THE UNIVERSITY OF SYDNEY

Artificial Intelligence: Australia's Ethics Framework

Submission to the Department of Industry, Innovation and Science

Dr Carolyn McKay

University of Sydney Law School

31 May 2019

Artificial Intelligence: Australia's Ethics Framework

Dr Carolyn McKay, Lecturer in Law, The University of Sydney Law School

Thank you for the opportunity to comment briefly on the Artificial Intelligence (AI) Discussion Paper.[1] The focus of this submission is on the impact of AI on the criminal justice system and specifically the use of AI risk assessment instruments in various criminal justice decision points[2] that affect a person's liberty and/or legal status. Only Question 1 has been addressed in this submission.

At the outset of this submission, it is important to understand that there are various statistical, data-driven predictive tools increasingly used in criminal procedure risk assessments that are not strictly AI, rather they are 'actuarial' or 'algorithmic' instruments. However, similar ethical concerns arise in these instances requiring inclusion and consideration in the proposed ethical framework.

1. Are the principles put forward in the discussion paper the right ones? Is anything missing?

The Principle of Legitimacy

The ethics of using AI in criminal procedure is significant given the high stakes decisions that can be made in that context, for instance, potential criminalization, punishment and loss of liberty.[3] A variety of data-driven actuarial, algorithmic or AI tools may be applied in risk assessments undertaken for the purposes of bail, sentencing and parole hearings as well as in determining whether an offender's detention should be continued or their release be supervised.[4]

The core principles for AI identified in the Discussion Paper that are particularly pertinent to criminal procedure are Fairness, Contestability, Transparency & Explainability and Accountability. It is submitted that flowing from the application of these principles is the missing principle of Legitimacy: the inherent integrity and authority of a democratic criminal justice system that is fundamental for citizens' voluntary compliance and trust.[5] The legitimacy of the law is premised upon a 'moral authority, which in turn depends on law's ability to justify its requirements'[6], closely connected with the Rule of Law that, amongst other things, underpins the

---

[1] Dawson, D. and Schleiger, E., Horton, J., McLaughlin, J., Robinson, C., Quezada, G., Scowcroft, J., and Hajkowicz, S. (2019). *Artificial Intelligence: Australia's Ethics Framework*. Data61 CSIRO, Australia, hereinafter referred to as the Discussion Paper.

[2] Barabas, C., Dinakar, K., Ito, J., Virza, M., & Zittrain, J. (2017). Interventions over predictions: Reframing the ethical debate for actuarial risk assessment. *arXiv preprint arXiv:1712.08238*.

[3] Eckhouse, L., Lum, K., Conti-Cook, C. and Ciccolini, J. (2019) Layers of Bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior, vol. 46, No. 2, 185-209*.

[4] For example, in New South Wales, risk assessments are framed by legislation including: *Bail Act 2013* (NSW), *Crimes (High Risk Offenders) Act 2006* (NSW), *Terrorism (High Risk Offenders) Act 2017* (NSW), *Crimes (Sentencing Procedure) Act 1999* (NSW).

[5] Tyler, T.R. (1994). Governing amid diversity: The effect of fair decision-making procedures on the legitimacy of government. *Law and Society Review*, 28, 809-831; Centre for Justice Innovation (2018). *Just technology: Emergent technologies and the justice system…And what the public thinks about it.* https://justiceinnovation.org/publications/just-technology-emergent-technologies-and-justice-system-and-what-public-thinks-about; Stobbs, N., Hunter, D. and Bagaric, M. (2017). 'Can Sentencing Be Enhanced by the Use of Artificial Intelligence?' 41 *Criminal Law Journal* 261; Goel, S., Perelman, M., Shroff, R., & Sklansky, D. A. (2017). Combatting police discrimination in the age of big data. *New Criminal Law Review: In International and Interdisciplinary Journal*, 20(2), 181-232; Rowden, E., 2018. Distributed Courts and Legitimacy: What do we Lose When we Lose the Courthouse? *Law, Culture and the Humanities*, 14(2), 263–281.

