University of New South Wales Law Research Series

ARTIFICIAL INTELLIGENCE: AUSTRALIA'S ETHICS FRAMEWORK SUBMISSION TO THE DEPARTMENT OF INDUSTRY, INNOVATION AND SCIENCE

MONIKA ZALNIERIUTE AND OLIVIA GOULD-FENSOM

[2019] *UNSWLRS* 40

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

Artificial Intelligence: Australia's Ethics Framework

Submission to the Department of Industry, Innovation and Science

Dr Monika Zalnieriute
Olivia Gould-Fensom
Allens Hub for Technology, Law and Innovation
UNSW Faculty of Law, Sydney, Australia

About us

We are scholars working at the Allens Hub for Technology, Law and Innovation ('the Allens Hub') - an independent community of scholars based at UNSW Sydney. As a partnership between Allens and UNSW Law, the Allens Hub aims to add depth to research on the diverse interactions among technology, law, and society. The partnership aims to enrich academic and policy debates and drive considered reform of law and practice through engagement with the legal profession, the judiciary, government, industry, civil society and the broader community. More information about the Allens Hub can be found at http://www.allenshub.unsw.edu.au/. The opinions expressed in this submission are the views of the authors, and not reflect or present the views or positions of the Allens Hub or UNSW Law.

About this submission

This submission seeks to respond to the questions raised in the Department of Industry, Innovation and Science's consultation hub regarding 'Artificial Intelligence: Australia's Ethics Framework' discussion paper prepared by the CSIRO Data61 data innovation group. As scholars working at the intersection of law and technology, we are delighted to participate in consultation process led by the Department of Industry, Innovation and Science, because we believe that policy in this area is important to shape Australia's future. In this submission, we draw upon some of the research conducted by the Allens Hub researchers to suggest how the current Ethical Framework, proposed in the discussion paper, could be improved. We note that this research does not relate to all the questions raised in the terms of reference, and so we only set out answers in relation to those matters where our research may be relevant, namely, Questions 1, 2 and 7. While this submission considers the potential impacts of AI generally, it will take as its primary consideration the legal and judicial context. In particular, we examine the guidance that rule of law values offer for the future development and use of AI. We aim to provide a meaningful contribution to the Australian Government's objective of developing an ethical framework for the design and application of AI with the goal of promoting trust in AI amongst Australian citizens. In making this submission we hope to raise some additional legal issues related to AI. We are grateful for the opportunity to present our views and hope this submission will assist in informing the Government's approach to developing policy in relation to AI ethics in Australia. The opinions expressed in this submission are the views of the authors, and not reflect or present the views or positions of the Allens Hub or UNSW Law.

Answer to Questions 1,2 and 7

Overview of the principles put forward in the discussion paper

The 'Artificial Intelligence: Australia's Ethics Framework' discussion paper identifies eight 'core principles' which are to be used as an ethical framework to guide organisations in the development and use of AI systems. [1] Some of these proposed principles have a degree of overlap with the key rule of law values identified by the Allens Hub researchers Monika Zalnieriute, Lyria Bennett Moses and George Williams in their recent work on the rule of law and automation of government decision-making.[2] These values are transparency, accountability, predictability and consistency, and equality before the law. Together they represent, what we call, the 'minimalist' version of the rule of law. [3] It overlaps with five of the eight principles identified in the 'Artificial Intelligence: Australia's Ethics Framework' discussion paper:

> **Principle 3: Regulatory and legal compliance.** The AI system must comply with all relevant international, Australian Local, State/Territory and Federal government obligations, regulations and laws.
>
> **Principle 5: Fairness.** The development or use of the AI system must not result in unfair discrimination against individuals, communities or groups. This requires particular attention to ensure the "training data" is free from bias or characteristics which may cause the algorithm to behave unfairly.
>
> **Principle 6: Transparency and explainability.** People must be informed when an algorithm is being used that impacts them and they should be provided with information about what information the algorithm uses to make decisions.
>
> **Principle 7: Contestability.** When an algorithm significantly impacts a person there must be an efficient process to allow that person to challenge the use or output of the algorithm.

---

1 D Dawson et al, 'Artificial intelligence: Australia's Ethics Framework' Data61 CSIRO, Australia (2019) 57 [7].

2 Monika Zalnieriute, Lyria Bennett Moses and George Williams, 'Rule of Law and Automation in Government Decision-Making' *Modern Law Review,* Vol 82(3).

3 Ibid.

> **Principle 8: Accountability.** People and organisations responsible for the creation and implementation of AI algorithms should be identifiable and accountable for the impacts of that algorithm.[4]

We respectfully submit that while there is overlap between the key principles identified by Zalnieriute, Bennett Moses and Williams and Data 61 discussion papers proposed ethical framework, further elaboration of bearing of rule of law on AI would be valuable. Our analysis focuses on highlighting challenges that the use and development of AI systems pose to the fundamental public law values and principles, and in particular the concept of the rule of law.[5] Even though the rule of law framework is generally discussed in the context of governments, its core principles or values are also often followed by private actors – either voluntarily or because they are embedded into legislative frameworks governing their activities.

