Discrimination by Machine: An Evaluation of Australia's Regulatory Approach to Machine Learning Discrimination

Research Thesis 76040, Autumn 2020

Word count: 10709

Submitted in fulfilment of the requirements for 76040 Research Thesis, Autumn Semester 2020, at the University of Technology Sydney

CONTENTS

ABSTRACT

The increasing use of Machine learning systems has given rise to the greater risk of Machine learning discrimination against the most vulnerable of society. Machine learning's reliance on data and its learning nature makes the permeation of human bias in its decisions almost inevitable. In response to this increased threat, this article outlines the dangers of Machine learning bias and assesses Machine learning induced barriers to the existing Australian anti-discrimination framework. It then compares the individual rights and accountability and oversight approaches to Machine learning regulations. It determines which approach is most appropriate for the Australian context by considering which approach would most effectively address the barriers while balancing the commercial interests of Australia's Machine learning industry.

I INTRODUCTION

The use of Machine learning ('ML') systems has grown exponentially and shows no sign of slowing down. The global ML market is expected to exceed $30 billion USD by 2024;[1] meaning its presence in, and effect on, everyday life will become more prevalent, posing a challenge for regulators to mitigate threats of discrimination. ML is a subset of Artificial Intelligence ('AI') and is distinct from other algorithm-based programs due to its learning capability which improves its ability to make predictions.[2] To do so, ML relies on historical data and data collected from interactions with other systems and humans.[3] However, in most cases, this data is riddled with human biases and the machine's output will likely reflect these biases.[4] This poses the inevitable threat of discrimination by machine where this bias is left unchecked and the output informs a human's decision.[5] Internationally, discrimination in ML-informed decisions has been recognised by regulators in the United States ('US') and the European Union ('EU') who have attempted to mitigate discriminatory impacts. Australia has a much less developed regulatory framework and has yet to address the implications of ML on decision-making processes. Any such regulations created to address ML discrimination must be appropriate for the Australian context, protect vulnerable individuals and future-proof Australia's ML industry. This article argues that discrimination caused by ML-informed decisions is inadequately addressed in Australia's existing anti-discrimination framework and provides a suggested approach for redress via ML regulations.

Previous studies of ML regulations conducted by researchers, academics and governments can be grouped into two categories. The first group champions an individual rights approach as a means of creating transparency.[6] The second group focuses on an accountability and oversight approach, which instead imposes restrictions and obligations on those using ML systems.[7] Recently, international scholars have considered how the use of ML interacts with existing legal frameworks. In doing so, they identify barriers to the application and procedural efficacy of laws which are created or exacerbated by ML.[8] While these studies evaluate existing regulations, few have discussed how gaps in existing anti-

[1] Louis Columbus, 'Roundup of Machine Learning Forecasts and Market Estimates: 2020', *Forbes* (online at 19 January 2020) <https://www.forbes.com/sites/louiscolumbus/2020/01/19/roundup-of-machine-learning-forecasts-and-market-estimates-2020/>.
[2] Brent Daniel Mittelstadt et al, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3(2) *Big Data & Society* 1, 3.
[3] Ignacio N Cofone, 'Algorithmic Discrimination Is an Information Problem' (2019) 70(6) *Hastings Law Journal* 1389, 1424.
[4] Ibid 1426.
[5] Jon Kleinberg et al, 'Discrimination in the Age of Algorithms' (2018) 10(1) *Journal of Legal Analysis* 1, 3.
[6] See, eg, Margot E Kaminski, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability' (2019) 92(6) *Southern California Law Review* 1529, 1553.
[7] See, eg, Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 *Duke Law & Technology Review* 18.
[8] See, eg, Shlomit Yanisky-Ravid and Sean K Hallisey, 'Equality and Privacy by Design: A New Model of Artificial Intelligence Data Transparency via Auditing, Certification, and Safe Harbour Regimes' (2019) 46(2) *Fordham Urban Law Journal* 428, 446.

discrimination frameworks should be approached.[9] In Australia, the Department of Industry, Science, Energy and Resources published an AI ethics framework in November 2019 containing eight principles which developers are encouraged to adhere to when creating and implementing AI.[10] The Australian Human Rights Commission ('AHRC') is currently conducting public consultation with stakeholders to formulate proposed ways of addressing select issues presented by AI,[11] including the discriminatory effects of ML systems.[12] While the recent Australian work acknowledges the dangers of discrimination, it fails to do two things. Firstly, it has not identified the barriers to the existing anti-discrimination framework in Australia. Secondly, it has failed to evaluate how different approaches may vary in effectiveness in Australia, particularly with respect to their success in mitigating discrimination.

This article contributes to the literature in three ways. First, it groups the existing literature into two categories. Although other classifications of the literature may exist, the distinction made in this article hinges on the enforceability of the anti-discrimination framework; thereby providing a necessary consideration for regulators moving forward. Second, it identifies the barriers inhibiting the functionality of the Australian anti-discrimination framework. Third, it assesses the effectiveness of the regulatory approaches in addressing these barriers. This article's unique classification of existing strategies into the accountability and oversight approach and individual rights approach is an essential addition to the Australian discourse. The classifications appropriately broaden the discussion beyond strictly implementing existing EU or US strategies. By understanding the purpose and outcomes of such regulatory strategies, Australia can map out the effectiveness of each approach. Given Australia's early stages of regulating ML, designing a purposeful regulatory regime is of critical importance. This article focuses on an Australian application of ML discrimination and algorithmic bias literature. However, the approach it takes to identify barriers to the existing legal framework and evaluate regulatory approaches can be applied similarly around the world or within other areas of Australian law, such as competition and consumer law.[13] While some of the discussion in this article applies to other AI systems and algorithmic decisions, this article focuses on ML given the significant reliance on data and substantial size of the ML industry, which currently leads all AI funding worldwide.[14]

In Part II, this article will demonstrate how ML bias occurs and the discrimination that ensues. Part III of this article analyses the barriers to justice created where ML participates in decision-making with reference to the existing Australian anti-discrimination framework. These gaps illustrate that applying

---

[9] Kleinberg et al (n 5).
[10] 'AI Ethics Principles', *Department of Industry, Science, Energy and Resources* (Web Page) <https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles>.
[11] 'Consultation', *Human Rights & Technology* (Web Page) <https://tech.humanrights.gov.au/consultation>.
[12] See Sophie Farthing et al, 'Human Rights and Technology Discussion Paper' (Paper, Australian Human Rights Commission, December 2019) 78.
[13] Celine Castets-Renard, 'Accountability of Algorithms in the GDPR and Beyond: A European Legal Framework on Automated Decision-Making' (2019) 30(1) *Fordham Intellectual Property, Media & Entertainment Law Journal* 91, 125.
[14] Columbus (n 1).

the existing anti-discrimination framework to ML reduces its usefulness and is contrary to the objectives of the framework. Part IV evaluates the effectiveness of an individual rights approach compared to an accountability and oversight approach in regulating ML, specifically in light of addressing the ML-induced barriers while promoting the innovation and growth of Australia's ML industry and capabilities.

II THE DANGERS OF MACHINE LEARNING BIAS AND DISCRIMINATION

ML is a subset of AI and is a 'set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty'.[15] There are three critical aspects of ML to understand in order to appreciate the unique dangers of ML discrimination. Firstly, ML systems are reliant on large amounts of historical data.[16] By learning from historical data used to train the system, ML will process inputted data in accordance with rules set by a processing algorithm and provide its prediction (the output).[17] Kleinberg et al provide the example of an ML system used for recruitment.[18] The system predicts applicants' likely success in the position being recruited by analysing patterns in the performance of existing employees.[19] The machine's processing algorithm would identify attributes that it considers determinative of an individuals' success and then process each applicant's CV against this criteria.[20]

Secondly, ML systems have a learning algorithm which enables them to alter the processing algorithm to reflect their learnings from historical data or human feedback.[21] As a machine processes data, the learning algorithm alters the variables, or weight attached to certain variables, to improve predictions.[22] It is said that this allows ML to increase the accuracy of predictions without necessarily requiring the oversight, recoding or input of humans.[23] In some types of ML, such as reinforcement learning, a feedback loop created by human intervention is what guides the system towards greater accuracy.[24] The feedback loop can be intentionally created by quality checking and providing feedback

[15] Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (Massachusetts Institute of Technology Press, 2012) 1.
[16] Ethem Alpaydin, *Introduction to Machine Learning* (Massachusetts Institute of Technology Press, 3rd ed, 2014) 3; Yanisky-Ravid and Hallisey (n 8) 442.
[17] Alpaydin (n 16) 2–3.
[18] Kleinberg et al (n 5) 42.
[19] Ibid.
[20] Ibid.
[21] Kleinberg et al (n 5) 20; Yanisky-Ravid and Hallisey (n 8) 450.
[22] Emre Bayamlioglu, 'Contesting Automated Decisions' (2018) 4(4) *European Data Protection Law Review* 433, 439.
[23] David Lehr and Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn about Machine Learning' (2017) 51 *University of California, Davis Law Review* 653, 684–85.
[24] Rashida Richardson, Jason Schultz and Kate Crawford, 'Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice' (2019) 192 *New York University Law Review* 192, 218.

to the system if it is making incorrect predictions.[25] On the other hand, a feedback loop may be unintentionally created if the system is linked to, or can access data from, other systems, social media or the internet.[26] In the case of the recruitment system example, the feedback loop could be created by human interviewers or recruiters noting the outcomes of interviews conducted by humans. The machine will learn from this data by assessing the attributes of the successful candidate and adjusting its processing algorithm to ensure that future predictions reflect any patterns it identifies and thereby considers indicators of success.[27]

Finally, there are three types of ML systems, which all use methods that process and make predictions using the historical data and learning algorithm in different ways. In supervised learning models, systems are provided with labelled datasets which act as training examples, which over time allow the system to identify patterns between the labelled data and any processed data to provide better predictions that align with the training examples.[28] Unsupervised learning models provide historical data to the systems and allow the algorithms to identify the patterns independently.[29] The final type is reinforcement learning which can be used in conjunction with supervised or unsupervised models. It focuses on a feedback loop by way of 'occasional reward or punishment signals' which guides the learning of the system.[30] The different types have multiple algorithms which can be used within them, thereby resulting in varying complexities of ML systems.

