.html> [2.1].

Laws Should Be Changed

Secondly, we suggest that, rather than creating a body to regulate AI, it may be more helpful to begin by asking how the law might be changed to better protect individuals against some of the harms associated with AI where existing formulations are insufficient.[6] As the abovementioned example of student assessment illustrates, AI is not inherently in need of regulation. There are, however, ways in which the law fails to protect individuals against some of the harms associated with particular applications of AI. The question of how laws need to be *changed* is a better starting place than the question of how a particular technology ought to be governed or regulated, as this allows for the possibility that the best response to AI is better privacy laws, anti-discrimination laws, administrative law protections, and so forth.[7] Asking how the law needs to change also allows for the introduction of laws subjecting specific applications of AI (e.g. a law banning the exclusive use of automated systems for military purposes) to human rights obligations, but does not limit the solution space to such possibilities.

Furthermore, we suggest that treating AI as a targeted object for regulation is also problematic in the sense that, while this may achieve better policy coherence in relation to AI (assuming that the term is clearly defined), this hinders policy coherence more broadly. For example, just as there are debates in Australia about having special privacy laws for drones, consumer data, automated vehicles, etc., so too can special privacy laws to 'regulate' AI be introduced. However, it is far more useful and more coherent to ask what privacy law should be for *all* potential applications. Although privacy laws should work effectively for all AI applications, this does not mean that we need specific privacy laws regulating AI per se or an AI regulator to enforce them.

Regulatory Goals Should Be Contextual

Any regulatory or policy body created needs to understand the implications of AI in particular contexts. The most crucial context is the automation of decision-making with potential for negative impacts on individual lives. In particular, it is important to ensure that *appropriate* automation tools are used in the making of decisions or the drawing of inferences that become inputs into decisions. This was previously emphasised in the Allens Hub submission to the Issues Paper on Human Rights and Technology,[8] where it was remarked that AI tools need to be assessed at the micro-level, rather than at a vague macro-level.[9]

We also highlight the importance of factual accuracy and the empirical background in which regulation should be introduced. For example, the White Paper states that 'we have seen allegations of AI entrenching bias and discrimination in… policing in Australia' but the cited material does not confirm the use of AI in the design of the STMP system.[10] Inaccuracies such as these confirm our earlier point about the vagueness of the concept of AI and the fact that human rights issues may be more closely associated with a different category (such as government or law enforcement decision-making). In the example of STMP, it is not to the point that AI has the

---

[6] See Lyria Bennett Moses, 'Regulating in the Face of Sociotechnical Change' in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press, 2017) 573.

[7] See generally Bennett Moses, above n 6.

[8] The Allens Hub for Technology, Law and Innovation, Submission to Australian Human Rights Commission, *Human Rights and Technology Issues Paper*.

[9] Ibid 6.

[10] Australian Human Rights Commission and World Economic Forum, *Artificial Intelligence: governance and leadership* (White Paper, January 2019) 5.

capacity to entrench bias and discrimination – rather, it is biased and discriminatory decision-making that needs to be targeted, regardless of whether or not it includes an element that can properly be characterised as AI. A further example of this is the White Paper's mischaracterisation of the Centrelink 'robodebt' controversy as being based on 'the adoption of automated systems with efficiency as a primary goal'.[11] The 'robodebt' controversy stemmed mainly from Centrelink's implementation of a flawed algorithm which was subsequently automated without clear processes and explanations to deal with predictable exceptions, rather than from the fact of automation per se.[12] On this basis, it is clear that the manner in which AI issues are framed is vital in understanding how the law needs to be changed to protect against the harms which AI can cause.

In this respect, we suggest that it is also important to note that human rights law is not the only relevant framework for the regulation of AI. For instance, the interaction between AI and the rule of law brings with it another set of considerations, such as predictability and consistency, transparency and accountability, and equality before the law, that may overlap with, but are not completely identical to, human rights considerations.[13]

[11] Ibid 13.

[12] Monika Zalnieriute, Lyria Bennett Moses and George Williams, 'The Rule of Law and Automation of Government Decision-Making', (2019) 82(3) *Modern Law Review* (advance) <https://ssrn.com/abstract=3348831>.