[6] Stern, S. (2018). Introduction: Artificial intelligence, technology, and the law. *University of Toronto Law Journal*, 68(supplement 1), 1-11, 4, citing Sheppard, B. 'Warming up to Inscrutability: How Technology

need for legal decision-makers to have authority to determine outcomes and for the decision-making process to be examinable and contestable.[7] It centres on the foundational checks and balances that provide standing to citizens, enabling them to challenge violations of fundamental rights.[8] To be legitimate, the law and its application should be 'accessible and so far as possible intelligible, clear and predictable'.[9] The principle of legitimacy is a significant issue in situations when state power is being used to criminalise or punish. How, then, is the legitimacy of criminal procedure challenged when decision-making authority is 'ceded to the algorithm'?[10]

The principles of Fairness, Contestability, Transparency & Explainability, Accountability and, I submit, Legitimacy are brought to the fore in one of the Discussion Paper case studies.[11] The United States case of *State of Wisconsin v Loomis* illustrates the challenges in utilising algorithmic risk assessment tools in sentencing procedure, specifically the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) designed by Northpointe (now Equivant), a private, for-profit company.[12] There is considerable literature critiquing *Loomis* in the context of in-built bias, that is, the profiling of an individual accused or offender against a predictive algorithm and group data that may be premised on embedded bias.[13]

In particular, the case demonstrates the impenetrability and opacity of algorithms that are subject to proprietary interests. While Mr Loomis could verify certain inputs to the COMPAS risk assessment tool, the concluding score, that recommended a significant prison sentence, could not be contested as the internal structure and formulae of the algorithm was based on proprietary information. The Court accepted that Northpointe treated COMPAS as a trade secret and therefore it did not need to disclose how risk scores or factors were determined and weighed.[14] Mr Loomis was not entitled to access the formulae or factors that were determinative of his high risk score and his punishment.[15] The inscrutability of COMPAS has been analysed as being morally corrupt.[16] As the Honourable Justice Martin argues regarding

---

Could Challenge Our Concept of Law' (2018) 68: *Suppl UTLJ* 36, citing Raz, J., *The Authority of Law: Essays on Law and Morality* (Oxford: Clarendon Press, 1979).

[7] Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128), 20170359.

[8] Hildebrandt, M. (2018). Law as computation in the era of artificial legal intelligence: Speaking law to the power of statistics. *University of Toronto Law Journal*, 68(supplement 1), 12-35.

[9] The Hon. M. Gordon, 2017. Courts and the Future of the Rule of Law. **http://www.hcourt.gov.au/assets/publications/speeches/current-justices/gordonj/gordonj21Jul2017.pdf** , 2, citing Lord Bingham, "The Rule of Law", speech delivered as the Centre for Public Law's Sixth Sir David Williams Lecture, 16 November 2006 at 6.

[10] Stern, S. (2018). Introduction: Artificial intelligence, technology, and the law. *University of Toronto Law Journal*, 68(supplement 1), 1-11, 3.

[11] Discussion Paper, 40-41.

[12] *State of Wisconsin v Loomis* 881 N.W.2d 749 (Wis. 2016); The Hon. M. Gordon, 2017. Courts and the Future of the Rule of Law. http://www.hcourt.gov.au/assets/publications/speeches/current-justices/gordonj/gordonj21Jul2017.pdf; *State v. Loomis:* Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing. Harvard Law Review. https://harvardlawreview.org/2017/03/state-v-loomis/

[13] Eckhouse, L., Lum, K., Conti-Cook, C. and Ciccolini, J. (2019) Layers of Bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior,* vol. 46, No. 2, 185-209, 190 citing analysis by ProPublica: Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016) Machine bias. *Propublica.* **https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing** ; Carlson, A. (2017) 'The Need for Transparency in the Age of Predictive Sentencing Algorithms', *Iowa Law Review*, Vol. 103: 303-329; Case Note, (2017) State v Loomis: Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessment in Sentencing https://harvardlawreview.org/2017/03/state-v-loomis/

[14] Carlson, A. (2017) 'The Need for Transparency in the Age of Predictive Sentencing Algorithms', *Iowa Law Review*, Vol. 103: 303-329.

