Submission to the Australian Human Rights Commission's Human Rights and Technology Project

Castan Centre for Human Rights Law, Monash University

Authors: Eleanor Jenkin, Hannah Osborne

October 2018

1. Introduction

'What is in any event needed is more interaction among human-rights and AI communities so the future is not created without the human-rights community. (There is no risk it would be created without the AI community.)'[1]

- *Mathias Risse*

1.1. Scope of this submission

The Castan Centre for Human Rights Law welcomes the opportunity to make this submission to the Australian Human Rights Commission's (AHRC) project on human rights and technology. This submission responds to the Issues Paper, released by the AHRC in July 2018. It focuses on exclusively on human rights issues related to artificial intelligence (AI), and in particular on how AI technologies can, and ought to be, regulated (consultation questions 3, 4, 5, 6 and 7).

The term 'AI' is used throughout this submission in its broadest sense. There is considerable variation in how the term AI, and related terms such as machine learning and algorithmic decision-making, are used.[2] We do not believe that a precise definition is necessary to explore the general issues raised in this submission.

This submission comprises six parts. Part 2 of this submission outlines the challenges associated with regulating AI technologies. Part 3 explores the role which human rights can play in this regard, with a particular focus on the benefits and shortcomings of a human rights framework compared with an ethics approach. Part 4 surveys the current regulatory approach to both AI and human rights, and Part 5 examines options and international trends in regulating AI technologies. Finally, Part 6 recommends to the Australian government and the Australian human rights community (including the AHRC) a series of next steps to improve the protection and promotion of human rights in the emerging AI world.

Our recommendations reflect two overarching conclusions. The first is that better regulation is needed if human rights are to be protected and promoted as AI technologies become pervasive. An 'expanded regulatory toolbox' should be employed to achieve flexible, effective regulation, with a particular

---

[1] Mathias Risse, 'Human Rights and Artificial Intelligence: An Urgently Needed Agenda', Discussion Paper, Carr Centre for Human Rights Policy, (May 2018) 30.

[2] See, e.g.: Matthew Scherer, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' (2016) 29(2) *Harvard Journal of Law and Technology,* 354-400*;* Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (2009); Matt Chessen, 'What is Artificial Intelligence? Definitions for Policy-Makers and Non-Technical Enthusiasts', *Medium* (4 Apr. 2017) < https://medium.com/artificial-intelligence-policy-laws-and-ethics/what-is-artificial-intelligence-definitions-for-policy-makers-and-laymen-826fd3e9da3b>.

focus on co-regulatory approaches. Secondly, human rights must move from the periphery to the centre of AI regulation. This will require not only a shift from regulators and the AI community, but also much greater engagement – and considerable work – from the human rights community.

2. Challenges in regulating AI

AI technologies exhibit a number of features which make them devilishly difficult to regulate effectively. Many of these difficulties arise in the regulation of new technologies generally, although they may manifest in particular (and often exaggerated) ways in the case of AI.

The first challenge is one which arises in the regulation of many new technologies – how to regulate sufficiently to minimise social risks and protect human rights, without stifling innovation and progress. This balancing act is important because AI technologies have the potential to contribute in positive ways to the realisation of human rights, and social justice.[3] Any regulation must therefore strike a balance between offering space and flexibility for novel developments while constraining those developments in such a way that social threats are mitigated.

Additional challenges arise from the characteristics of AI itself. These challenges relate less to how *should* we regulate AI, but instead to *how can we* regulate AI? In what ways do the specific features of AI respond to, confound and escape the regulatory tools we have at our disposal? Some of these features are discussed in detail below.

A. The 'Pace Problem'

The pace of innovation in AI has far outstripped the pace of innovation in regulatory tools that might be used to govern it.[4] This 'pacing problem', while not uncommon in the regulation of new technologies generally, is particularly acute with the development of AI, which is moving at blistering speed.