A *Machine Learning Bias and Discrimination*

ML systems are often viewed as having greater objectivity and reliability than human decision-makers.[31] However, they rely on data that is highly likely to contain human bias which ultimately taints ML's predictions.[32] ML discrimination occurs when a person relies on the predictions of the bias-riddled ML system to make a discriminatory decision about an individual. Mittelstadt explains that biased decisions are inevitable in algorithmic decision-making, given that bias permeates systems because of 'pre-existing social values'.[33] These emerge from the institutions or cultures that algorithms develop within, as well as the 'technical constraints and emergent aspects of a context of use'.[34] Throughout the operation of an ML system, there are thereby three sources of bias; historical data used to train the

[25] Yanisky-Ravid and Hallisey (n 8) 442.
[26] Ibid 451.
[27] Ibid.
[28] Alpaydin (n 16) 11.
[29] Ibid.
[30] Murphy (n 15) 2.
[31] Mireille Hildebrandt, 'Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning' (2019) 20(1) *Theoretical Inquiries in Law* 83, 106.
[32] Richardson, Schultz and Crawford (n 24) 224.
[33] Mittelstadt et al (n 2) 7.
[34] Ibid.

system, the development process of the ML and the feedback loop. Historical data used to train systems will contain the biases of past decisions or actions of humans.[35] A recruitment system used by Amazon embodied this when its predictions favoured male applicants because of the historical data used.[36] Amazon exists in a historically male-dominated industry, and the system inadvertently learned to reflect the biases of past human decision-makers.[37] Secondly, an ML system may contain the biases of its engineers and software developers.[38] Systems will inadvertently learn to emulate the biases of the human taggers and engineers,[39] or may even be confined to operate in a way which inadvertently disadvantages marginalised groups (for example, if the engineers have not considered the impact of the system design on these groups). Finally, for reinforcement models, a positive feedback loop may continue to reinforce biases, thereby further skewing the data.[40] For example, in the recruitment ML system, the system could provide reasonable recommendations in the first round of interviews that specific candidates are suitable for a role. If management proceeds with interviews and ultimately hire a male for the role as opposed to a female (whether or not this decision was the result of their implicit bias), the system may learn that being a male is an indicator of success. Consequently, bias within ML systems can produce discriminatory outcomes for individuals where the ML system learns to make predictions in ways which breach anti-discrimination laws.

B *Australia's Need to Champion Non-Discrimination in Machine Learning Regulations*

Recognising the existence of bias in ML and its potential discriminatory outcomes is the first step in ensuring the protection of fundamental human rights while balancing the world's need for innovative solutions to existing and future problems. The use of ML raises several issues related to harm caused by breaches of privacy and the misuse and illegal use of information; issues which regulators often seek to bundle together with discrimination. However, this Part will demonstrate that extensive consideration of the discriminatory impacts of ML is of paramount importance when regulating ML. By understanding the risks of failing to address ML discrimination, regulators and legal professionals can better appreciate how and why they must champion non-discrimination when drafting general ML regulations, and further consider where discrimination mitigation fits within that framework.

The legal principles surrounding society's distaste for discrimination stems from the international law principle that 'all are entitled to equal protection against any discrimination'.[41] It is recognised that

[35] Cofone (n 3) 1426.
[36] Ibid 1397–1398.
[37] Ibid.
[38] Mittelstadt et al (n 2) 7.
[39] Ibid.
[40] Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104(3) *California Law Review* 671, 135.
[41] *Universal Declaration of Human Rights*, GA Res 217A (III), UN GAOR, UN Doc A/810 (10 December 1948) art 7.

failing to prohibit discrimination reinforces 'long-entrenched inequality' of the most vulnerable of society.[42] The disadvantages that result from discrimination range from financial adversities to educational and social segregation.[43] ML systems present many opportunities for improvements to human life, including the opportunity to mitigate discriminatory decisions and address the impacts of discrimination.[44]

Nonetheless, the need for action against discrimination grows daily with the increasing use of ML across all aspects of life. ML systems are currently used to improve and, in some cases, replace tasks previously performed by humans.[45] Recently AI systems, including ML, have been used for legal research,[46] in the development of self-driving cars,[47] for credit scoring,[48] for military operations[49] and for predictive policing.[50] Two pertinent examples that demonstrate the dangers of ML discrimination are the use of ML during COVID-19 and in human resources. Firstly, several ML-driven solutions have been developed in the past few months in response to problems arising from COVID-19. Since January 2020, the world has seen ML assist in the search for a COVID-19 vaccine,[51] use security camera footage to assess whether social distancing measures are being adhered to in workplaces[52] and predict the impact of COVID-19 on smaller cities using publicly available health data.[53] Despite examples of ML assisting during the pandemic, the use of ML systems is certainly not without danger. Reliance on data from more than 6 million reported cases of COVID-19[54] could result in the infiltration of biased or incorrect

[42] Matthew Adam Bruckner, 'The Promise and Perils of Algorithmic Lenders' Use of Big Data' (2018) 93(1) *Chicago-Kent Law Review* 3, 4.
[43] Castets-Renard (n 13) 99.
[44] Kleinberg et al (n 5) 3.
[45] Jim Shook, Robyn Smith and Alex Antonio, 'Transparency and Fairness in Machine Learning Applications' (2018) 4(5) *Texas A&M Journal of Property Law* 443, 458.
[46] Agnieszka McPeak, 'Disruptive Technology and the Ethical Lawyer' (2019) 50(3) *University of Toledo Law Review* 457, 462; 'Clio and ROSS Intelligence Join Forces to Redefine Legal Research', *ROSS Intelligence* (Blog Post, 21 October 2019) <https://blog.rossintelligence.com/post/clio-and-ross-intelligence-join-forces-to-redefine-legal-research>.
[47] Jack Stilgoe, 'Machine Learning, Social Learning and the Governance of Self-Driving Cars' (2018) 48(1) *Social Studies of Science* 25, 25.
[48] Mikella Hurley and Julius Adebayo, 'Credit Scoring in the Era of Big Data' (2016) 18 *Yale Journal of Law and Technology* 148, 151.
[49] Tejaswi Singh and Amit Gulhane, '8 Key Military Applications for Artificial Intelligence in 2018', *Market Research Blog* (Blog Post, 3 October 2018) <https://blog.marketresearch.com/8-key-military-applications-for-artificial-intelligence-in-2018>.
[50] Richardson, Schultz and Crawford (n 24) 196.
[51] Swami Sivasubramanian, 'How AI and Machine Learning Are Helping to Fight COVID-19', *World Economic Forum* (Web Page, 28 May 2020) <https://www.weforum.org/agenda/2020/05/how-ai-and-machine-learning-are-helping-to-fight-covid-19/>.
[52] Karen Hao, 'Machine Learning Could Check If You're Social Distancing Properly at Work', *MIT Technology Review* (Blog Post, 17 April 2020) <https://www.technologyreview.com/2020/04/17/1000092/ai-machine-learning-watches-social-distancing-at-work/>.
[53] Mary L Martialay, 'Machine Learning Models Predict COVID-19 Impact in Smaller Cities', *Rensselaer* (Blog Post, 17 April 2020) <https://news.rpi.edu/content/2020/04/17/machine-learning-models-predict-covid-19-impact-smaller-cities?utm_source=miragenews&utm_medium=miragenews&utm_campaign=news>.
[54] Martin Farrer, 'Global Report: Coronavirus Cases Pass 6 Million as Donald Trump Postpones G7', *The Guardian* (online at 31 May 2020) <http://www.theguardian.com/world/2020/may/31/global-report-coronavirus-cases-pass-6-million-as-donald-trump-postpones-g7>.

data. Failing to appropriately account for the specific context of each reported case (for example, cultural factors contributing to higher infection rates) could also lead to the misuse of this data.

Further, relying on outcomes produced by the system without human oversight could cost lives, especially when using the ML system to answer ethical conundrums. For example, as Italy's death toll increased, doctors queried whether hospitals would need to refuse care to older patients to manage Italy's limited medical resources.[55] Deciding to refuse care on the basis of age could be discriminatory in any event. However, an ML system could be used to inform this difficult decision by predicting the likelihood of a patient's recovery, thereby seemingly impartially addressing the issue. While this may assist in maximising available resources, the ML system could incorrectly identify other protected attributes, such as race, as a necessary contributor to recovery rates. Patients could therefore be refused treatment based on their race but under the guise of the maximisation of resources. On the other hand, if these types of ML systems perform accurately, they could save countless lives.

