Centre for Policy Futures
8 March 2019

THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Submission to Australian Human Rights Commission and World Economic Forum’s ‘Artificial Intelligence: governance and leadership’ White Paper

Responses to Select Consultation Questions

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

Government regulations of AI and similar digital decision-making tools should be directed towards upholding the values and principles of human rights and citizen/consumer protections already in place when engaging with government, non-government and commercial sectors.

It is important to acknowledge that there is an important distinction in the dynamics of these rights and protections, and expectations of regulation, between businesses and government (and organisations delivering government funded services). As citizens, there is a higher expectation and requirement for accountability and transparency by the state, as it operates on behalf of the people. In contrast, the modus operandi of businesses is for making a profit.

We wish to draw attention to three key areas in which government regulation needs to occur, how it may occur, the principles, and challenges in operationalising them.

- **Protection from discrimination** based on sex/gender, race/ethnicity, religion, sexuality, etcetera is an important human rights principle. A challenge for AI (and other algorithmic decision making) is that what counts as discrimination and what counts as appropriate differential treatment becomes harder to harder to conceptualise and detect. To illustrate, consider a judge that consistently gives higher sentences to indigenous youths compared to anglo youths for similar crimes. This would suggest bias and discrimination. However, if an algorithm (or AI) based on statistical analysis (or big data) demonstrates that indigenous youths were more likely to be recidivists, then such 'scientific', 'objective' knowledge could justify differential treatment, which is held to be 'non-discriminatory'. It could be counter argued that the algorithm is biased because it has been trained on past statistics, which are an outcome of bias. Understanding the difference between when differential treatment is valid, and when it is discrimination is scarcely been tackled (but see Henman 2005). The reality is that such differential treatments on the basis of data, which reinforce (not ameliorate) inequalities, is widespread. We know that online markets such as Amazon and TripAdvisor differentiate prices by sex/gender, postcode, and device used. We do not know if they use other social categories (e.g. race/ethnicity, religion, sexuality) to differentiate prices. In rental tenancy databases used to assess potential 'quality' of potential renters, it is suspected that age and race/ethnicity is used. The difficult regulatory question is 'are these forms of differentiation discrimination?'. This is where a distinction between commercial and government needs to occur. As citizens there is an expectation of equality, unless it can be justified, whereas in the commercial sector, the drive to maximise profits may override a concern that such practices are unfair and discriminatory. Accordingly, the onus of proof of and the threshold test for discrimination may need to be different for these different sectors.
- The principle of freedom of speech has been an important principle in liberal democracies, including Australia. In the age of disinformation and 'fake news', as well as rise of online bullying, there is often a refrain that regulating free speech is a breach of human rights. It is widely recognised that human rights are at time in tension with each other, and drawing a balance between competing rights is an important and ongoing program. In relation to free speech, it is helpful to draw a parallel with the principle of free markets that also underpin liberal democratic capitalist societies, such as ours. Just as it is widely recognised that within free market societies it is acceptable and appropriate to regulate free markets – for example, there regulations about food standards and labelling, requirements about false advertising, and restrictions on advertising in some industries (e.g. tobacco, pharmaceuticals) – to enhance economic and social outcomes, so too do we need to regulate free speech.

- **The right to an explanation, appeal, correction and redress**. State accountability and transparency – right to appeal. Responsibility and accountability. Ever since governments and businesses started using digital tools for automating decision making there have emerged problems in law arising from citizen and consumer rights in understanding these decisions, seeking an appeal and receiving correction and redress. For example in the early 1980s, a computer of the then Department of Social Security automatically cancelled benefit payment to a beneficiary, as a form was not recorded as being returned. The case was appealed and went through the Social Security Appeals Tribunal, the Australian Administrative Appeals Tribunal and eventually the Federal Court. The final legal decision was that the decision to cease payment was not correct, but no remedy was available as the decision was deemed not appealable because a computer, and not a human acting on behalf of the Secretary, made it. The law in this case, and in similar occasions over time, was subsequently amended to ensure that computer made decisions are regarded as similar to human made ones. This example is repeated in recent debates about AI, particularly as machine learning algorithms decrease the level of transparency and accountability, with the EU acknowledging a right to explanation in its General Data Protection Regulation (Edwards & Veale 2017; Watcher et al 2017). Recognising and operationalising a right to an explanation for automated decisions, alongside associated right to appeal, correction and redress is particularly pertinent in public services (whether delivered by government or non-government organisations) as it reinforces the principle of government accountability and transparency within liberal democracies. It is also important for customers of commercial decisions to ensure a well-functioning free market.