[15] Carlson, A. (2017) 'The Need for Transparency in the Age of Predictive Sentencing Algorithms', *Iowa Law Review*, Vol. 103: 303-329.

[16] Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A:*

inscrutability, 'this is entirely inconsistent with the common law requirement that a decision maker must expose his or her reasoning':[17] the legitimacy of the decision-making process is thereby impugned. Moreover, there is evidence that the mere existence of a predictive risk score can overly influence a judge's decision by focusing attention on the risk of recidivism over and above other relevant factors.[18] Studies suggest that given the proprietary nature of algorithmic tools, it would be difficult or impossible for a judge to meaningfully critique or challenge the risk assessment.[19]

Common Law Principles

The application of AI tools in criminal procedure can be analysed through the common law principle of procedural justice. This term, encapsulating fairness and due process, is derived from natural justice and its elements include open justice, equality before the law, the presumption of innocence and the right to hear and answer a case brought by the state.[20] Regarding open justice, criminal proceedings should be subject to public oversight as a means to counteract abuses of power and to promote transparency, accountability and ultimately, the Rule of Law.[21] But open justice may be undermined when defendants, courts and society are denied the oversight of AI tools that are used in determining a person's legal status and liberty. AI needs to be 'testable and contestable.'[22]

Equality of arms is a key principle in procedural justice meaning that the defendant should not be at a disadvantage compared with the prosecuting state.[23] However, there is no level playing field if the prosecution uses opaque risk assessment instruments and undisclosed input data against a defendant.

The presumption of innocence is the 'golden thread' that runs through the criminal justice system to ensure that the legal onus of proof remains on the prosecution.[24] If there is no way of proving or disproving an algorithm's formulae or methodology, the prosecution's burden of proving a case against a defendant beyond reasonable doubt seems compromised. In addition, the presumption of innocence – as well as the concept of individualized justice[25] - are

---

Mathematical, Physical and Engineering Sciences, 376(2128), 20170359; Polson, N. and Scott, J. 2018 *AIQ: how artificial intelligence works and how we can harness its power for a better world*. London, UK: Bantam Press.

[17] Martin, G.C. (2017) How far has technology invaded the criminal justice system? **http://classic.austlii.edu.au/au/journals/QldJSchol/2018/10.pdf**, 22.

[18] Eckhouse, L., Lum, K., Conti-Cook, C. and Ciccolini, J. (2019) Layers of Bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior,* vol. 46, No. 2, 185-209; Harcourt, B. (2010) Risk as a Proxy for Race 2 (John M Olin & Econ Working Paper No 535 (2d Series) & Pub. Law & Legal Theory Working Paper No 323, 2010).

[19] Case Note, (2017) *State v Loomis*: Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessment in Sentencing **https://harvardlawreview.org/2017/03/state-v-loomis/**; Christin, A., Rosenblat, A. and Boyd, D. (2015) Courts and Predictive Algorithms *Data & Society* http://www.datacivilrights.org/pubs/2015-1027/Courts_and_Predictive_Algorithms.pdf

[20] McKay, C. (2018) *The Pixelated Prisoner: Prison video links, court 'appearance' and the justice matrix*. Routledge; Bronitt, S. and McSherry, B. (2017) *Principles of Criminal Law*. Lawbook Company; Mulcahy, L. (2013) Putting the defendant in their place: Why do we still use the dock in criminal proceedings? *British Journal of Criminology* 53(6), 1139-56; Hildebrandt, M. (2018). Law as computation in the era of artificial legal intelligence: Speaking law to the power of statistics. *University of Toronto Law Journal*, 68(supplement 1), 12-35.

[21] Resnik, J. (2015) The contingency of courts: changing the experiences and logics of the public's role in court-based ADR *Nevada Law Journal* 15, 951; McKay, C. (2018) *The Pixelated Prisoner: Prison video links, court 'appearance' and the justice matrix*. Routledge.