Rule of Law Values Relevant for AI

We argue that the use of rule of law as a guiding framework for the development of Australian AI policy is valuable as it is both a widely accepted standard for gauging governmental behaviour worldwide,[6] and reflects the fundamental values of Australian society. Furthermore, as we suggest in some of our work, it is widely understood to be flexible and responsive to social and legal developments.[7] As academics and jurists have long argued, the rule of law can be understood as a goal or an ideal, a state in which the legal system is free from certain dangers or pathologies.[8] The use of this long established and widely endorsed framework is in line with the goals of the discussion paper, namely, that the aim of developing an ethics framework for AI is not about rewriting existing laws or standards. Instead it is about 'updating them to ensure that existing laws and ethical principles can be applied in the context of new AI technologies.'[9] Drawing primarily on the work of Hub researchers Zalnieriute, Bennett Moses and Williams in this area, we wish to highlight five key principles which have a significant bearing on the ethics of AI: transparency, accountability, equality before the law,

---

4 Dawson, above n 1, 57 [7].

5 Zalnieriute, Bennett Moses and Williams, above n 2.

6 See International Congress of Jurists, 'The Rule of Law in a Free Society' (Report of the International Commission of Jurists, New Delhi, 1959), [1].

7 Monika Zalnieriute, Lisa Burton Crawford, Janina Boughey, Lyria Bennett Moses, Sarah Logan, 'From the Rule of Law to Statue Drafting: Legal Issues for Algorithms in Government Decision-Making,' in Woodrow Barfield (ed.) *Cambridge Handbook on Law and Algorithms,* Cambridge University Press, UK, 2019.

8 Martin Krygier, 'The Rule of Law: Legality, Teleology, Sociology' in Gianlugi Palomblla and Neil Walker (ed), Relocating the Rule of Law (Hart Publishing, 2009).

9 Dawson, above n 1, 5-6.

predictability and consistency. [10] In the process of examining the importance of these principles we hope to raise some additional ethical issues related to AI.

Transparency and accountability

AI can pose serious challenges to transparency and accountability. Academic Jenna Burrell has convincingly categorised these challenges as falling within three 'forms of opacity' in relation to machine learning. [11] First, within this frame of reference, intentional secrecy due to algorithms being treated as a commercial or state secret may inhibit transparency. [12] The discussion paper cited the cases study of the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) system currently in use in many United States courts to advise judges on sentencing and probation decisions. [13] This case study is used as illustrative of issues of potential bias, particularly racial bias, in such systems. To elaborate on this point, Northpointe Inc (now 'Equivant'), [14] which created and owns COMPAS, has not publicly disclosed its methods in developing this tool, as it claims that it considers its algorithms to be commercial secrets. [15] This is particularly troubling as the COMPAS tool plays a significant role in judicial sentencing, an activity that has a direct impact on individual liberty. Along with other scholars, in our work we have strongly advocated for open source software where the decision-making can have such significant impact, such as on an individual's liberty. [16]

Technical illiteracy amongst users can also pose a challenge to transparency and AI. As noted by Burell, even if operational information is disclosed, extracting useful knowledge from that information may be inaccessible to much of the population, creating a second level of opacity.[17] Furthur, Burell identifies a third level of opacity brought about by a fundamental difference in the reasoning process between humans and machines. Even with suitable training, humans aren't always able to interpret the interactions among algorithms and data. [18] This, we suggest, poses a serious challenge to transparency,

---

10 Zalnieriute, Bennett Moses and Williams, above n 2. We also discuss them in Zalnieriute, Burton Crawford, Boughey, Bennett Moses, and Logan, n 7.

11 Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3(1) *Big Data & Society*, 1.

12 Ibid.

13 Dawson, above n 1, 40, [5.2.1].

14 Equivant' at http://www.equivant.com/ (last accessed 15 April 2019).

15 Noted in the United States case *State v. Loomis* 881 N.W.2d 749 (Wis. 2016), at [144]. See generally Pasquale, Frank, *The black box society* (Harvard University Press, 2015).

16 See, e.g., Monika Zalnieriute and Felicity Bell, 'Technology and Judicial Role' forthcoming in Gabrielle Appleby and Andrew Lynch (eds.), *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia*, Cambridge University Press, UK, 2020. Danielle Keats Citron, 'Technological Due Process' (2008) 85 *Washington University Law Review* 1249; D. K. Citron and F. Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89 *Washington Law Review* 1.

17 Burrell, n 11, p. 4.

18 Ibid, 4-5.

a crucial value for preserving the rule of law, as machine learning systems inevitably become more complex over time.