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?

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

We agree that there will be significant economic and social value for Australia in creating a Responsible Innovation Organisation (RIO). The Commission has provided a strong outline in its White Paper of the significant benefits of having such an Organisation and costs of not having one. The urgent need for such an Organisation is due to the continuous and rapidly evolving nature of new technologies, which constantly challenge and disrupt government's policy and regulatory settings, as well as long-standing and well-respected policy and legal principles. Without a proactive involvement of regulators, policy, social and ethics researchers and policy makers, new technologies have had a tendency to force change without considered public discussion and deliberation about whether what it is changing is beneficial or not.

The case of the shared/gig economy – such as Uber, Amazon's Mechanical Turk and Freelancer – provides an illustration. These digital platforms have generated new opportunities and opened up access to a wider range of people seeking more flexible work arrangements and avenues to unlock their abilities and assets. At the same time, they have variously challenged labour laws on minimum wages, blurred the boundaries between employees and self-employed individuals, and (in the case of Uber) have operated illegally in some jurisdictions in competition to taxi services.

However, we recommend that a RIO needs to extend its remit beyond AI and Machine Learning (ML) to fully capture the economic and social value it can offer to Australia. The remit of a RIO needs to be extended in three ways:

1. The Organisation needs to oversee non-AI/ML digital technologies and tools. Many of the issues that are now being grappled with in relation to AI (e.g. privacy, data protection, inequality, bias, discrimination, accountability, transparency) also relate to more standard forms of human coded

algorithms (e.g Denrick et al 2018; Eubanks 2018; Henman 2005; Redden 2018). For example, for over a decade many child protection services in Australia and globally have made use of risk assessment tools built on the basis of statistical analysis and knowledge of risks apparently associated with child abuse or neglect. These systems have variously been critiqued along the same lines that people express concerns about introducing AI based risk assessment into child protection systems (Eubanks 2018; Gillingham 2017; 2019). Introducing AI based tools does not fundamentally change these human rights concerns, however, they do decrease the transparency of such systems from what is already a very low level. Centrelink's Robodebt is another example of a similarly basic, non-AI algorithm that has generated significant human rights challenges (Henman 2017; Carney 2018). Consequently, the issues that a RIO will deal with in relation to AI have applicability to pre-AI digital tools, and a focus on the latter with all the previous research and studies in this area helps to inform how to respond to emerging uses of AI.

2. A RIO needs to extend its reach to oversee non-digital technologies, such as genetic technologies and synthetic biology. A plethora of new technologies and tools are rapidly being developed. While originally based within particular scientific fields, the boundaries are blurring, with digital tools and big data, becoming enmeshed with genetic technologies, bio-medical and other sensors, and forms of manufacture. AI will be increasingly involved in doing research within these fields and also interlinked with resulting technologies. As such it is unnecessarily limiting to have a RIO focusing entirely on AI.
3. A RIO needs work with science research funding agencies to catalyse innovation in responsible innovation, in areas such as but not limited to explainable AI, adversarial machine learning, continuous monitoring supporting data governance in cloud computing (Ko 2014) and distributed environments (e.g mobile devices, IoT), and accountable computing systems (Ko et al 2011). We encourage the engagement and discussions with the academic community, which are at the forefront of developing algorithms for AI and other technologies.

Responsible Innovation provides a framework and approach that readily informs a human rights based organisation addressing the challenges and problems, and the regulatory and policy implications, arising from AI and other emerging technologies.