Additionally, the use of ML in human resources has recently been criticised and cautioned because of Amazon's ML recruiting tool.[56] The system viewed any CVs that included the word 'women', and that listed certain all-women colleges, unfavourably.[57] While Amazon has amended the system to attach neutrality to these terms and variables, given the often unpredictability of the learning path machines may take, this does not guarantee that similar discrimination will not occur in the future.[58] An added danger of discrimination in these circumstances is that candidates may lose autonomy over their expression and representation; needing to stiffly portray themselves in a way which meets the requirements that they believe systems will favour.[59]

It is evident why academics argue that the bias in ML systems can shift power to governments, authorities and companies using ML without providing justifications for decisions,[60] create an overall lack of accountability for any decision-maker relying on ML[61] and threaten the rule of law.[62] Some have also grappled with additional questions relating to the differing levels of autonomy and the relevant responsibility that should be apportioned to those using AI,[63] as well as the increasing use of algorithms

---

[55] Bevan Shields, 'Italian Doctors Propose Intensive Care Age Limit to Save Younger Patients', *The Sydney Morning Herald* (online at 12 March 2020) <https://www.smh.com.au/world/europe/italian-doctors-propose-intensive-care-age-limit-to-save-younger-patients-20200312-p5499t.html>.
[56] Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women', *Reuters* (online at 10 October 2018) <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.
[57] Ibid.
[58] Ibid.
[59] Mittelstadt et al (n 2) 9.
[60] Ari Ezra Waldman, 'Power, Process, and Automated Decision-Making' (2019) 88(2) *Fordham Law Review* 613, 616.
[61] Castets-Renard (n 13) 105.
[62] Emily Berman, 'A Government of Laws and Not of Machines' (2018) 98(5) *Boston University Law Review* 1277, 1283.
[63] Shook, Smith and Antonio (n 45) 459–462; Iria Giuffrida, 'Liability for AI Decision-Making: Some Legal and Ethical Considerations' (2019) 88(2) *Fordham Law Review* 439; Michael Callier and Harly Callier, 'Blame It on the Machine: A Socio-Legal Analysis of Liability in an AI World' (2018) 14(1) *Washington Journal of Law, Technology & Arts* 49.

by lawyers and judges.[64] Evidently, the legal issues relating to ML discrimination are complex and widespread. Australia needs to regulate before the impacts of unchecked ML discrimination become so detrimental that trust in the technology is lost. When considering potential regulations, Australia must ensure that mitigation of discrimination and accessibility to justice under the current framework is, at a minimum, maintained.

III MACHINE LEARNING INDUCED BARRIERS TO THE AUSTRALIAN ANTI-DISCRIMINATION FRAMEWORK

Australia does not currently have a clear approach to regulating discrimination caused by ML bias. Given that ML systems are not legal entities in Australia, this article considers situations where a human uses ML, and where the human's decision is subject to the existing anti-discrimination framework. It is noted that when applying the existing framework to ML, many additional questions arise relating to the identification of liability under the framework. While further research is required to assess how to identify liability, this Part considers whether discrimination is detectable when using ML and how the relevant evidence is accessed and interpreted by individuals or courts to prove discrimination.

A *The Australian Anti-Discrimination Framework*

The Australian anti-discrimination framework consists of statutes at the Commonwealth, State and territory levels, which prohibit discrimination against individuals based on protected attributes.[65] These statutes are mostly consistent with respect to direct discrimination and the underlying purpose of each act.[66] This framework exists to ensure that discrimination does not go unchecked in areas of public life.[67] In order to understand how the regulation of ML interacts with the Australian anti-discrimination framework, this article focuses on direct discrimination as an example. It first explains the elements of direct discrimination, how they might apply to ML and the process of lodging discrimination claims, before considering the impact of ML on the operation of the anti-discrimination framework in section B.

[64] Frank Fagan and Saul Levmore, 'The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion' (2019) 93(1) *Southern California Law Review* 1; Chris Chambers Goodman, 'AI/Esq: Impacts of Artificial Intelligence in Lawyer-Client Relationships' (2019) 72(1) *Oklahoma Law Review* 149.
[65] Including, but not limited to, race, sex, age, disability, colour, ethnic origin and sexuality.
[66] Neil Rees, Dominique Allen and Simon Rice, *Australian Anti-Discrimination Law* (Federation Press, 2nd ed, 2014) 3.
[67] Ibid.

1 *Elements of Direct Discrimination*

The anti-discrimination framework prohibits direct discrimination,[68] which is an act involving a distinction, exclusion, restriction or preference based on a protected attribute for the purpose of treating an individual differently to others.[69] There are thereby two elements of direct discrimination; differential treatment and the causation test. Establishing differential treatment requires 'that there be two situations or sets of circumstances, the actual and the hypothesized, so that it can be determined by a comparison whether treatment in the former is "less favourable" than in the latter'.[70] A complainant is thereby required to identify real or hypothetical comparators to demonstrate that in comparable circumstances, an individual without their protected attribute would have been treated more favourably.[71] In most discrimination cases, this comparator is notional.[72] When using ML, the task of identifying real comparators would not significantly vary, but identifying notional comparators would likely require counterfactuals. Counterfactuals operate by altering variables in the ML system to prove that, had an individual in the same circumstances without the protected attribute been processed by the system, the outcome would have varied.[73]

The question of causation is not a 'but for' test, but instead a question of fact as to what the real reason for differential treatment was.[74] The High Court's most recent interpretation of the test indicates that the motive, purpose and effect of a decision bear on the question of 'why was the aggrieved person treated as he or she was?', but they are not substitutes for the statutory expressions of 'because of' or 'reasons for'.[75] Multiple causes of differential treatment do not preclude the unlawfulness of the conduct.[76] In all jurisdictions other than Victoria, Queensland and South Australia, the discriminatory reason does not need to be dominant or significant.[77] Victoria, Queensland and South Australia require that the protected attribute is a 'substantial reason' for the discrimination.[78] With respect to ML, a claimant would require proof that the protected attribute or a proxy for the protected attribute was a reason for the decision. Although the court is yet to apply the causation test to ML, the variables affecting the output should be considered as reflecting the 'real reason' for differential treatment. Given the varying complexities of ML models, the ML system in question would need to be examined to

---

[68] The legislation also provides exceptions to discrimination and prohibit discrimination in certain areas of life which are not discussed in this article.
[69] See, eg, *Racial Discrimination Act 1975* (Cth) s 9(1); *Anti-Discrimination Act 1977* (NSW) s 7; *Equal Opportunity Act 2010* (VIC) s 8.
[70] *Boehringer Ingelheim Pty Ltd v Reddrop* [1984] 2 NSWLR 13, 19.
[71] Rees, Allen and Rice (n 66) 80.
[72] Ibid.
[73] See Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31(2) *Harvard Journal of Law & Technology* 841, 848.
[74] *Purvis v New South Wales (Department of Education and Training)* (2003) 217 CLR 92, 106.
[75] Ibid 163 (Gummow, Hayne and Heydon JJ).
[76] See, eg, *Anti-Discrimination Act 1977* (NSW) s 4A.
[77] Rees, Allen and Rice (n 66) 106.
[78] Ibid.

determine whether the protected attribute was a variable which influenced the decision. For example, in deep learning neural networks, if the protected attribute was assigned weighting, this could indicate that it had a bearing on the final decision. It would be insufficient to assume that a protected attribute was a reason for a decision merely on the basis that the data related to the protected attribute was presented to the system. When applying this test in Victoria, Queensland and South Australia, a threshold weighting may be required to prove that it was a substantial reason for the decision.

2 *Lodging Claims under the Anti-Discrimination Framework and Remedial Action*

An individual alleging discrimination must lodge a complaint with their relevant anti-discrimination authority,[79] and in all states besides Victoria, individuals do not have a right to direct access to a court or tribunal.[80] Complaints must be lodged within 12 months of the alleged discriminatory conduct,[81] and complaints must provide a minimum level of content such as the identity of the alleged offender.[82] The authority may accept and investigate the complaint to determine if the conduct was unlawful,[83] and if it is unlawful, the parties may take part in a conciliation process.[84] If a complaint is referred to the administrative tribunal, a tribunal may order a range of remedies; however, the most sought after remedy in litigated cases is compensatory damages.[85]

B *Scoping the Barriers to Justice*

When ML is used in decision-making, the current anti-discrimination framework falls short of the objectives it sets out to achieve in two ways. Firstly, there is a higher chance of discrimination going undetected because of the veil of objectivity of ML systems. Secondly, ML systems increase the difficulty in proving a complaint relating to unlawful discrimination because of the inaccessibility of evidence and its lack of interpretability.

1 *Undetected Discrimination and the Veil of Objectivity*

Discrimination cases often rely on circumstantial evidence, making it difficult to establish the elements of the offence on the balance of probabilities.[86] ML decisions complicate this further because of the perceived objectivity of machines.[87] Kleinberg notes that this perception is created by humans viewing machines as technocratic and dispassionate.[88] The sources of bias discussed in Part II mostly permeate the system during the earlier implementation phases, thereby restricting the ability of complainants to

[79] Ibid 8.
[80] Ibid.
[81] See, eg, *Anti-Discrimination Act 1977* (NSW) s 89B(2)(b).
[82] Rees, Allen and Rice (n 66) 735.
[83] See, eg, *Anti-Discrimination Act 1977* (NSW) s 90.
[84] Rees, Allen and Rice (n 66) 753.
[85] Ibid 8.
[86] Ibid 144.
[87] Cofone (n 3) 1392.
[88] Kleinberg et al (n 5) 26.

identify the 'smoking gun' evidence of any resulting discrimination.[89] As a result, discrimination may be invisible to humans,[90] unless the differential treatment is intuitively apparent to an individual (for example, if an individual is aware that the other candidates for a job position were significantly inexperienced but males). Currently, individuals are not notified of the role of ML in a decision about them in Australia, and as a result of the 12-month timeframe, if an individual does not recognise or report discrimination within this time, no recourse is available. Additionally, an individual may have insufficient information to lodge a claim where they are unable to identify the source of the discrimination and the alleged perpetrator.