B. Information asymmetry and the 'black box' of AI

Policymakers find themselves at a serious informational and knowledge disadvantage when faced with AI technologies. As Guihot et al note, 'even if

[3] See, e.g.: Sherif Elsayed-Ali, 'Can Technology Help Solve Human Rights Challenges? We Believe it Can' *Amnesty International* (19 Dec. 2016) < https://www.amnesty.org/en/latest/research/2016/12/technology-can-help-solve-human-rights-challenges/>.

[4] Michael Guihot, et al, 'Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence' (2018) 20(2) *Vanderbilt Journal of Entertainment and Technology Law,* 421. See also: Gary Marchant, et al (eds.) *The Growing Gap Between Emerging Technologies And Legal-Ethical Oversight: The Pacing Problem* (2011); Kenneth Abbott, 'Introduction: The Challenges of Oversight for Emerging Technologies', in Kenneth Abbott, et al (eds.) *Innovative Governance Models For Emerging Technologies* (2014) 1–16.

lawmakers are able to obtain technical information from developers, most non-technical folk will still be at a loss to understand a product, let alone predict what impacts it may have on individuals, societies and economies.'[5] This, combined with the pacing problem, can result in what is known as the Collingridge Dilemma; that regulation during the early stages of a technology's development is hampered by a lack of information (in particular about the technology's future impacts), while regulation once a technology has become entrenched faces increased resistance to regulatory change, from users, developers and investors. In other words, 'when change is easy, the need for it cannot be foreseen; when the need for change is apparent, change has become expensive, difficult and time consuming.'[6]

While this problem is common across emerging technologies, it is exacerbated in the case of AI by the inherently opaque nature of algorithmic decision-making technologies. Put simply, it is possible to observe incoming data (input) and outgoing data (output) in algorithmic systems, but their internal operations are poorly understood. This is commonly known as the 'black box' problem. Indeed, in some cases, it is effectively impossible to understand these inner workings, and where the algorithms are proprietary, it may be difficult to gain access to them anyway. This presents a serious challenge to achieving transparency.

C. Definitional quandaries

As mentioned in the introduction, AI itself escapes tight definition. While academics and practitioners in a range of fields have sought to pin down definitions of AI and related concepts (such as machine learning), common definitions accepted across all industries and fields of study remain elusive.[7] This presents a range of challenges to regulation efforts, not least of which is the difficulty of achieving a cohesive regulatory approach across industries and technologies, when key concepts and terminologies are used in different ways.

D. Problems of scope

AI applications reach across industries and the globe. AI technologies are infused into a wide range of sectors, including but not limited to healthcare, transport, policing and the justice system, provision of public services, education, manufacturing, communications, and IT and social media. Any regulatory approach must be capable of being tailored to respond to the contours of a technology and its specific uses, while also maintaining a degree of coherence and logic across environments.

---

[5] Guihot et al, above n 5, 421-2.
[6] David Collingridge, *The Control of Technology* (1980) 11.
[7] Gary Lea, 'Why We Need a Legal Definition of Artificial Intelligence', *The Conversation* (3 Sept. 2015) < https://theconversation.com/why-we-need-a-legal-definition-of-artificial-intelligence-46796>.

The regulation of AI is also complicated by its cross-border nature. The development of an AI technology is often a multi-jurisdictional process; code is easily shared, the companies driving innovation are large multinationals with operations in many countries, and global data flows are unprecedented. Its deployment and impacts AI also defy borders. The extraterritorial reach of any regulatory approach is therefore relevant to its potential effectiveness.[8]

E. Control and liability, or, whose problem is it anyway?

Another feature of AI which creates regulatory challenges is the difficulty in determining who has control, and therefore responsibility, for the impacts of a technology. This is the product of several characteristics of AI. The first is the *discreteness* and *diffusion* of AI development; that is, different components of an AI system may be developed and built separately from each other, by different entities, in different places.[9] It is plausible that no one entity will understand the design or operation of all the components of the final system. The second is that due to the complex (and self-learning) nature of the algorithms, the process by which the AI came to a particular decision is not always clear (even to the developers themselves). This makes it difficult to measure and assign responsibility for harm which may arise.[10] It is therefore unclear how existing torts such as negligence might apply, or indeed how liability generally might be ascertained (let alone distributed).