[22] Hildebrandt, M. (2018). Law as computation in the era of artificial legal intelligence: Speaking law to the power of statistics. *University of Toronto Law Journal*, 68(supplement 1), 12-35, 34

[23] McKay, C. (2018) *The Pixelated Prisoner: Prison video links, court 'appearance' and the justice matrix*. Routledge; Roberts, P and Zuckerman, A (2010) *Criminal Evidence*. Oxford University Press.

[24] *Woolmington v DPP* [1935] AC 462

[25] Martin, G.C. (2017) How far has technology invaded the criminal justice system? **http://classic.austlii.edu.au/au/journals/QldJSchol/2018/10.pdf**

challenged by assessing an individual defendant against aggregate group data of other 'like' offenders.

Finally, the hearing rule, *audi alteram partem*, requires that the defendant be enabled to hear and comprehend the case being brought against them.[26] The use of unintelligible, inscrutable, secret proprietary information against a citizen is at odds with the right to be enabled to understand and defend a case.

Human Rights Principles

AI tools used in determinative criminal procedures can also be analysed through a human rights framework.[27] Following from the above analysis of AI in criminal procedure and the impacts on common law principles, there is the potential for AI to violate human rights, specifically, equality before the courts and the right to a fair and public hearing heard by a competent, independent and impartial tribunal.[28] Other human rights measures that may be challenged by AI include the presumption of innocence, [29] general procedural fairness requirements including the presentation of an understandable case against the defendant,[30] and protection from discrimination.[31]

Conclusion

In conclusion, where private, for-profit organisations are involved in providing instruments for essential public functions such as the criminal justice system, their products should be subject to democratic disclosure and freedom of information requirements. [32] On this basis, various scholars support the idea of a regulatory body to oversee and audit algorithms and thereby ensure transparency, accountability and procedural justice.[33]

---

[26] Butt, P. and Hamer, D. (2011) *LexisNexis Concise Australian Legal Dictionary* LexisNexis Butterworths.
[27] Australian Human Rights Commission (2018) *Human Rights and Technology Issues Paper*, July 2018. https://www.humanrights.gov.au/sites/default/files/document/publication/AHRC-Human-Rights-Tech-IP.pdf; Pasquale, F. & Cashwell, G., (2018) 'Prediction, Persuasion, and the Jurisprudence of Behaviourism' 68:Suppl *UTLJ* 63 at 65; Stern, S. (2018). Introduction: Artificial intelligence, technology, and the law. *University of Toronto Law Journal*, 68(supplement 1), 1-11; Aletras, N. et al, 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective' (2006) 2 *PeerJ Computer Science* 92;
[28] *International Covenant on Civil and Political Rights* opened for signature 16 December 1966, 999 UNTS 171 (entered into force 23 March 1976) art 14(1).
[29] Ibid art 14(2).
[30] Ibid art 14(3)(a), art 14(3)-(7).
[31] Ibid art 26.
[32] Carlson, A. (2017) 'The Need for Transparency in the Age of Predictive Sentencing Algorithms', *Iowa Law Review*, Vol. 103: 303-329; Keats Citron, D. and Pasquale, F. (2014) The scored society: due process for automated predictions. *Washington Law Review* 89:1. https://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1318/89WLR0001.pdf
[33] Hogan-Doran, D. (2017) 'Computer says 'no': Automation, algorithms and artificial intelligence in Government decision-making', 13 *The Judicial Review*; Pasquale, F. (2017) Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society. *Ohio State Law Journal*, Vol 78, U of Maryland Legal Studies Research Paper No. 2017-21. Available at SSRN: https://ssrn.com/abstract=3002546; Balkin, J. (2017) The Three Laws of Robotics in the Age of Big Data. *Ohio State Law Journal* Vol 78: 5, 1217. https://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?referer=https://www.google.com.au/&httpsredir=1&article=6160&context=fss_papers