Not only does the inability to detect discrimination and the difficulty in lodging a claim affect the individual complainant; it also reinforces systemic discrimination as undetected discrimination perpetuates existing social structures which are adverse to the most vulnerable of society.[91] If a female or person identifying as a woman applies for a job position and is rejected on the basis of their sex or gender by the ML system (and if they have no reason to suspect discrimination), the system may continue to reject other women for similar positions. Over time, if the developers or managers of the system do not detect and rectify this bias, the machine will continue to shift the demographic and potentially the culture of the company or organisation. As more males are hired, it may become more difficult for females to progress through the company. Similar issues occurring across different industries may perpetuate the existing pay gaps and gendered stereotypes of what success in business ought to embody.[92] Interestingly, Mittelstadt also considers the disruption of personalisation as a result of the pressure to conform with decisions which are deemed favourable to a machine.[93] Existing oppressive practices reinforced by ML decision-making threaten the autonomy of vulnerable people, forcing conformity to appease standards set by machines.

2 *Proving Discrimination*

The limitations ML places on accessing and interpreting relevant evidence of discrimination creates significant barriers in lodging a case and proving discriminatory conduct.

(a) Inaccessibility of Relevant Evidence

In order to effectively establish unlawful discrimination, it is vital that all necessary evidence from the users or developers of ML is accessible. ML evidence is not as readily available as human evidence (ie verbal comments made to an individual) as the disclosure of algorithms is often protected by trade secrets laws, contractual obligations or intellectual property claims.[94] Bayamlioglu states that there exists a lack of transparency in the processes of algorithms through 'a culture of confidentiality and

[89] Ibid 17.
[90] Cofone (n 3) 1440.
[91] Castets-Renard (n 13) 94.
[92] Kleinberg et al (n 5) 43.
[93] Mittelstadt et al (n 2) 9.
[94] Andrea Roth, 'Machine Testimony' (2017) 126(7) *Yale Law Journal* 1972, 2028.

secrecy promulgated by the businesses, government or other organisations of interest, or in the form of legal claims primarily based on intellectual property rights and in particular trade secrets'.[95] While the reluctance to voluntarily disclose information detailing the processes of the system may be intended to prevent disclosure to their competitors,[96] it makes it incredibly difficult for those discriminated against to obtain sufficient evidence for a complaint to be accepted. In cases where the ML user procures the system from a third-party provider, evidence may even be inaccessible for the accused as a result of third party intellectual property rights.[97] Without voluntary disclosure, complainants would have great difficulty progressing a complaint to the conciliation phase. On appeal to a tribunal or court, the court can compel the disclosure of evidence which an ML user claims contains trade secrets and may choose to order a suppression order to prevent public disclosure of the protected information.[98] Since a complainant is only able to access evidence on appeal to a court, ML evidently frustrates the intended operation of the conciliation process of the anti-discrimination framework.

(b) Interpretability of Evidence

The difficulty of interpreting evidence of discrimination where ML is involved results from the inscrutable nature of ML. However, the extent to which ML systems are inscrutable is disputed by many. On the one hand, it is believed that most ML systems create opacity which results in an inability to apply human intuition to understand the decision made.[99] Shook, Smith and Antonio note that there are instances where 'even the creators may not understand how decisions are being made',[100] which results in ML systems being compared to 'black boxes'.[101] Inscrutability is contended to distinguish ML decision-making from human decision-making as a result of the inability to question the decision-maker under legal and compliance frameworks; thereby limiting the opportunities to understand how the decision was made.[102] In contrast, others suggest that human-decision makers can also be 'black boxes' meaning that inscrutability is not unique to automated decision-making.[103] This argument, however, is often accompanied by the view that opportunities exist to 'regulate the data in ways that one cannot do for human decision-makers'.[104] Kleinberg further argues that, in some cases, humans, unlike ML systems, cannot 'easily simulate counterfactuals';[105] making ML decisions easier to interrogate. Lehr and Ohm provide a balanced middle ground argument, noting that there are versions of explanations that can be made available, but that these explanations vary in usefulness and are

[95] Bayamlioglu (n 22) 436.
[96] Kleinberg et al (n 5) 41.
[97] Ibid.
[98] *Australian Broadcasting Commission v Parish* (1980) 29 ALR 228, 233 (Bowen CJ).
[99] Shook, Smith and Antonio (n 45) 446.
[100] Ibid.
[101] Frank Pasquale, *The Black Box Society* (Harvard University Press, 2015) 3.
[102] Shook, Smith and Antonio (n 45) 446.
[103] Cofone (n 3) 1439; Kleinberg et al (n 5) 18.
[104] Cofone (n 3) 1438.
[105] Kleinberg et al (n 5) 18.

specific to the individual ML system used.[106] This approach is the most comprehensive as it acknowledges the varying complexity of ML methods and processes, a mindset that is critical when considering the impact of ML systems holistically. Taking the Lehr and Ohm approach, the inscrutability of ML systems must be considered as existing on a spectrum of interpretability, with each systems' place being informed by its complexity. Two issues arise as a result of the interpretability of ML. Firstly, given the inconsistency of interpretability across different ML systems, understanding and interpreting which data is relevant to identifying discrimination is difficult for complainants. Secondly, in many cases, the evidence available requires specialised knowledge, thereby increasing the burden of pursuing a discrimination claim.

(i) Application of the Law

Inscrutability creates a significant hurdle to interpreting meaning from ML data. It is challenging to provide meaningful explanations of how the system works as the complete disclosure of the source code and processes is not always feasible. Arguably, ML systems present a greater opportunity for complainants to identify discrimination in comparison to cases involving only human decision-makers.[107] As Kirby J noted, the motivations of humans are complex and 'much discrimination occurs unconsciously, thoughtlessly or ignorantly'.[108] However, ML does not provide the requisite information to conclusively state that a particular decision made by the system relied on a protected attribute. In some cases, the algorithm used in the ML system may be less complicated and have fewer variables, meaning that it is easier for users to map out its operation to individuals.[109] In other cases (for example, where deep learning is used), the algorithm may become unexplainable even to its developers and the weight given to certain variables may not be as easy to identify.[110] The learning nature of ML also complicates the process of simulating counterfactuals. As the ML system learns, the algorithm processing the data is altered over time.[111] If a comprehensive version control system is not maintained, the task of facilitating counterfactuals at a later date becomes nearly impossible.[112] Consequently, while a complainant may have access to the inputted historical data or outputs of the ML system, in many cases, this offers little insight into the weight given to the variables affecting the decision.

The second element; causation; is more difficult to establish given the uncertainty as to what information is required from an ML system to satisfy the test. To establish causation, a complainant requires access to the inner workings of the system to examine the variables, or in some cases, the weighting attached to a variable relating to a protected attribute. Shook, Smith and Antonio note that

[106] Lehr and Ohm (n 23) 692.
[107] Kleinberg et al (n 5) 40.
[108] *IW v Perth* (1997) 191 CLR 1, 59.
[109] Andrew D Selbst and Solon Barocas, 'The Intuitive Appeal of Explainable Machines' (2018) 87(3) *Fordham Law Review* 1085, 1095–1096.
[110] Shook, Smith and Antonio (n 45) 446.
[111] Kleinberg et al (n 5) 20; Yanisky-Ravid and Hallisey (n 8) 450.
[112] Shook, Smith and Antonio (n 45) 446.

ML systems often do not create logs or have capabilities to produce reports to explain their processes.[113] Bayamlioglu further argues that the full disclosure of ML models is not 'legally possible or technically or economically feasible' as a result of the complexity of these data-centric systems.[114] Therefore, it could be difficult for a court and complainants to determine whether the mere existence of the protected attribute in the data or neural network is sufficient to establish causation, especially as there are likely multiple variables affecting the decision. Proxies may also be the reason for discrimination; for example, postcodes can be used as a proxy for low-socioeconomic status or ethnic backgrounds.[115] However, these are more difficult to identify and consequently overlooked when interrogating the reasons for discrimination.[116]

As a result, it will be unclear to complainants what evidence would sufficiently satisfy the elements of direct discrimination when lodging their claim or contesting a decision of the anti-discrimination authority. Due to complainants not having a right to legal representation during the complaints process,[117] a complaint declined by the authority would unlikely be contested if the complainant is unable to assess any available evidence and determine if an appeal is worth pursuing, particularly given the legal costs of an appeal and the potentially minimal compensation awarded.

(ii) Requirement of Specialised Knowledge

If the relevant evidence is not available by way of a non-technical explanation, the complainant requires specialised knowledge of the ML system to interpret and present the evidence to the relevant authority.[118] The inability to comprehend the decision-making process, in addition to the limited access to information, contributes to reduced transparency and trust in ML systems.[119] Experts with specialised knowledge would need to act as expert witnesses,[120] however; sometimes, this knowledge may still be insufficient.[121] The learning nature of ML and its ability to develop beyond a point where its creator can understand its processes means that specialised knowledge may still limit the interpretability of the evidence.[122] Given that the onus is on the complainant to prove discrimination,[123] the burden of obtaining expert evidence may ultimately discourage individuals from voicing their adverse treatment by pursuing a complaint.

---

[113] Ibid.
[114] Bayamlioglu (n 22) 443.
[115] Yanisky-Ravid and Hallisey (n 8) 449.
[116] Mittelstadt et al (n 2) 8.
[117] See, eg, *Anti-Discrimination Act 1977* (NSW) s 91B.
[118] Selbst and Barocas (n 109) 1090.
[119] Ibid 1093.
[120] Kleinberg et al (n 5) 5.
[121] Joshua Kroll et al, 'Accountable Algorithms' 165(3) *University of Pennsylvania Law Review* 633, 638.
[122] Ibid.
[123] Rees, Allen and Rice (n 66) 142.