3. The value (and limits) of a human rights framework for regulating AI

Despite these challenges, there are growing calls for the regulation of AI in order to maximise the social benefits, and minimise the risks, associated with these new technologies.[11] To date, these conversations have been framed largely around the related concepts of 'ethical AI' and 'fairness, accountability and transparency (FAT)'.[12] Human rights norms and organisations have

[8] A significant feature of the GDPR is its 'aspiration to global jurisdiction', although the extent to which this constitutes extraterritorial reach is still being debated (see e.g.: Kurt Wimmer, 'Free Expression and EU Privacy Regulation: Can the New GDPR Reach U.S. Publishers?' (2018) 68 *Syracuse Law Journal,* 547). See also: Paul de Hert and Michal Czerniawski, 'Expanding the European Data Protection Scope Beyond Territory: Article 3 of the General Data Protection Regulation in its Wider Context' (2016) 6(3) *International Data Privacy Law*, 230–243.

[9] Scherer, above n 2, 369-372.

[10] Andrew Tutt, 'An FDA for Algorithms' (2017) 69 *Administrative Law Review,* 105.

[11] See, e.g.: World Economic Forum, *The Global Risks Report 2017* (12th ed.) (2017) 45-46; Elon Musk, CEO of Tesla and SpaceX argues that, 'AI is the rare case where I think we need to be proactive in regulation instead of reactive. Because I think by the time we are reactive in AI regulation, it'll be too late.' (Samuel Gibbs, 'Elon Musk: Regulate AI to Combat 'Existential Threat' Before it's Too Late', *The Guardian* (17 Jul. 2017) < https://www.theguardian.com/technology/2017/jul/17/elon-musk-regulation-ai-combat-existential-threat-tesla-spacex-ceo>).

[12] See e.g.: Corinne Cath, et al, 'Artificial Intelligence and the 'Good Society': the US,

played a surprisingly limited role. It is important, therefore, to consider the potential and limitations of a human rights based approach to regulating AI – what value does it add, and what heavy lifting can human rights do that an 'ethics and AI' framework can't?

3.1. The 'value add' of a human rights framework

Applying a human rights framework to the regulation of AI has several advantages. The first is that it provides a shared language that lowers barriers to entry and engagement, 'which in turn can generate more diverse, creative thinking and enhance both the effectiveness and legitimacy of outcomes.'[13] The 'vernacular' of human rights is familiar to a range of groups and individuals who may lack technical understandings of AI, and offers an entry-point to joining the discussion on the design, use, and regulation of AI.

This vernacular is rooted in human rights norms which have been developed and refined over decades through international legal mechanisms and jurisprudence. While these norms are frequently contested, they offer a reasonably solid, commonly understood normative foundation. Parts 3.4 and 6.2 of the Issues Paper identify some of the human rights most commonly impacted by AI technologies: rights to life and human dignity; the right to privacy; and the right to non-discrimination. However, human rights are relevant to this issue not only because they are affected by AI technologies, but also because the application of a human rights based approach can change the way in which we regulate AI. That is, human rights not only help us understand the problem, but also the solution. The right to an effective remedy, the right to receive information, as well as human rights standards relating to participation, consultation, and transparency, all have the potential to shape the way in which AI is regulated.

The norms and language of human rights also have the benefit of being commonly understood across borders. Lan Xue, Professor and Dean at Tsinghua University's School of Public Policy and Management, describes 'fragmentation of ethics' - the difficulty of aligning moral and ethical values in the global context - as one of the key challenges facing better regulation of AI.[14] As a framework with universal application, human rights is uniquely positioned to respond to the cross-border nature of AI technologies.