> *Responsible innovation acknowledges the power of innovation to create the future (and associated with these uncertainties) and asks how we can and should meaningfully engage as a society with the futures innovation seeks to create, futures that are being created unintentionally or by design.* (Owen and Pansera 2018)

New technologies are continuously and rapidly emerging, constantly giving rise to new human rights, policy, and regulatory challenges.

Legislation backing the RIO is key. For example, the RIO would require law and policy makers to develop next generation legislation allowing the regulation. For example, the state of California in the USA has recently made attempts to regulate in the responsible innovation space through, for example:

- A law making it "unlawful for any person to use a bot to communicate or interact with another person in California online with the intent to mislead the other person about its artificial identity for the purpose of knowingly deceiving the person about the content of the communication in order to incentivize a purchase or sale of goods or services in a commercial transaction or to influence a vote in an election." (http://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001)
- A bill beginning on January 1, 2020, that "would require a manufacturer of a connected device, as those terms are defined, to equip the device with a reasonable security feature or features that are appropriate to the nature and function of the device, appropriate to the information it may collect, contain, or transmit, and designed to protect the device and any information contained therein from

> unauthorized access, destruction, use, modification, or disclosure, as specified." (https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB327)

From another angle, standards and design guidelines are also critical in supporting and guiding the best practices promoted by the RIO. There should be consideration for a tiered approach to levels of assurance instead of a one-dimensional approach commonly found in several IT standards and regulations. The RIO guidelines should feature layered frameworks such as those in the cloud computing industry, e.g. Cloud Security Alliance's STAR Registry, and the Singapore government's Multi-tiered Cloud Security standard (https://cloudsecurityalliance.org/star/#_overview; https://www.imda.gov.sg/industry-development/infrastructure/ict-standards-and-frameworks/mtcs-certification-scheme)

That said, the standards, frameworks and legislations should not be developed in an overly-prescriptive way which will then impede the pace of innovation - channeling potential innovation overseas to countries with lesser barriers to entry. The case of Facebook removing shifting users from the EU to the USA following the execution of the GDPR is an example (https://www.theregister.co.uk/2018/04/19/facebook_shifts_users/)

In summary, Australia will benefit from a RIO that has wide remit to encourage and manage innovation to strong economic and social outcomes while upholding human rights.

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

No response.

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

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?

The White Paper articulates a Responsible Innovation Organisation (RIO) that has strong ombudsman, complaints handling and investigatory roles. We argue that such an approach, with its significant resourcing requirements, will lead to a quite large organisation which will likely slow the overall responsiveness to emerging challenges with human rights and new technologies. We suggest a different organisational model and focus is more beneficial.

Given the rapid nature of new and emerging technologies**, we propose a lean and nimble organisation, based on a start-up culture**. A government-based start-up organisational model can is evidenced in a number of digital units operating in governments (e.g. 18F and United States Digital Services (USDS) in the USA; Canadian Digital Service; Australia's Digital Transformation Office[1]). These units operate with small numbers of staff (often a few dozen to 100). They are designed to provide expertise that support and assist government agencies either by providing advice to a government agency about their digital tools or strategy, or come into an organisation to work collaboratively with agency staff for brief periods. They are discrete project focused in operation. Their architectural settings are designed as highly interactive and flexible. Employers are often entrepreneurial and operate cross-disciplinary. For example, the USDS states:

> *USDS deploys small, responsive groups of technology experts to work with and empower civil servants. These multi-disciplinary teams bring best practices and new approaches to untangle some*

[1] https://18f.gsa.gov/, https://www.usds.gov/, https://digital.canada.ca/. https://www.dta.gov.au/

> *of our nation's most important problems. Our staff comes from all corners of the technology industry, nonprofit world, and government to serve 'tours' of service, bringing a steady influx of fresh perspectives into government. Tours typically last between six months and two years, with a maximum length of four years. Most of our staff have backgrounds in design, engineering, or product management. We also hire strategists, recruiters, procurement experts, attorneys, communications specialists and others* [2]

We believe such a start-up model is a very fruitful model for a proposed Responsible Innovation Organisation.