[13] Ibid.

Question 2) Considering how artificial intelligence is currently regulated and influenced in Australia:

(a) What existing bodies play an important role in this area?

There are *many* organisations and collectives internationally that are involved in projects related to AI governance, AI standards and AI ethics. These include:

- The Artificial Intelligence, Ethics and Society conference;
- The Institute of Electrical and Electronics Engineers' (IEEE) Global Initiative on Ethics of Autonomous and Intelligent Systems and the P7xxx standards development program;
- The International Standards Organisation's JTC1/SC42 standardisation program;
- Standards Australia;
- The 'Artificial Intelligence Roadmap and Ethics Framework' project by Australia's Data61;
- The new Future AI Forum being convened by KPMG;
- The FAT conference, which has adopted its own Principles for Accountable Algorithms and a Social Impact Statement for Algorithms;
- Statement on Artificial Intelligence, Robotics and 'Autonomous Systems' by the European Group on Ethics in Science and New Technologies;
- French Data Protection Authority CNIL paper on rules for Artificial Intelligence;
- UNI Global Union Top 10 Principles for Ethical AI;
- Montreal Declaration for Responsible AI draft principles;
- ACM US Public Policy Council's Principles for Algorithmic Transparency and Accountability;
- The UK Centre for Data Ethics and Innovation; and
- The Algo Aware project of the European Commission.

Although the abovementioned organisations and collectives generally have no capacity to create enforceable laws, they have done considerable work in proposing regulatory frameworks in relation to the use of AI. Their recommendations can be implemented in conjunction with other legislation. For example, JT 1 standards have been recognised by Standards Australia, which is the peak non-governmental standards-writing body recognised by the Australian Commonwealth and adopted by Australian regulators, including the Australian Communications and Media Authority. As such, any new work by the AHRC and WEF is entering a very noisy space where common themes are abundant (such as transparency, accountability and fairness, etc.) but agreed formulations of principles are rare. At the moment, each of the abovementioned bodies, conferences and organisations are working independently. There is little value in adding to the noise in the absence of an attempt to bring existing efforts together.

There is a small body of literature which the AHCR and WEF may find useful on transnational industry rule-making and the mechanisms that are used by transnational industry actors in an attempt to render their rule-

making processes legitimate.[14] Much of this literature is relevant to consider to the AHRC and WEF agenda as they seek to improve the accountability and responsiveness of private rule-making in the interests of democratic accountability and public policy. New initiatives need to work alongside, rather than seek to replace or add to activity in this emerging research space.

(b) What are the gaps in the current regulatory system?

In addition to the principles, standards, roadmaps etc. mentioned in our response to Question 2(a) above, general commercial laws relating to contracts, consumer rights, directorial responsibility, copyright and patents, as well as anti-discrimination and other human rights laws, could apply to AI technologies to the extent that the AI systems to which they apply interact with such laws. However, some of these laws would need to be amended in order to effectively address the challenges posed by AI given the outdated sociotechnical context within which they have been designed.[15] For example, in the context of digital consumer manipulation, see the work of Allens Hub member Kayleen Manwaring.[16] Moreover, at least for the moment, the state-centred human rights legal framework fails to cover private actors,[17] and economic incentives largely act against *voluntary* protection of human rights by informal and non-state actors in the digital age.[18]

The exercise of identifying gaps in the current regulatory system would, in Australia, normally fall to a law reform commission. There are advantages to these kinds of permanent bodies, including institutional memory in identifying legal issues from previously 'new' technological contexts (such as genetic privacy). There are also disadvantages, including a primarily legal focus. However, the solution could be to combine some of the benefits of law reform commissions with those of technology assessment bodies to create a new standing body focussing on legal and regulatory responses to technological change.[19]

---

[14] See, eg, Karen Lee, *The Legitimacy and Responsiveness of Industry Rule-making* (Hart Publishing, 2018); Sigrid Quack, 'Law, expertise and legitimacy in transnational economic governance: An Introduction' (2010) 8(1) *Socio-Economic Review* 3; Kristina Tamm Hallström, *Organising International Standardisation: ISO and the ISAC in Quest of Authority* (Edward Elgar Publishing, 2004); Nicolas Hachez and Jan Wouters, 'A Glimpse at the Democratic Legitimacy of Private Standards: Assessing the Public Accountability of GlobalG.A.P (2011) 14(3) *Journal of International Economic Law* 677; Klaus Dingwerth, *The New Transnationalism: Transnational Governance and Democratic Legitimacy* (Palgrave MacMillan, 2007).