EU, and UK Approach' (2018) 24 *Science and Engineering Ethics,* 505-528; ACM Conference on Fairness, Accountability, and Transparency (ACM FAT*) < https://fatconference.org/>; Fairness, Accountability, and Transparency in Machine Learning <http://www.fatml.org/>.

[13] Jason Pielemeier, 'The Advantages and Limitations of Applying the International Human Rights Framework to Artificial Intelligence', *Points* (6 Jun. 2018) < https://points.datasociety.net/the-advantages-and-limitations-of-applying-the-international-human-rights-framework-to-artificial-291a2dfe1d8a>.

[14] Comments at the AI for Good Global Summit, (7-9 Jun. 2017, Geneva, Switzerland), reported in 'Scientists and Stakeholders in Geneva for Good Artificial Intelligence', *Synced* (13 Jun. 2017) <https://medium.com/syncedreview/scientists-and-stakeholders-in-geneva-for-good-artificial-intelligence-5e09e7dcafa9>.

The international human rights regime also offers an institutional architecture which can assist rights-holders and their representatives to compel compliance with human rights standards by states and others. As van Veen describes:

> Today, there is a global network of United Nations human rights bodies, human rights NGOs, human rights defenders, courts and national human rights institutions that provide spaces in which human rights disputes caused by the development and use of AI systems can be aired and addressed constructively, ensuring violators are held to account. These human rights bodies, procedures, and institutions are more responsive than is often believed.[15]

This institutional architecture is supplemented by a wide range of tools and processes designed to assist duty-bearers to operationalise human rights standards. In recent years a particular - and in the case of AI applications, potentially useful - set of tools has been developed to assist business entities wishing to operate in accordance with the responsibilities set out in the UN's Guiding Principles on Business and Human Rights.[16] This includes tools such as human rights impact assessments and guidance for compliance audits,[17] guidance on effective consultation with affected communities,[18] and more controversially, operational-level grievance mechanisms.[19]

Critically, human rights law and language provide a means to surface power asymmetries, and to challenge them. Many of the risks relating to AI concern the entrenchment of the disadvantage experienced by marginalised and vulnerable groups. A human rights analysis requires the identification of duty bearers, and empowers rights holders with helpful analytical, normative and institutional tools to hold them to account. Even when formal institutional

---

[15] Christiaan van Veen, 'Artifical Intelligence: What's Human Rights Got to Do With It?', *Points* (14 May 2018) < https://points.datasociety.net/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d5>. See also Pierlemeier, above n 16.

[16] Special Representative of the Secretary-General on the Issue of Human Rights and Transnational and other Business Enterprises, *Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework*, Human Rights Council, UN Doc. A/HRC/17/31 (21 Mar. 2011).

[17] See: James Harrison, 'Human Rights Measurement: Reflections on the Current Practice and Future Potential of Human Rights Impact Assessment' (2011) 3(2) *Journal of Human Rights Practice,* 162-187. See also: Danish Institute of Human Rights, 'Human Rights Impact Assessment Guidance and Toolbox' < https://www.humanrights.dk/business/tools/human-rights-impact-assessment-guidance-and-toolbox>; Shift, *From Audit to Innovation: Advancing Human Rights in Global Supply Chains* (2013).

[18] See e.g.: Oxfam America, *Community Voice in Human Rights Impact Assessments* (2015).