IV EVALUATING THE APPROACHES TO MACHINE LEARNING REGULATION

The barriers identified above reduce the effectiveness of the Australian anti-discrimination framework by making cases of discrimination more difficult to detect and reprimand. As Australia is in its early stages of ML and AI regulating, a strong foundation must be created by approaching ML regulation in a purposeful manner. As Gaon and Stedman note, sweeping laws are not appropriate for regulating AI given how rapidly technology can develop and change.[124] The AHRC has proposed recommendations which mirror aspects of international regulatory approaches to ML and AI.[125] However, the AHRC's proposals reflect a scattergun approach; combining the creation of individual rights with proposals for an oversight regime. This approach ultimately fails to appreciate the difference in the effectiveness of these distinct approaches and to explicitly consider how each approach would address ML discrimination. This article addresses this gap by creating a classification of the two main approaches to ML and AI regulation and assesses the effectiveness of the approaches with respect to how sufficiently they bridge the barriers identified in this article.

Accordingly, this Part will consider whether it would it be more effective for ML regulations to bridge the barriers by: (i) empowering individuals with rights to contest, and access information from, ML-informed decisions; or (ii) creating a regulator responsible for overseeing ML and identifying, reporting and penalising discriminatory outcomes. The best approach is that which most effectively addresses the barriers to the anti-discrimination framework, is appropriate in the context of the existing legal system and is considerate towards the future of the Australian ML industry, irrespective of changes to the anti-discrimination framework. It is not suggested that these approaches cannot be used together, nor that they are not interrelated, or that they are the only approaches available. Instead, it considers existing approaches to regulation already being contemplated in Australia in a manner which does not preclude the introduction of new techniques or research methods which may result in new methods of regulating ML discrimination which have yet to be implemented. While this article focuses on discrimination, this evaluation should be performed across any existing legal frameworks which are rendered less effective as a result of ML.

A *Individual Rights versus Accountability and Oversight*

The primary purpose of an individual rights approach is to empower individuals to contest ML-informed decisions to which they are subjected. An Australian individual rights approach would likely consist of a 'right to explanation' constructed through rights to access information from users of ML.[126] The AHRC proposes that an individual should be able to 'demand: (a) a non-technical explanation of the

[124] Aviv Gaon and Ian Stedman, 'A Call to Action: Moving Forward with the Governance of Artificial Intelligence in Canada' (2019) 56(4) *Alberta Law Review* 1137, 1164.
[125] Sophie Farthing et al (n 12) 189-192.
[126] Sophie Farthing et al (n 12) 190.

AI-informed decision, which would be comprehensible by a lay person, and (b) a technical explanation of the AI-informed decision that can be assessed and validated by a person with relevant technical expertise'.[127] Where a system is unable to generate 'reasonable explanations', the AHRC proposes that those machines are not deployed 'where decisions could infringe the human rights of individuals'.[128] This approach could also consist of rights for individuals to challenge ML-informed decisions and request human oversight. These proposals mirror similar rights created under the European Union's General Data Protection Regulation (the 'GDPR').[129] Article 22 provides a right to not 'be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her'.[130] There are, however, exceptions to this right; including where a data controller implements measures to safeguard the individual's rights, freedoms and legitimate interests or where consent is obtained.[131] The GDPR's right to explanation has been critiqued for the impediments created by trade secrets protections,[132] the limited remedies available to individuals discriminated against and its failed uniformity across the Member States of the EU.[133]

On the other hand, an accountability and oversight approach focuses on enforcing the law via regulations and obligations imposed on ML users. It is distinct from an individual rights approach as it seeks to enforce the law through accountability to the government, as opposed to accountability to an individual. It places greater emphasis on ongoing compliance and uses penalties as a punitive measure to prevent non-compliance. An accountability and oversight regime in Australia would likely require ML users to remain accountable to a central oversight body. Commonly proposed accountability and oversight measures include mandatory audits and reports,[134] impact statements[135] or compliance with design and industry standards prescribed by the oversight body.[136] The oversight body can also broaden the ambit of enforceability through mandating standards for developers or suppliers of ML. The accountability and oversight approach has been heavily advocated by US researchers and academics, with an FDA-like regulator proposed to maintain oversight of AI and ML.[137] A lighter touch 'policy-first' approach has been proposed in Canada, with governance and accountability as the central focus.[138] Certain articles of the GDPR, such as the required data controller risk reports, data protection impact

[127] Ibid.
[128] Ibid.
[129] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)* [2016] OJ L 119/1.
[130] Ibid art 22(1).
[131] Ibid arts 22(2)-(3).
[132] Edwards and Veale (n 7) 53.
[133] Castets-Renard (n 13) 137.
[134] See Bryce Goodman, 'Discrimination, Data Sanitisation and Auditing in the European Union's General Data Protection Regulation' (2016) 2(4) *European Data Protection Law Review* 493.
[135] See Barocas and Selbst (n 40) 712.
[136] See Gaon and Stedman (n 124) 1160.
[137] Andrew Tutt, 'An FDA for Algorithms' (2017) 69(1) *Administrative Law Review* 83.
[138] Gaon and Stedman (n 124) 1164.

assessments and voluntary certifications;[139] while not necessarily specific to automated decision-making, are also reminiscent of the mechanics of an accountability and oversight approach.

B *Which Approach Would Most Effectively Address the Barriers?*

In order to ensure that the Australian anti-discrimination framework maintains its current effectiveness and continues to meet its objectives, Australia must focus on an approach to regulation which addresses the barriers identified in Part III.

1 *Undetectability*

An effective regulatory approach needs to enable the identification of differential treatment and the reasons for that treatment. An individual rights approach could address the undetectability of ML discrimination if the government invests in educating the public on how and when they can exercise these rights.[140] It is argued that providing individuals with a right to an explanation allows them to identify the misuse of their information.[141] However, an inherent issue with an individual rights approach is that it requires individuals to exercise their rights actively. Given ML's somewhat covert operation and the reduced presence of human intuition to voice concerns, the active exercise of rights could be difficult to achieve without creating public paranoia or requiring mandatory notification of ML's involvement in a decision-making process. Further, even where an individual did exercise these rights, the information disclosed may not provide information relating to comparators and may not allow an untrained eye to identify the reasons for discrimination.[142] ML is also unique to other information already required to be disclosed under Australian law. Unlike privacy statements which must be provided to individuals by companies subject to privacy laws,[143] ML users would need to provide individualised non-technical and technical explanations; which is comparatively a more burdensome than the AHRC's proposed rights. Further, notifications requiring users to explain the purpose of, or types of data processed by, the ML system could inadvertently portray the system as objective.[144] Therefore, these notifications fail to address the lack of human intuition which would ordinarily detect discrimination. Without such intuition, discrimination in ML may continue to go undetected and impact the lives of numerous individuals.

Conversely, accountability measures can provide transparency of the processes of the system and the outputs to an oversight body,[145] thereby rendering an accountability and oversight approach more

[139] Goodman (n 134) 498.
[140] See Castets-Renard (n 13) 136.
[141] Yanisky-Ravid and Hallisey (n 8) 454.
[142] Wachter, Mittelstadt and Russell (n 73) 848.
[143] See *Privacy Act 1988* (Cth) sch 1 cl 1.3.
[144] Danielle Keats Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89(1) *Washington Law Review* 1, 14.
[145] Ibid 20.

effective in detecting and mitigating discrimination. Auditing would be particularly useful as a part of the accountability and oversight approach because it can create consistent oversight of decisions informed by ML. ML systems can also be audited for their compliance with social values,[146] thereby identifying discrimination or potential risks of discrimination within systems and providing ML users with the opportunity to rectify non-compliance.

Although audits have proven to be effective at identifying discrimination,[147] the frequency and scope of audits could impact the effectiveness of this strategy in Australia.[148] To sufficiently detect discrimination, audits would need to be conducted during the system's implementation and periodically or continuously post-deployment.[149] The latter is essential to ensuring that complaints can be lodged within the 12 months required by the anti-discrimination framework. A benefit of auditing and similar accountability measures is that they provide flexibility and can be tailored depending on the type of ML system used and the risk of discrimination posed by the individual system.[150] For example, for simpler systems, auditing may not be considered necessary where the ML will not make a significant decision about a person. Further, if auditing is considered necessary, periodic auditing can be used for less complex systems given the practicality of maintaining version control.[151] This would also alleviate the financial burden of continuous auditing for these simpler systems.[152] For more complex systems, continuous auditing could ensure that appropriate records are maintained, data for counterfactuals is available, and discrimination is quickly identified. However, the time and financial burdens associated with continuous auditing may act as a disincentive to the use of more complex systems, thereby reducing the availability of more advanced, accurate and useful ML systems.