[15] Roger Clarke, 'Guidelines for the Responsible Business Use of AI Foundational Working Paper', *Roger Clarke's Web-Site* (Web Page, 20 February 2019) <http://www.rogerclarke.com/EC/GAIF.html> [5.3].

[16] Kayleen Manwaring, 'Will Emerging Information Technologies Outpace Consumer Protection Law? The Case of Digital Consumer Manipulation' (2018) 26 *Competition and Consumer Law Journal* 141.

[17] See Monika Zalnieriute and Stefania Milan, 'Internet Architecture and Human Rights: Beyond the Human Rights Gap' 11(1) *Policy & Internet* 6 <https://onlinelibrary.wiley.com/doi/10.1002/poi3.200>.

[18] See Monika Zalnieriute, 'From Human Rights Aspirations to Enforceable Obligations by Non-State Actors in the Digital Age: The Example of Internet Governance and ICANN' 21 *Yale Journal of Law & Technology* (forthcoming) <https://ssrn.com/abstract=3333532>.

[19] See Lyria Bennett Moses, 'Bridging distances in approach: Sharing ideas about technology regulation' in Ronald Leenes and Eleni Kosta (eds), *Bridging distances in technology and regulation* (Wolf, 2013) 37.

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

In our opinion, to ask for a 'business case' for a policy intervention is to uncomfortably mix opposing ideas, in that it places more emphasis on demonstrating feasibility or profitability rather than a broader public good. The question should thus really be whether a Responsible Innovation Organisation has an important policy role to play. There are various roles such an organisation could play, some of which were discussed at the 5 March 2019 workshop. These include the following:

- A body that analyses gaps in existing law and proposes changes (analogous to a law reform commission, but created for the specific context of AI, likely with greater in-house technical expertise);
- An advocacy organisation shining light on AI-related human rights issues;
- A regulatory agency, with powers to ensure that AI meets particular standards (with powers of enforcement);
- A body offering practical advice on best practice and compliance with existing law (perhaps analogous to the standards being developed by IEEE but with an Australian focus);
- A body that conducts engagement and education campaigns with government, industry, civil society and the broader community in order to feed into AI-related policy discussions (similar to the Science and Technology Engagement Pathways (STEP) framework that has been proposed by the National Enabling Technology Strategy – Public Awareness and Community Engagement program (NETS-PACE) with respect to nanotechnology and other 'enabling' technologies of the previous decade);[20] or
- A body that promotes the development of an AI industry in Australia.

Before asking about the benefits of any of these potential roles, it is important to understand the rationale for creating a new body to undertake one or more of these roles and, in particular, the limitations of existing bodies that might otherwise fulfil them. It is also important to recognise that one body cannot necessarily achieve *all* of the abovementioned goals and there are conflicts in having one body be both a regulatory agency and a promoter of industry.[21]

[20] See Department of Industry, Innovation, Science, Research and Tertiary Education, *Science and Technology Engagement Pathways (STEP): Community involvement in science and technology decision making* (2010).

[21] See also Graham Greenleaf, Roger Clarke and David Lindsay, 'Does AI need governance? – The potential roles of a 'Responsible Government Organisation' in Australia (Submission to the Australian Human Rights Commissioner on the White Paper *Artificial Intelligence: Governance and Leadership*', *SSRN* (Web Page, 6 March 2019) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3346149> 1.

Question 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?

We address questions 6(a),6(b), 6(c) and 6(d) together as they are clearly linked, and all depend on the type of organisation being proposed.