[19] Sarah Knuckey and Eleanor Jenkin, 'Company-Created Remedy Mechanisms for Serious Human Rights Abuses: A Promising New Frontier for the Right to Remedy?' (2015) 19 *International Journal of Human Rights,* 801-827; Emma Wilson and Emma Blackmore, *Dispute or Dialogue?: Community Perspectives on Company-led Grievance Mechanisms* (2013).

processes are unable to achieve a positive outcome for rights holders, framing grievances as rights violations can give them weight which influences the behaviour of more powerful actors; 'human rights, as a language and legal framework, is itself a source of power because human rights carry significant moral legitimacy and the reputational cost of being perceived as a human rights violator can be very high.'[20]

These advantages can be contrasted with the 'ethics' paradigm which currently dominates discussions of AI. Key concepts such as 'ethical' or 'good' corporate behaviour, and 'unfair' actions, remain slippery and ill-defined.[21] Moreover, the process for defining such terms – and most importantly, *who* defines them – is haphazard. Consequently, it is corporations engaged in AI development which are currently leading efforts to define ethical approaches to AI.[22] While the initiative of industry is to be applauded, these efforts are insufficiently inclusive, participatory and representative. Many commentators – including from within industry – have called for national and international discussions about the social risks of AI.[23] Industry lacks the institutional structures, normative frameworks, and legitimacy, to act as the locus for this discussion. Human rights institutions – at the international level, and in Australia – are uniquely positioned to perform this role.

3.2. The (not insurmountable) limits of a human rights based framework

Despite these advantages, there are limits to how heavily we can rely on human rights norms and frameworks to constrain the risks presented by AI. Human rights norms develop slowly, and human rights mechanisms sometimes move at a similarly crawling pace. This presents significant problems in light of the 'pace problem' discussed in section .2

Another limitation facing human rights in addressing the consequences of AI is the weakness of the obligations placed on businesses. Human rights law is traditionally concerned with the actions of states. As the power of corporations has increased – along with the human rights impacts of their operations – there have been efforts to bring them within the ambit of human rights law. The most significant step is this regard was the adoption of the Guiding Principles on Business and Human Rights, which posit that business enterprises have a responsibility to respect human rights, meaning that they should avoid infringing on the human rights of others and should address

[20] Van Veen, above n 15.
[21] AI Now, a leading research center on the social implications of AI has gone so far as to declare that 'current framings of AI ethics are failing…' (Alex Campolo, et al, *AI Now 2017 Report,* AI Now (2017) 34).
[22] For examples of corporate initiatives, see below n 58.
[23] Toby Walsh, 'Elon Musk is Right: We Need to Talk About Artificial Intelligence', *The Conversation* (30 Oct. 2014) < https://theconversation.com/elon-musk-is-right-we-need-to-talk-about-artificial-intelligence-33577>; Urvashi Aneja 'What We Need to Talk About When We Talk About Artificial Intelligence' *Digital Policy Portal* (7 Mar. 2017) < http://www.digitalpolicy.org/need-talk-talk-artificial-intelligence/>.

adverse human rights impacts with which they are involved.[24] The principles which operationalise this responsibility – in particular the requirement to make a policy commitment, and to undertake human rights due diligence – have been taken up widely by businesses.

However, the responsibilities placed on businesses – compared to those which apply to states – are modest. The scope of business responsibilities is much narrower than the obligations placed on states. The weight of the responsibilities placed on businesses is also much lighter. While states are subject to enforceable obligations under international law ('musts'), the responsibility of businesses to respect human rights as articulated in the Guiding Principles is only a 'should'. This means that compliance with this responsibility, and adoption of those measures which operationalise it, are entirely voluntary. This presents a major challenge in the context of AI technologies, which are primarily developed and deployed by (often large, wealthy and powerful) corporate entities.

At the normative level, some have queried whether human rights has the conceptual capacity to adequately account for AI.[25] This is critical if human rights is to form a comprehensive framework for shaping AI. While human rights law provides a firm foundation, it is not yet clear what these norms have to say about the specific conditions created by the development and use of AI technologies. There is a tendency in general discourse - and even within the human rights field - for the term 'human rights' to be used loosely, similar to the way 'ethics' is often used. For example, President of the Electronic Privacy Information Center (EPIC), Marc Rotenberg, is quoted on EPIC's website as stating, "At the intersection of law and technology - knowledge of the algorithm is a fundamental human right."[26] Without any supporting evidence or analysis, the statement is at best a stretch, at worst a blatant mischaracterisation. As noted by the Electronic Frontier Foundation, it is not at all settled — at least in terms of international agreements and similar law — how many key international law and human rights principles should be applied to various AI technologies and applications.[27]

Examples of issues requiring clarity range from the narrow to the deeply fundamental. At the narrow end, the inference of personal information using non-sensitive data and the use of big data profiling to sort, score, categorise, assess and rank individuals present difficulties for prevailing understandings of the human right to privacy and discrimination.