Despite this, mandatory auditing could importantly result in increased public trust in ML.[153] With the growing emphasis on the corporate social responsibility of companies[154] and the increasing role of ML in health systems, as seen in its role during the COVID-19 pandemic, greater public trust in ML is imperative. Increasing trust is the most beneficial way to regulate from an economic welfare approach,[155] as the economic benefits of consumer trust will offset the cost of auditing and the cost of training ML sufficiently.[156] Failure to take this approach could result in the slowing of the ML industry

[146] Waldman (n 60) 632.
[147] Bryan Casey, Ashkon Farhangi and Roland Vogl, 'Rethinking Explainable Machines: The GDPR's Right to Explanation Debate and the Rise of Algorithmic Audits in Enterprise' (2019) 34(1) *Berkeley Technology Law Journal* 143, 182.
[148] Goodman (n 134) 503.
[149] Ibid 504.
[150] See Gaon and Stedman (n 124) 1161.
[151] Goodman (n 134) 504.
[152] Ibid.
[153] Yanisky-Ravid and Hallisey (n 8) 475.
[154] Michael R Siebecker, 'Making Corporations More Humane through Artificial Intelligence' (2019) 45(1) *Journal of Corporation Law* 95, 141.
[155] Bronwen Morgan and Karen Yeung, *An Introduction to Law and Regulation* (Cambridge University Press, 2007) 18.
[156] Yanisky-Ravid and Hallisey (n 8) 479.

and developments,[157] thereby causing Australia to fall behind the international competition. It is, however, noted that the public trust created by an accountability and oversight approach might not apply where government organisations are using ML systems.[158] This is particularly an issue in Australia given the recent misuse of technology by the government.[159] However, governments already have the means to facilitate auditing of technology,[160] and doing so transparently and independently could repair existing mistrust. As audit non-compliance is ordinarily penalised, this can work to incentivise the industry to maintain rigorous oversight of their systems past the implementation stage.[161] There are some issues with auditing measures such as the possible gaming of periodic reviews and the cost of continuous auditing,[162] however, these issues can be addressed through the collaboration of industry and the oversight body to better understand the limitations of the technology. Further, to address any mistrust of government uses of ML, the AHRC could be heavily involved in shaping auditing requirements and providing their human rights expertise where necessary (ie to inform the social standards for which systems are audited).

2 *Inaccessibility of Relevant Evidence*

The most effective approach to addressing the accessibility barrier will not only ensure greater accessibility to evidence, but also ensure that this evidence is accessible from any relevant parties (whether that be the user, developer or provider of ML). Where an individual suspects discrimination, an individual rights approach would allow them to request direct access to the information. However, individuals must be granted access to the *relevant* evidence required to lodge and prove a complaint. Similar access rights are not foreign to Australia, particularly when it concerns government decision-making.[163] However, extending explanation and access rights to companies using ML systems would be more complicated and likely viewed unfavourably by the ML industry considering that greater access increases the risk of competitors reverse-engineering their systems.[164] Access rights may also be redundant if the right conflicts with intellectual property or trade secret laws which ultimately prevent access.[165] Further, the effectiveness of the right to explanation under the GDPR is contingent on a series of rights afforded to individuals relating to data protection and privacy.[166] Australia's privacy laws create obligations relating to privacy and penalise companies failing compliance.[167] Requiring

[157] Ibid 475.
[158] Berman (n 62) 1324.
[159] See Luke Henriques-Gomes, 'Robodebt: Government Admits It Will Be Forced to Refund $550m Under Botched Scheme', *The Guardian* (online at 27 March 2020) <https://www.theguardian.com/australia-news/2020/mar/27/robodebt-government-admits-it-will-be-forced-to-refund-550m-under-botched-scheme>.
[160] Waldman (n 60) 631.
[161] Cofone (n 3) 1442.
[162] Goodman (n 134) 504.
[163] For example, a right of access exists under s 11 of the *Freedom of Information Act 1982* (Cth).
[164] Bayamlioglu (n 22) 444.
[165] Patrick W Nutter, 'Machine Learning Evidence: Admissibility and Weight' (2019) 21(3) *University of Pennsylvania Journal of Constitutional Law* 919, 942.
[166] Goodman (n 134) 495.
[167] See, eg, *Privacy Act 1988* (Cth) ss 14, 13G.

individuals to actively exercise rights would be inconsistent with the existing privacy framework and could result in misunderstandings of how these rights interact or conflict with existing obligations.

In contrast, an accountability and oversight approach is unlikely to broaden accessibility to complainants. However, it could instead allow greater access for the relevant discrimination authority, courts and the oversight body. The oversight body could penalise those in breach of discrimination laws which could deter discrimination. Punitive action could result in developers and users giving greater consideration to potential ML discrimination when designing and deploying systems.[168] However, to ensure individuals can access remedies afforded in an anti-discrimination case, any suspected contravention of the anti-discrimination framework would need to be reported to the discrimination authority, and a complaint must be lodged on the individual's behalf. This approach presents the opportunity to address the accessibility barrier while also implementing measures to reduce the presence of discriminatory outcomes. While it may not necessarily increase accessibility for individuals, it is essential to consider the purpose of the anti-discrimination laws. On the one hand, it provides individuals with the ability to access remedies where they have been discriminated against; however, its primary purpose is to reduce and mitigate discrimination. As Cofone argues, prevention of harms of discrimination is more valuable than remedying them.[169] While individuals may have less involvement in this process under the accountability and oversight approach which some believe reduces individual autonomy,[170] it is still the more effective way to reduce and reprimand discrimination in Australia by increasing the accessibility of evidence by the court and discrimination authorities.

Additionally, a practical approach would need to identify from whom information can be accessed. If access or auditing is restricted to the party using the system for a decision, information may be inaccessible given the contractual obligations between the party and its ML supplier. The likely exceptions to an access right (ie where disclosure would be inconsistent with trade secret laws) would likely not cater to these circumstances. For auditing or mandatory disclosures, on the other hand, parties could opt to contractually agree to permit disclosure of their source code where required by law or regulators. While auditing would allow a broader scope of accessibility, requiring all procurers and suppliers of ML systems to audit could deter the use of ML systems because of the associated financial burden.[171] Deterrence from using ML systems could reduce competition, causing a monopoly of the ML market by larger companies with the necessary resources to audit. This is primarily because the audit costs could outweigh the opportunity costs of developing ML for smaller businesses or start-up ventures.[172] Nonetheless, an accountability and oversight approach vests power in an oversight body to amend and extend auditing requirements to ensure wide-reaching enforceability of ML regulations. An oversight body could require differing levels of auditing or standards compliance for the various types

[168] Shook, Smith and Antonio (n 45) 461.
[169] Cofone (n 3) 1440.
[170] Kaminski (n 6) 1541.
[171] Kleinberg et al (n 5) 41.
[172] Sophie Farthing et al (n 12) 38.

of ML. These auditing requirements could depend on the complexity of the system and the risk assessment of the impact of a system's decisions.[173] Even where not all users in the supply chain are regulated via mandatory auditing, creating a clear compliance framework will ensure all members of the chain are held accountable (for example, by prescribing mandatory standards for engineers of ML).[174] Ultimately, assigning the responsibility to review and account for ML discrimination to users and developers, rather than individuals, can foster an industry culture of transparency and socially responsible uses of technology.[175]

3 *Interpretability of Evidence*

Closing the gap of technological illiteracy is a costly task, and even the most comprehensive public education campaign may not provide the specialised knowledge required to interpret ML.[176] Individual rights would, therefore, be ineffective in addressing this issue. Further, diluting explanations of ML's processes to allow for easier interpretation can compromise the performance of the system.[177] On the other hand, an accountability and oversight approach could bring together the knowledge of industry experts and well-trained officials. This knowledge could ensure that the oversight body can assist or provide guidance where state or territory authorities are unable to interpret data relating to an anti-discrimination complaint. Having an oversight body equipped with specialised knowledge will also ensure that future developments in the law are well-informed.[178]

With respect to the application of the law, individual rights would only be sufficient at addressing barriers to establishing differential treatment if they entail broad access rights. As discussed, individuals would require access to data which allows them to facilitate counterfactuals or be privy to the outcomes produced by the system for other individuals. Providing such access to the latter could render users of the systems in breach of privacy obligations owed to other individuals. It would also not be appropriate in all circumstances, for example, where the consent of the other individuals is required or where disclosure would be prejudicial to others. Alternatively, an accountability and oversight approach could ensure that a central body has the necessary expertise to examine and facilitate counterfactuals. Some scholars argue that understanding the purpose and the 'normativity embedded in their behaviour/action' can provide equally useful understandings of the functioning of the system.[179] Even in the case where lesser information is considered to be necessary to interpret the system, an audit body could ensure that the minimum amount necessary to run such counterfactuals is provided.

---

[173] Goodman (n 134) 506.
[174] Gaon and Stedman (n 124) 1160. See also KPMG Australia, *Human Rights and Technology in 2020 and Beyond* (Submission, March 2020) 5.
[175] Michael Guihot, Anne F Matthew and Nicolas P Suzor, 'Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence' (2017) 20(2) *Vanderbilt Journal of Entertainment & Technology Law* 385, 446.
[176] Cofone (n 3) 1439.
[177] Castets-Renard (n 13) 61.
[178] See Gaon and Stedman (n 124) 1164.
[179] Bayamlioglu (n 22) 434.

As discussed in Part III(B), access to the variables and understanding how they have been processed would be most useful when attempting to establish causation. Individual rights may provide access to information, but where that information is uninterpretable, it is of little use to complainants. The information provided would need to be 'sufficiently comprehensive' for individuals to understand why a decision was made.[180] High-level explanations may provide individuals with a superficial understanding of how the machine works, but would not necessarily allow them to identify the relevant variables and obtain the necessary evidence. It would also be difficult for individuals to identify which information has been used as a proxy for their protected attribute. Bayamlioglu states that clear requirements outlining what information must be interpretable or produced by the system need to be provided to system users from the date of development.[181] An oversight body could prescribe such requirements; however, if the information is inscrutable even to developers, it may not be useful at all. Further, eliminating variables or simplifying systems to maximise explainability would result in the loss of essential and relevant information that is critical for improving future predictions.[182] Mandatory requirements could ensure mitigation measures identify, consider or address proxy variables, without eliminating them. This could potentially deter the use of more complex systems which may be more appropriate for the intended purpose or function of the ML system. From a commercial perspective, compliance with such requirements could be costly, poorly understood by developers and result in them falling behind the international competition.