There are various choices that need to be made at the outset as to the type of organisation in question. One possibility, as suggested above, is a body that takes the benefits of a law reform commission but enhances those by bringing in some insights from technology assessment processes, including the need for deep technical expertise that has the capacity to undertake 'systematic expert valuation of technological possibilities to determine benefits as well as potential harm' of AI in the reform process.[22]

On the other hand, if what is sought is a regulatory body, then it may be useful to consider Roger Clarke's Co-Regulation framework, in which he proposes that a code negotiation and maintenance institution should have the power to establish a code (which is to be accepted either by an independent Commissioner or a Minister in accordance with a power to do so delegated by Parliament) that regulates industry behaviour in accordance with a consultative process with other stakeholders, and a *separate* regulatory organisation should have the following powers and functions:

- Overview of consultative processes;
- Supervision of compliance with the code;
- Conduct of own-motion and complaint-based investigations;
- Imposition of penalties on those who infringe the code;
- Prosecution of offenders;
- Research into technological and environmental changes;
- Provision of an information clearing house; and
- Provision of a focal point for adaptation of the law.[23]

In accordance with Clarke's proposed framework, the code negotiation and maintenance institution should rely on comprehensive consultative processes being undertaken with various stakeholders, which include those

[22] Lyria Bennett Moses, 'Bridging distances in approach: Sharing ideas about technology regulation' in Ronald Leenes and Eleni Kosta (eds), *Bridging distances in technology and regulation* (Wolf, 2013) 37, 41, 47.
[23] Roger Clarke, 'Regulatory Alternatives for AI', *Roger Clarke's Web-Site* (Web Page, 6 March 2019) <http://www.rogerclarke.com/EC/RAI.html> [4].

entities being regulated, as well as the affected public.[24] Varying levels of external technical expertise is likely to be made available to the code negotiation and maintenance institution from both of these groups through the consultative process. Such a body would also benefit from permanent internal expertise by part-time technical experts and ad hoc committees of experts.[25]

Furthermore, if what is proposed is an organisation to conduct consultations in order to feed industry, government, civil society and broader community concerns into a policy process, then a useful starting place is to consider work that has already been done by the government in this area. In particular, the STEP framework, as developed by the NETS-PACE program, provides a useful analysis of how such public consultation could be undertaken.[26] This form of community consultation is not currently captured in the White Paper. Further inspiration can be taken from the Matthew Kearnes et al submission to the Australian Council of Learned Academies report on 'The Future of Precision Medicine in Australia', which suggests, *inter alia*, that deliberative methods which engage in future trajectories of new technologies and approaches (such as deliberative polling, deliberative and multi-criteria mapping, and consensus conferences) are useful public engagement strategies.[27] These suggestions can be applied directly in the context of AI.[28]

[24] Roger Clarke, 'Regulatory Alternatives for AI', *Roger Clarke's Web-Site* (Web Page, 6 March 2019) <http://www.rogerclarke.com/EC/RAI.html> [3.1], [4].

[25] See also Graham Greenleaf, Roger Clarke and David Lindsay, 'Does AI need governance? – The potential roles of a 'Responsible Government Organisation' in Australia (Submission to the Australian Human Rights Commissioner on the White Paper *Artificial Intelligence: Governance and Leadership*', *SSRN* (Web Page, 6 March 2019) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3346149> 5.

[26] Department of Industry, Innovation, Science, Research and Tertiary Education, *Science and Technology Engagement Pathways (STEP): Community involvement in science and technology decision making* (2010).

[27] Matthew Kearnes et al, 'Public Engagement', *Australian Council of Learned Academies: The Future of Precision Medicine in Australia* (Web Page) <https://acola.org.au/wp/wp-content/uploads/IP-3-Public-engagement.pdf>.

[28] See also Jason Chilvers and Matthew Kearnes (eds), *Rethinking Participation: Science, Environment and Emerging Publics* (Routledge, 1st ed, 2016); Declan Kuch et al, 'Five Lessons from the #MyHealthRecord discussion for #PrecisionMedicine', *Medium* (Web Page, 13 August 2018) <https://medium.com/@declank/five-lessons-from-the-myhealthrecord-discussion-for-precisionmedicine-96821aba8a59>; Declan Kuch et al, 'An Energy Data Manifesto' in Angela Daly, Kate Devitt and Monique Mann (eds) *Good Data* (INC Press, 2019) 77.