---

[24] Guiding Principles on Business and Human Rights, principles 11-15.
[25] Helmut Aust, ''The System Only Dreams in Total Darkness': The Future of Human Rights Law in the Light of Algorithmic Authority' *German Yearbook of International Law* (forthcoming).
[26] Electronic Privacy Information Centre, 'Algorithmic Transparency: End Secret Profiling' <https://www.epic.org/algorithmic-transparency/>.
[27] Peter Eckersley, 'How Good Are Google's New AI Ethics Principles?', *Electronic Frontier Foundation* (7 Jun. 2018) <https://www.eff.org/deeplinks/2018/06/how-good-are-googles-new-ai-ethics-principles>.

At the more fundamental end, human rights are premised on (among other things) the notions of human agency and autonomy, which manifest in a heavy reliance on transparency and informed consent. However, AI technologies undermine these ideas in a range of ways: AI nudges our behaviour in ways we cannot be aware of; it influences the information we are presented with and shapes our reality; and its 'black box' and 'explainability' problems make transparency difficult, and potentially less useful.[28] It is unclear at present whether existing human rights concepts can be adapted to meet these challenges (such as through a 'right to explanation') or whether a more radical rethink or extension of human rights standards is required. Another 'big picture' conceptual issue which will need to be considered is what, if anything, human rights has to say about 'surveillance capitalism'. Surveillance capitalism refers to the 'monetization of data captured through monitoring people's movements and behaviours.'[29] What implications, if any, might it have for human dignity, for example?

While significant, these normative gaps are not insurmountable. Work has begun (if belatedly) on exploring the particular ways human rights norms might map on to AI applications. Civil society is leading the way, with Amnesty International and AccessNow launching the Toronto Declaration on protecting the rights to equality and non-discrimination in machine learning systems on 16 May 2018.[30] The Declaration makes some tentative, but important, steps in articulating actions which duty-bearers might take to protect human rights from harms caused by machine learning applications. As of September, 32 non-state groups and individuals have signed on to the Declaration.[31] At the UN level, the Special Rapporteur on the right to privacy has instigated a Taskforce on Big Data - Open Data (see case study below). While these initiatives are encouraging, a tremendous amount of work remains to be done in exploring, shaping and advancing human rights standards to meet the challenge of AI technologies. It is imperative that the human rights community takes up this work as a matter of urgency, or human rights may be left out of future AI regulation entirely.

---

[28] See e.g.: Mike Ananny and Kate Crawford, 'Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability' (2018) 20(3) *New Media & Society,* 973-989; Sandra Wachter, et al, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR (2018) 31(2) *Harvard Journal of Law and Technology*, 842-887; Sandra Wachter, et al, 'Transparent, Explainable, and Accountable AI for Robotics' (2017) 2(6) *Science Robotics* 1-2; Finale Doshi-Velez and Mason Kortz, 'Accountability of AI Under the Law: The Role of Explanation', Berkman Klein Center for Internet & Society, Working Paper (2017).

[29] Shoshana Zuboff, 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization' (2015) 30(1) *Journal of Information Technology,* 75). See also: Shoshanna Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (forthcoming 2019).

[30] Amnesty International and Access Now, 'Toronto Declaration: Protecting the Right to Equality and Non-discrimination in Machine Learning Systems' (2018).

[31] This comprises 21 civil society organisations, three academic institutions, two technology sector companies, and six individuals. (correspondence with authors).