Its remit would focus on more on the preventative end of human rights and responsible innovative at the front-end of innovation and adoption, rather than as a compliance/regulatory agency. It would do this in various ways.

- Engagement with government, academia, industry and civil society to help work through the human rights, responsible innovation and policy/regulatory issues associated with developing and deploying new tools *in situ*.
- By being involved at this more front-end of the innovation and adoption process, the Organisation would be very well placed to identify policy and regulatory issues and provide advice to government on how they may be addressed.
- A strong educational and community consultation role around new and emerging technologies, would help facilitate public acceptance of innovations that have been developed with an informed understanding of community attitudes and responsible innovation principles.
- Working with innovation funding agencies and funding responsible innovation such as explainable AI, the Organisation would be able to co-develop protocols and frameworks, an analytical toolset, for guiding and assessing AI and other innovations. This would include computer scientists engaged in reverse engineering and 'bias testing' AI/algorithms, especially propriety algorithms to evaluate their consistency with responsible innovation and human rights and AI principles.
- Designing and instituting a 'certification scheme' (such as a 'Turing Stamp') for human rights compliant AI development and use, and issuing such certificates based on the RIO's assessment (which could include reverse engineering and bias testing as mentioned above).
- It would **not** operate as a compliance or regulatory organisation, nor one that investigates complaints or breaches, or determines liability and issues fines. We believe that such regulatory and investigatory roles are best located within current organisations under current (or expanded) arrangements. For example, issues of privacy should be regulated and investigated by an appropriately resourced *Office of the Information Commissioner* and equivalent state bodies. False and misleading advertising or unconscionable practices by businesses should be incorporated within the *Australian Competition and Consumer Commission* and equivalent state bodies. Considerations of matters relating to political advertising and targeting would fall under the state and federal electoral commissions. Similarly, emerging health technologies could be dealt with through the *Therapeutic Goods Administration* (with appropriate extensions for what may fall through their current remit).

We envisage a relatively small non-hierarchical organisation, a Director, with short-term input and leadership from specialist part-time Commissioners relating to specific projects or programs, international expert advisors and a Board reflective of leading government, commercial and civil society stakeholders. It would be a statutory independent body reporting to Parliament, not to a specific Minister. Accordingly, like the Parliamentary Library, it would be able to provide advice to all Parliamentarians and Senators when sought, while publishing all such advice.

[2] https://www.usds.gov/how-we-work

(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?

About UQ's Centre for Policy Futures and Authors

The University of Queensland, Centre for Policy Futures

Created in 2017, The University of Queensland's Centre for Policy Futures (CPF) aims to enhance the University's position as a key source of ideas and insights on the policy priorities that matter to Australia and the Pacific region. It does this through robust, rigorous and timely research and sustained policy engagement. The Centre's researchers, affiliated senior associates and visiting fellows pursue a vibrant research program focused on independent and peer-reviewed research, as well as commissioned reports, discussion papers, and policy briefs. Working closely with governments, international organisations, and key stakeholders, the Centre specialises in three policy areas:

- Science, Technology and Society
- Sustainable Development Goals and Capacity- Building
- Trade, Foreign & Security Policy

In addition to its research program, the Centre provides policy engagement and studies, as well as executive education involving academics across UQ and beyond. This approach enables the Centre to be flexible and responsive to policy matters as they arise.

The Centre is leads a multi-million dollar **CSIRO-UQ research collaboration on responsible innovation**. This work covers questions of regulation relating to a wide range of emerging technologies, including AI and digital technologies, synthetic biology and DNA manipulation, hydrogen and nuclear energy cycles, and health monitoring and detection technologies. At UQ, this collaboration involves a Principal Research Fellow, a Postdoctoral Research Fellow for Digital Human Rights, a Postdoctoral Research Fellow on the governance and regulation of synthetic biology, and eight PhD students involved in various projects relating to responsible innovation of new and emerging technologies being developed by CSIRO.