Equality before the law

Equality before the law is a central tenet of the rule of law. The adoption and development of AI by governments and the judiciary may further challenge the notion that, irrespective of their status, all individuals must be treated equally before the law.[19] This concern has been explored in the work of scholars from various disciplines. [20] The COMPAS case study is illustrative of this point as a notable lack of transparency has led to a compromise in due process rights in the United States. Further, as noted in the discussion paper, inherent biases in the data used when creating algorithms can fundamentally compromise an individual's ability to seek equal treatment before the law. [21] Equally troubling is AI's potential, through the use of particular data sets, to further intrench existing biases and inequalities

Predictability and consistency

It has been argued that AI has the potential to enhance consistency between government action and the law. Given the complexity of modern legislative frameworks and the frequency with which changes are made it can be difficult for the most conscientious government actors or agent to always act consistently with the law. Scholars, business and policymakers have all noted that the use of algorithms in government decision-making could improve consistency with the law. [22] This argument is coherent with the idea that pre-programmed systems, unlike humans, are unable to disregard pre-determined rules. Therefore, as noted by Zalnieriute, Bennett Moses and Williams, in some circumstances AI offers the potential to enhance consistency in decision-making. [23]

---

19 Zalnieriute and Bell, n 16; Zalnieriute, Bennett Moses and Williams, n 2.

20 See Barocas, Solon, and Andrew D. Selbst. "Big data's disparate impact." *Calif. L. Rev.* 104 (2016): 671. People have particularly strongly objected to courts systematically imposing more severe sentences on defendants who are poor or uneducated or from a certain demographic group: see G. Kleck, 'Racial Discrimination in Criminal Sentencing: A Critical Evaluation of the Evidence with Additional Evidence on the Death Penalty' (1981) 46 *American Sociological Review* 783; L. Wacquant, 'The Penalisation of Poverty and the Rise of Neo-Liberalism' (2001) 9 *European Journal on Criminal Policy and Research* 401; C. Hsieh and M.D. Pugh, 'Poverty, Income Inequality, and Violent Crime: A Meta-Analysis of Recent Aggregate Data Studies' (1993) 18 *Criminal Justice Review* 182.

21 Dawson, above n 1, 40-41 [5.2.1].

22 See for example C Coglianese and D Lehr, 'Regulating by Robot: Administrative Decision Making in the Machine-Learning Era' (2017) 105 *Georgetown Law Journal* 1147.

23 Zalnieriute, Bennett Moses and Williams, above n 2.

However, challenges to consistency and coherence may arise when the rules put in place to govern AI, such as in a pre-programmed system, are inconsistent with legal requirements. A case that is illustrative of this point, and raises important ethical issues, is the Australian Robo-debt programme used by the Department of Human Services, beginning in 2016. [24] While it is yet to be determined whether the government's actions were legal, it has now become clear that many people were incorrectly advised that there was a discrepancy between their reported income and their legal entitlements. [25] The issue in this case was not that errors were made, human decision making often produces errors, it was that errors were made on a scale far greater than if the system were human driven. The program had a high error rate due to a fundamental error in the assumption made in its calculations – that fortnightly incomes could be calculated by averaging from annual incomes. However, this assumption did not apply to many welfare recipients, namely those with fortnightly incomes that varied. Fundamental errors in designing this system, attributable to human error, ultimately lead to decision-making that may be proven to be inconsistent with the law.

Values of the Australian public

The discussion paper appears to broadly characterise the values of the Australian public as encapsulated by the colloquial motto 'everyone deserves a fair go.'[26] In the paper statements concerning the intent of developing an ethical framework for AI are framed in terms of 'a fair go', for example, as argued by the paper, '[e]nsuring that AI systems are operating in a *fair* and balanced ways across the diverse Australian population is a cornerstone of ethical AI' or industry standard should be developed 'to support the use of *fair* algorithms'.[27] This umbrella concept of 'fairness' can be extrapolated to encompass many of the rule of law principles discussed above. We submit that the fundamental public law principles and values, particularly the values of the rule of law, are the cornerstone of the Australian legal system and reflect many of the core values of the Australian public.

---

[24] Luke Henriques-Gomes, 'Centrelink cancels 40,000 robodebts, new figures reveal' *The Guardian* (online) 6 February 2019, cited 30 May 2019 < https://www.theguardian.com/australia-news/2019/feb/06/robodebt-faces-landmark-legal-challenge-over-crude-income-calculations>.

[25] Senate Community Affairs References Committee, Parliament of Australia, *Design, Scope, Cost-Benefit Analysis, Contracts Awarded and Implementation Associated with the Better Management of the Social Welfare System Initiative* (2017) at [2.88].

[26] Dawson, above n 1, 38 [5].

[27] Ibid.