Case study: United Nations Special Rapporteur on the right to privacy's Taskforce on Big Data - Open Data

The current (and first) UN Special Rapporteur on the right to privacy, Prof. Joseph Cannataci, has convened a taskforce to consider the right to privacy in the context of big data and open data. The taskforce is led by David Watts, Adjunct Professor of Law at Latrobe University and at Deakin University, and previously Victorian Commissioner for Privacy and Data Protection. The taskforce has produced an initial report (presented to the Human Rights Council as part of the Special Rapporteur's 2017 report to the UN General Assembly[32]) in which it mapped key issues. The taskforce and Special Rapporteur are consulting widely in 2018, and intend to release a final report in or after 2018. This initiative demonstrates the growing engagement of the human rights community with the human rights issues surrounding new technologies, and represents a timely opportunity to begin articulating the precise ways in which human rights standards apply (or should apply) to the development and use of these technologies.

[32] Joseph Cannataci, Report of the Special Rapporteur on the right to privacy, UN Doc A/72/43103 (19 Oct. 2017).

4. Current regulatory landscape

4.1. Protections for human rights

The Issues Paper sets out some of the ways in which AI can impact human rights, and the ways in which human rights are protected in Australia, and we will refrain from restating these. Nonetheless, we consider it important to note the inadequacy of Australia's protections for human rights. While some of the key human rights affected by AI (such as privacy, and the right to non-discrimination) are protected to varying degrees under state or federal legislation (and by human rights legislation in Victoria and the ACT),[33] without full incorporation of international human rights obligations into domestic law protection remains patchy and contingent.

When considering how the harmful human right impacts of AI technologies might be mitigated, the emphasis will naturally fall on identifying appropriate forms of technological regulation, and the incorporation of human rights norms and practices into these frameworks. This is necessary and proper, in particular given the lead role played by the private sector in the development and deployment of AI technologies. However, it remains the case that one of the best ways to increase the realisation of human rights in the context of emerging technologies is to enhance protections for human rights generally. We submit that this is best achieved through the enactment of a comprehensive, judicially enforceable federal Human Rights Act. Such an Act would not only ensure actions and decision by government authorities were consistent with human rights, but would also contribute to a stronger human rights culture in Australia, and to the refinement (through application and judicial consideration) of human rights norms relevant to AI technologies.

4.2. Laws relating to AI

There is no specific regulatory framework relating to artificial intelligence in Australia (or indeed, anywhere else that we are aware of). Various harms arising from the technology may be captured under a patchwork of existing avenues for legal recourse. For example, actions in tort may be lie in some situations (such as manufacturer negligence), as well as existing protections found in consumer and privacy laws. However, it is unclear how these laws might apply to rapidly developing areas such as AI. Two key examples are offered below:

A. Australian Consumer Law

If AI is used in a technology which falls within the scope of the Australian Consumer Law (that is, for personal, domestic or household

[33] *Charter of Human Rights and Responsibilities Act 2006* (Vic); *Human Rights Act 2004* (ACT).

use, or less than $40,000),[34] then existing protections may apply. For example, consumers would be able to rely upon manufacturer liability for product defects. Where difficulty arises is if the technology is not 'ordinarily acquired' for personal or domestic use - which in early stages of commercialisation may be difficult to establish. This is especially relevant for technology expected to ordinarily exceed the $40,000 threshold, such as self-driving vehicles - thus precluding access to these protections.

B. Privacy Act

The *Privacy Act 1988* (Cth) provides further protection for individuals, if an organisation using AI applications handles 'personal information'.[35] The Privacy Act applies to the collection, use and disclosure of information about identified or identifiable individuals and requires compliance with the Australian Privacy Principles (APP). Relevantly, for algorithmic decision making by AI, the APP require transparency about an organisation's information handling practices and provide for access and correction rights.