Associate Professor Paul Henman

Paul Henman is Associate Professor of Digital Sociology and Social Policy, School of Social Science, and Principal Research Fellow, Centre Policy Futures at the University of Queensland. In the latter role is leads the Science, Technology and Society research program, and the CSIRO-UQ Responsible Innovation partnership. As outlined below, he is ideally placed to provide expert advice into this White Paper process.

Paul has over 20 years of active research interest in digital technologies and public governance. His research covers the use of digital technologies by government for the operation of government (including policy making, service delivery, governance of agencies), as well as the use of digital technologies for governing and governance. Whilst Paul's research has focused on governments' use of digital technologies, his work also provides insights for the private and NGO sectors.

In particular, Paul's research has investigated the ways in which new digital technologies have shaped the types of policy and services that can be and are enacted. His work predates current concerns about algorithms in profiling and targeting by over a decade. In the early 2000s, he identified the policy, social and ethical dynamics associated with digital technologies' disruption of public policy and administration principles, often leading to increased inequalities (e.g. Henman 1997; 1999; 2002; 2004; 2006; 2010; Henman & Adler 2003)

Significantly, Paul's research rests on interdisciplinary training in computer science (holding an award winning first class honours degree, 1989), and in sociology of technology and social policy (PhD, 1996). This has provided him with insights not typically open to people without such interdisciplinary training. To date, he has received almost $3 million in research funding, including from the Australian Research Council, IBM, and the former National Office for the Information Economy. He has published over 4 books and over 70 academic papers. He is currently lead an international comparative study of government web portals in 10 countries.

Importantly, Paul has also worked in government as a policy analyst (1996-99) thereby providing him with important insights into the way in which governments operate. Consequently, he has regularly contributed to government and independent inquiries regarding regulation of new technologies, including the Australian Law Reform's 2003 inquiry into genetic testing, the 2009 *Government 2.0 Taskforce*, and the Parliamentary Joint Committee on Intelligence and Security *Identity-matching Services Bill 2018* Inquiry.

Professor Ryan Ko

Professor Ryan Ko is Chair and Director of UQ Cyber Security at the University of Queensland, Australia. His applied research in cyber security focuses on 'returning control of data to cloud computing users'. His research reduces users' reliance on trusting third-parties and focusses on (1) provenance logging and reconstruction, and (2) privacy-preserving data processing (homomorphic encryption). Both his research foci are recognised nationally and internationally, receiving conference Best Paper Awards (2015, 2017), and technology transfers locally and internationally.

Prior to academia, he was a lead computer scientist with Hewlett Packard Labs where his innovation on cloud data provenance and data accountability were commercialised into HP ArcSight security information and event management (SIEM) products – deployed in critical infrastructure worldwide, including the USA Treasury, IRS and the Singapore government cloud.

He serves as Technology Advisory Board member of the NZX-listed (NZE:LIC) Livestock Improvement Cooperation (LIC), Nyriad, and expert advisor to INTERPOL, NZDF, NZ Minister for Communications' Cyber Security Skills Taskforce, and one of four nationally-appointed Technical Adviser for the Harmful Digital Communications Act 2015, Ministry of Justice.

Within the ISO/IEC JTC 1/SC 27 technical committee, Prof Ko served as Editor, ISO/IEC 21878 "Information technology -- Security techniques -- Security guidelines for design and implementation of virtualized servers", and hosted the ISO/IEC JTC 1/SC 27 meetings at Hamilton, New Zealand, in 2017. He has published more than 100 publications, including books, refereed conference papers, journal papers, book chapters, encyclopaedia entries, technical reports and international patents (PCT). He served in technical programme committees for more than 30 IEEE conferences/workshops, associate editor for 6 journals, and series editor for Elsevier's security books.