Nonetheless, an accountability and oversight approach would be more effective than the 'meaningless transparency' which would be created through individual rights.[183] It would allow greater flexibility for collaboration between industry experts and the oversight body to develop ways to produce interpretable information.[184] As Castets-Renard explains, regulations should place greater focus on the creation of better systems and the empowerment of agencies to review systems for bias, rather than merely attempting to challenge ML decisions on an individual by individual basis.[185] Oversight processes can also shape how existing laws are amended by informing legislators of what information is interpretable and meaningful.[186] As the application of the causation test to ML systems develops to address issues of identification of liability and the application of the test to machines, the newfound knowledge of the oversight body can shape 'which features should be considered' in determining how decisions are to be made.[187] As a result, an accountability and oversight focused approach means that the industry can continue to use ML systems to their fullest potential without limiting their use of ML, while also championing human rights. Overall, accountability regimes would reduce the issue of

[180] Castets-Renard (n 13) 120.
[181] Bayamlioglu (n 22) 442.
[182] Edwards and Veale (n 7) 61.
[183] Ibid 23.
[184] Gaon and Stedman (n 124) 1164.
[185] Castets-Renard (n 13) 23.
[186] Goodman (n 134) 506.
[187] Bayamlioglu (n 22) 442.

undetectability, provide greater access to relevant information required to establish a discrimination claim and encourage the creation of fair and equitable machines.

V CONCLUSION

This article has considered how the increasing presence of ML in everyday life threatens to increase undetected and unchecked discrimination. Through classifying the literature and international approaches into two categories; those which promote individual rights and those which advocate for an accountability and oversight approach; it has evaluated which of these approaches is most appropriate for the Australian context. These early stages of the law's development provide a notable opportunity for policymakers to create regulations which champion existing protections of individual human rights while also future-proofing the ML industry. While human rights may not be the only area of concern, it is one of critical importance. Breaches of human rights by machines pose threats to the autonomy of humans and can have detrimental impacts on the most vulnerable groups of society.[188] Academics, industry and the government must work together to conduct similar detailed enquiries into the impacts of ML on other existing legal frameworks. The risk of grouping issues together is that it does not allow for thorough consideration of the impacts of ML on existing legal frameworks. This article has concluded that with respect to discrimination in Australia, an accountability and oversight approach would most appropriately address the ML barriers while still fostering innovation in Australia's growing ML sector.

[188] Ibid 445.

BIBLIOGRAPHY

A *Articles/Books*

Ajunwa, Ifeoma, 'Age Discrimination by Platforms' (2019) 40(1) *Berkeley Journal of Employment and Labor Law* 1

Alpaydin, Ethem, *Introduction to Machine Learning* (Massachusetts Institute of Technology Press, 3rd ed, 2014)

Altman, Micah, Alexandra Wood and Effy Vayena, 'A Harm-Reduction Framework for Algorithmic Fairness' (2018) 16(3) *IEEE Security & Privacy* 34

Bambauer, Jane and Tal Zarsky, 'The Algorithm Game' (2018) 94(1) *Notre Dame Law Review* 1

Barocas, Solon and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104(3) *California Law Review* 671

Bayamlioglu, Emre, 'Contesting Automated Decisions' (2018) 4(4) *European Data Protection Law Review* 433

Berman, Emily, 'A Government of Laws and Not of Machines' (2018) 98(5) *Boston University Law Review* 1277

Bruckner, Matthew Adam, 'The Promise and Perils of Algorithmic Lenders' Use of Big Data' (2018) 93(1) *Chicago-Kent Law Review* 3

Callier, Michael and Harly Callier, 'Blame It on the Machine: A Socio-Legal Analysis of Liability in an AI World' (2018) 14(1) *Washington Journal of Law, Technology & Arts* 49

Calo, Ryan, 'Artificial Intelligence Policy: A Primer and Roadmap' (2017) 51(2) *University of California, Davis Law Review* 399

Casey, Bryan, Ashkon Farhangi and Roland Vogl, 'Rethinking Explainable Machines: The GDPR's Right to Explanation Debate and the Rise of Algorithmic Audits in Enterprise' (2019) 34(1) *Berkeley Technology Law Journal* 143

Castets-Renard, Celine, 'Accountability of Algorithms in the GDPR and Beyond: A European Legal Framework on Automated Decision-Making' (2019) 30(1) *Fordham Intellectual Property, Media & Entertainment Law Journal* 91

Citron, Danielle Keats and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89(1) *Washington Law Review* 1

Cofone, Ignacio N, ‘Algorithmic Discrimination Is an Information Problem’ (2019) 70(6) *Hastings Law Journal* 1389

Coglianese, Cary and David Lehr, ‘Regulating by Robot: Administrative Decision Making in the Machine-Learning Era’ (2017) 105(5) *Georgetown Law Journal* 1147

Cuatrecasas, Carlota, ‘Legal Challenges of Artificial Intelligence (AI)’ (2020) 1(1) *Global Privacy Law Review* 8

Desai, Deven R and Joshua A Kroll, ‘Trust but Verify: A Guide to Algorithms and the Law’ (2017) 31(1) *Harvard Journal of Law & Technology* 1

Edwards, Lilian and Michael Veale, ‘Slave to the Algorithm? Why a “Right to an Explanation” Is Probably Not the Remedy You Are Looking For’ (2017) 16 *Duke Law & Technology Review* 18

Fagan, Frank and Saul Levmore, ‘The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion’ (2019) 93(1) *Southern California Law Review* 1

Fuchs, Daniel James, ‘The Dangers of Human-Like Bias in Machine-Learning Algorithms’ (2018) 2(1–14) *Missouri S&T’s Peer to Peer* 15

Gaon, Aviv and Ian Stedman, ‘A Call to Action: Moving Forward with the Governance of Artificial Intelligence in Canada’ (2019) 56(4) *Alberta Law Review* 1137

Gillis, Talia B and Jann L Spiess, ‘Big Data and Discrimination’ (2019) 86 *University of Chicago Law Review* 459

Giuffrida, Iria, ‘Liability for AI Decision-Making: Some Legal and Ethical Considerations’ (2019) 88(2) *Fordham Law Review* 439

Goodman, Bryce, ‘Discrimination, Data Sanitisation and Auditing in the European Union’s General Data Protection Regulation’ (2016) 2(4) *European Data Protection Law Review* 493

Goodman, Chris Chambers, ‘AI/Esq: Impacts of Artificial Intelligence in Lawyer-Client Relationships’ (2019) 72(1) *Oklahoma Law Review* 149

Guihot, Michael, Anne F Matthew and Nicolas P Suzor, ‘Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence’ (2017) 20(2) *Vanderbilt Journal of Entertainment & Technology Law* 385

Hacker, Philip and Bilyana Petkova, ‘Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers’ (2017) 15(1) *Northwestern Journal of Technology and Intellectual Property* 1

Hertza, Vlad, 'Fighting Unfair Classifications in Credit Reporting: Should the United States Adopt GDPR-Inspired Rights in Regulating Consumer Credit?' (2018) 93(6) *New York University Law Review* 1707

Hewitt, Anne, 'Can a Theoretical Consideration of Australia's Anti-Discrimination Laws Inform Law Reform' (2013) 41(35) *Federal Law Review* 37

Hildebrandt, Mireille, 'Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning' (2019) 20(1) *Theoretical Inquiries in Law* 83

Howard, Ayanna and Jason Borenstein, 'The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity' (2018) 24(5) *Science and Engineering Ethics* 1521

Huq, Aziz Z, 'Racial Equity in Algorithmic Criminal Justice' (2019) 68(6) *Duke Law Journal* 1043

Hurley, Mikella and Julius Adebayo, 'Credit Scoring in the Era of Big Data' (2016) 18 *Yale Journal of Law and Technology* 148

Joh, Elizabeth E, 'Artificial Intelligence and Policing: First Questions' (2018) 41(4) *Seattle University Law Review* 1139

Jolls, Christine and Cass R Sunstein, 'The Law of Implicit Bias' (2006) 94(4) *California Law Review* 969

Kaminski, Margot, 'The Right to Explanation, Explained' (2019) 34(1) *Berkeley Technology Law Journal* 189

Kaminski, Margot E, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability' (2019) 92(6) *Southern California Law Review* 1529

Katyal, Sonia K, 'Private Accountability in the Age of Artificial Intelligence' (2019) 66 *University of California Law Review* 54

Keeling, Robert et al, 'Using Machine Learning on Legal Matters: Paying Attention to the Data Behind the Curtain' (2020) 11 *Hastings Science and Technology Law Journal* 9

Kim, Pauline T, 'Auditing Algorithms for Discrimination' (2017) 166 *University of Pennsylvania Law Review Online* 189

Kim, Pauline T, 'Big Data and Artificial Intelligence: New Challenges for Workplace Equality' (2019) 57 *University of Louisville Law Review* 313

Kleinberg, Jon et al, 'Discrimination in the Age of Algorithms' (2018) 10(1) *Journal of Legal Analysis* 1

Kroll, Joshua et al, 'Accountable Algorithms' 165(3) *University of Pennsylvania Law Review* 633

Larsson, Stefan, 'The Socio-Legal Relevance of Artificial Intelligence' (2019) 103 *Droit et Societe* 573

Lehr, David and Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn about Machine Learning' (2017) 51 *University of California, Davis Law Review* 653

Levendowski, Amanda, 'How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem' (2018) 93(2) *Washington Law Review* 579

Liu, Hin-Yan, 'Three Types of Structural Discrimination Introduced by Autonomous Vehicles' (2017) 51 *University of California, Davis Law Review* 149

MacCarthy, Mark, 'Standards of Fairness for Disparate Impact Assessment of Big Data Algorithms' (2017) 48(1) *Cumberland Law Review* 67