For his contributions to the field, he was elected Fellow of Cloud Security Alliance (CSA) (2016), the Singapore Government (Enterprise Singapore)'s Young Professional Award (2018), and awarded the inaugural CSA Ron Knode Service Award 2012. He is also recipient of the 2015 (ISC)2 Information Security Leadership Award.

References

Carney, T. (2018). Robo-debt illegality: The seven veils of failed guarantees of the rule of law?. *Alternative Law Journal*, 1037969X18815913.

Dencik, L., Hintz, A., Redden, J., & Warne, H. (2018). *Data scores as Governance: Investigating uses of citizen scoring in public services project report*. http://orca.cf.ac.uk/117517/1/data-scores-as-governance-project-report2.pdf.

Edwards, L., & Veale, M. (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke L. & Tech. Rev.*, *16*, 18.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Gillingham, P. (2017). Predictive risk modelling to prevent child maltreatment: insights and implications from Aotearoa/New Zealand. *Journal of public child welfare*, *11*(2), 150-165.

Gillingham, P. (2019). Decision Support Systems, Social Justice and Algorithmic Accountability in Social Work: A New Challenge. *Practice*, 1-14.

Gillingham, P., & Humphreys, C. (2009). Child protection practitioners and decision-making tools: Observations and reflections from the front line. *British Journal of Social Work*, *40*(8), 2598-2616.

Henman, P. (1997). Computer technology–a political player in social policy processes. *Journal of Social Policy*, 26(3), 323-340.

Henman, P. (1999). The bane and benefits of computers in Australia’s Department of Social Security. *International journal of sociology and social policy*, 19(1/2), 101-129.

Henman, P. (2002). Computer modeling and the politics of greenhouse gas policy in Australia. *Social science computer review*, 20(2), 161-173.

Henman, P. (2004). Targeted! Population segmentation, electronic surveillance and governing the unemployed in Australia. *International Sociology*, 19(2), 173-191.

Henman, P. (2005). E-government, targeting and data profiling: policy and ethical issues of differential treatment. *Journal of E-government* [now *Journal of Information Technology & Politics*], *2*(1), 79-98.

Henman, P. (2006). Segmentation and conditionality: technological reconfigurations in social policy. In C MacDonald and G Marston (eds) *Analysing social policy: A governmental approach*, Basingstoke: Palgrave.

Henman, P. (2010) *Governing Electronically: E-government and the reconfiguration of policy, public administration and power*, Basingstoke: Palgrave.

Henman, P. (2017). The computer says ‘DEBT’: Towards a critical sociology of algorithms and algorithmic governance, *Data for policy conference*, London, https://zenodo.org/record/884117#.WcTlEsh97IU

Henman, P., & Adler, M. (2003). Information technology and the governance of social security. *Critical Social Policy*, 23(2), 139-164.

Ko, RKL (2014), “Data accountability in cloud systems”, Chapter in *Security, Privacy and Trust in Cloud Systems,* pp. 211-238*,* Springer, Berlin, Heidelberg, https://link.springer.com/chapter/10.1007/978-3-642-38586-5_7

Ko, RKL, Lee, Bu Sung & Pearson, S. (2011) “Towards achieving accountability, auditability and trust in cloud computing”,Proceedings of International Conference on Advances in Computing and Communications, pp. 432-444, Springer, Berlin, Heidelberg, 2011/7/22, https://link.springer.com/chapter/10.1007/978-3-642-22726-4_45

Owen, R. & Pansera, M. (2018) Responsible Innovation and Responsible Research and Innovation, in: Kuhlmann, S., Canzler, W., Simon, D, Stamm, J. (Eds.) *Handbook of Science and Public Policy*, Edgar Elgar

Redden, J. (2018). Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society*, *5*(2), 2053951718809145.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. *Science Robotics*, *2*(6).

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, *7*(2), 76-99.

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Contact details

Associate Professor Paul Henman
T +61 7 **3443 3142**
M +61 **402 734 218**
E p.henman@uq.edu.au
W https://policy-futures.centre.uq.edu.au/