Mainka, Spencer M, 'Algorithm-Based Recruiting Technology in the Workplace' (2019) 5(3) *Texas A&M Journal of Property Law* 801

Mayson, Sandra G, 'Bias in, Bias Out' (2019) 128(8) *Yale Law Journal* 2218

McPeak, Agnieszka, 'Disruptive Technology and the Ethical Lawyer' (2019) 50(3) *University of Toledo Law Review* 457

Mittelstadt, Brent Daniel et al, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3(2) *Big Data & Society* 1

Molnar, Christoph, *Interpretable Machine Learning* (Lulu, 2019)

Morgan, Bronwen and Karen Yeung, *An Introduction to Law and Regulation* (Cambridge University Press, 2007)

Moses, Lyria Bennett and Louis de Koker, 'Open Secrets: Balancing Operational Secrecy and Transparency in the Collection and Use of Data by National Security and Law Enforcement Agencies' (2017) 41(2) *Melbourne University Law Review* 530

Murphy, Kevin P, *Machine Learning: A Probabilistic Perspective* (Massachusetts Institute of Technology Press, 2012)

Nutter, Patrick W, 'Machine Learning Evidence: Admissibility and Weight' (2019) 21(3) *University of Pennsylvania Journal of Constitutional Law* 919

O'Cinneide, Colm, 'The Uncertain Foundation of Contemporary Anti-Discrimination Law' (2011) 11 *International Journal of Discrimination and the Law* 7

Odinet, Christopher K, 'Consumer Bitcredit and Fintech Lending' (2018) 69(4) *Alabama Law Review* 781

Pasquale, Frank, *The Black Box Society* (Harvard University Press, 2015)

Peppet, Scott R, 'Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent' (2014) 93(1) *Texas Law Review* 85

Raymond, Anjanette H, Emma Arrington Stone Young and Scott J Shackelford, 'Building a Better HAL 9000: Algorithms, the Market, and the Need to Prevent the Engraining of Bias' (2018) 15(3) *Northwestern Journal of Technology and Intellectual Property* 215

Rees, Neil, Dominique Allen and Simon Rice, *Australian Anti-Discrimination Law* (Federation Press, 2nd ed, 2014)

Richardson, Rashida, Jason Schultz and Kate Crawford, 'Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice' (2019) 192 *New York University Law Review* 192

Roth, Andrea, 'Machine Testimony' (2017) 126(7) *Yale Law Journal* 1972

Selbst, Andrew D, 'Disparate Impact in Big Data Policing' (2017) 52(1) *Georgia Law Review* 109

Selbst, Andrew D and Solon Barocas, 'The Intuitive Appeal of Explainable Machines' (2018) 87(3) *Fordham Law Review* 1085

Selbst, Andrew D and Julia Powles, 'Meaningful Information and the Right to Explanation' (2017) 7(4) *International Data Privacy Law* 233

Shook, Jim, Robyn Smith and Alex Antonio, 'Transparency and Fairness in Machine Learning Applications' (2018) 4(5) *Texas A&M Journal of Property Law* 443

Siebecker, Michael R, 'Making Corporations More Humane through Artificial Intelligence' (2019) 45(1) *Journal of Corporation Law* 95

Snyder, Timothy M, 'You're Fired: A Case for Agency Moderation of Machine Data in the Employment Context' (2016) 24(1) *George Mason Law Review* 243

Stanila, Laura, 'Artificial Intelligence and Human Rights: A Challenging Approach on the Issue of Equality' (2018) 2018(2) *Journal of Eastern-European Criminal Law* 19

Stilgoe, Jack, 'Machine Learning, Social Learning and the Governance of Self-Driving Cars' (2018) 48(1) *Social Studies of Science* 25

Tutt, Andrew, 'An FDA for Algorithms' (2017) 69(1) *Administrative Law Review* 83

Wachter, Sandra, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31(2) *Harvard Journal of Law & Technology* 841

Waldman, Ari Ezra, 'Power, Process, and Automated Decision-Making' (2019) 88(2) *Fordham Law Review* 613

Yanisky-Ravid, Shlomit and Sean K Hallisey, 'Equality and Privacy by Design: A New Model of Artificial Intelligence Data Transparency via Auditing, Certification, and Safe Harbour Regimes' (2019) 46(2) *Fordham Urban Law Journal* 428

B *Cases*

Australian Broadcasting Commission v Parish (1980) 29 ALR 228

Boehringer Ingelheim Pty Ltd v Reddrop [1984] 2 NSWLR 13

IW v Perth (1997) 191 CLR 1

Purvis v New South Wales (Department of Education and Training) (2003) 217 CLR 92

C *Domestic Legislation*

Age Discrimination Act 2004 (Cth)

Anti-Discrimination Act 1977 (NSW)

Anti-Discrimination Act 1991 (Qld)

Anti-Discrimination Act 1992 (NT)

Anti-Discrimination Act 1998 (Tas)

Disability Discrimination Act 1992 (Cth)

Discrimination Act 1991 (ACT)

Equal Opportunity Act 1984 (SA)

Equal Opportunity Act 1984 (WA)

Equal Opportunity Act 2010 (Vic)

Freedom of Information Act 1982 (Cth)

Privacy Act 1988 (Cth)

Racial Discrimination Act 1975 (Cth)

Sex Discrimination Act 1984 (Cth)

D *Parliamentary Debates*

Commonwealth, *Parliamentary Debates*, House of Representatives, 13 February 1975

E *International Materials*

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1

Universal Declaration of Human Rights, GA Res 217A (III), UN GAOR, UN Doc A/810 (10 December 1948)

F *Discussion Papers/Government Reports/Research Papers*

Australian Human Rights Commission, 'Human Rights and Technology Issues Paper' (Paper, July 2018) <https://tech.humanrights.gov.au/our-work>

Davis, Nicholas et al, 'Artificial Intelligence: Governance and Leadership' (White Paper, Australian Human Rights Commission and World Economic Forum, 2019) <https://tech.humanrights.gov.au/our-work>

Farthing, Sophie et al, 'Human Rights and Technology Discussion Paper' (Paper, Australian Human Rights Commission, December 2019) <https://tech.humanrights.gov.au/our-work>

Hajkowicz, Stefan et al, 'Artificial Intelligence: Solving Problems, Growing the Economy and Improving our Quality of Life' (Report, Commonwealth Scientific and Industrial Research Organisation, 2019)

KPMG Australia, *Human Rights and Technology in 2020 and Beyond* (Submission Paper, March 2020)

Wachter, Sandra, 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising' [2019] *SSRN Electronic Journal* <https://www.ssrn.com/abstract=3388639>

G *Magazine Articles/Online Newspaper Articles*

Columbus, Louis, 'Roundup of Machine Learning Forecasts and Market Estimates: 2020', *Forbes* (online at 19 January 2020) <https://www.forbes.com/sites/louiscolumbus/2020/01/19/roundup-of-machine-learning-forecasts-and-market-estimates-2020/>

Dastin, Jeffrey, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women', *Reuters* (online at 10 October 2018) <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

Farrer, Martin, 'Global Report: Coronavirus Cases Pass 6 Million as Donald Trump Postpones G7', *The Guardian* (online at 31 May 2020) <http://www.theguardian.com/world/2020/may/31/global-report-coronavirus-cases-pass-6-million-as-donald-trump-postpones-g7>

Henriques-Gomes, Luke, 'Robodebt: Government Admits It Will Be Forced to Refund $550m Under Botched Scheme', *The Guardian* (online at 27 March 2020) <https://www.theguardian.com/australia-news/2020/mar/27/robodebt-government-admits-it-will-be-forced-to-refund-550m-under-botched-scheme>

Shields, Bevan, 'Italian Doctors Propose Intensive Care Age Limit to Save Younger Patients', *The Sydney Morning Herald* (online at 12 March 2020) <https://www.smh.com.au/world/europe/italian-doctors-propose-intensive-care-age-limit-to-save-younger-patients-20200312-p5499t.html>

H *Blogs/Websites*

'AI Ethics Principles', *Department of Industry, Science, Energy and Resources* (Web Page) <https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles>

'Clio and ROSS Intelligence Join Forces to Redefine Legal Research', *ROSS Intelligence* (Blog Post, 21 October 2019) <https://blog.rossintelligence.com/post/clio-and-ross-intelligence-join-forces-to-redefine-legal-research>

'Consultation', *Human Rights & Technology* (Web Page) <https://tech.humanrights.gov.au/consultation>

Hao, Karen, ‘Machine Learning Could Check If You’re Social Distancing Properly at Work’, *MIT Technology Review* (Blog Post, 17 April 2020) <https://www.technologyreview.com/2020/04/17/1000092/ai-machine-learning-watches-social-distancing-at-work/>

Martialay, Mary L, ‘Machine Learning Models Predict COVID-19 Impact in Smaller Cities’, *Rensselaer* (Blog Post, 17 April 2020) <https://news.rpi.edu/content/2020/04/17/machine-learning-models-predict-covid-19-impact-smaller-cities?utm_source=miragenews&utm_medium=miragenews&utm_campaign=news>

Singh, Tejaswi and Amit Gulhane, ‘8 Key Military Applications for Artificial Intelligence in 2018’, *Market Research Blog* (Blog Post, 3 October 2018) <https://blog.marketresearch.com/8-key-military-applications-for-artificial-intelligence-in-2018>

Sivasubramanian, Swami, ‘How AI and Machine Learning Are Helping to Fight COVID-19’, *World Economic Forum* (Web Page, 28 May 2020) <https://www.weforum.org/agenda/2020/05/how-ai-and-machine-learning-are-helping-to-fight-covid-19/>