However, the Privacy Act does not apply if the relevant data are not personal information, for example if data is de-identified. However, the unprecedented processing capability of Big Data and AI increases the possibility of re-identification of data that was stripped of personal attributes. Data traditionally regarded as 'de-identified' may therefore be inappropriately classified as non-personal information in the context of AI, and therefore outside the scope of protection. Further, companies may be unwilling to reveal the algorithms making use of this data (as protected trade secrets), thus restricting individuals' ability to review (and challenge) such decisions. Moreover, privacy rights are not directly enforceable in court, instead individuals must make a complaint to the Privacy Commissioner.

One of the primary disadvantages of the current legislative approach more generally is the lack of underlying explicit human rights protection. Where human rights are protected, this occurs inconsistently, and sometimes insufficiently. For example while the Privacy Act provides protection for the human right to privacy, its protections fall short of the requirements in international human rights law – the scope of the protection is narrower and is subject to exceptions and exclusions which arguably exceed those permitted under the International Covenant on Civil and Political Rights. Even where particular rights are presently protected, it is unclear whether these protections will be sufficiently flexible to cover future AI applications.

[34] *Competition and Consumer Act 2010* (Cth) Sch 2, s 3.
[35] *Privacy Act 1988,* s 6(1). See also: Office of the Australian Information Commissioner, *Guide to Data Analytics and the Australian Privacy Principles* (2018).

4.3. International comparison

We are not aware of any jurisdiction adopting laws to regulate AI generally. Perhaps the EU has come closest with the adoption of its General Data Protection Regulation ('GDPR'), which is discussed in the case study below. However, it is clear that many countries are turning their regulatory minds to the opportunities and risks presented by AI technologies. This is demonstrated by the burgeoning in the last two years of national AI strategies.[36] These often entail the creation or support of independent bodies to oversee the development of AI, with an explicit mandate regarding ethics or human rights. These strategies and bodies are often conceptualised as the 'first step' in defining a comprehensive approach to AI technologies, and tend to prioritise bringing together experts with diverse, specialised knowledge.

The role played by ethics varies across the different national strategies. For example, Canada has established a dedicated AI and Society program as a core part of the Pan-Canadian Artificial Intelligence Strategy.[37] Canada has placed particular emphasis on social responsibility, accessibility and inclusiveness within the development of AI, implemented throughout a range of public workshops, summer schools and university research funding.[38] Conversely, many strategies are concerned solely with positioning their country at the forefront of the 'AI revolution'. Japan's Artificial Intelligence Technology Strategy, which was released in March 2017, focuses exclusively on driving the development and use of Japanese AI technologies. Similarly, there was a notable shift away from ethics towards innovation from the US' 2016 AI Strategy report[39] to the President's approach to AI in 2018 (with a focus on deregulation rather than developing new ethical guidelines).[40]

[36] For a helpful overview of these strategies, see: Tim Dutton, 'Artificial Intelligence Strategies', *Medium* (29 Jun. 2018) <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>.
[37] Canadian Institute for Advanced Research, 'Pan-Canadian Artificial Intelligence Strategy' <https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy>.
[38] See: Canadian Institute for Advanced Research, 'CIFAR Congratulates Prime Minister Trudeau and President Macron on Historic Commitment to Create International Study Group on Inclusive and Ethical AI', *CISION* (7 Jun. 2018) < https://www.newswire.ca/news-releases/cifar-congratulates-prime-minister-trudeau-and-president-macron-on-historic-commitment-to-create-international-study-group-on-inclusive-and-ethical-ai-684842541.html>.
[39] National Science and Technology Council, *The National Artificial Intelligence Research and Development Strategic Plan* (2016).
[40] White House Office of Science and Technology Policy, *Summary of the 2018 White House Summit on Artificial Intelligence* (2018).

Case study: EU Data Protection

While the GDPR is not specifically concerned with AI, its article 22 establishes protections against algorithmic/automated decision-making - which forms the basis of many applications of AI. This provision creates a limited right to object to automatic decision-making and requires data controllers (such as AI developers handling personal data) to implement 'suitable measures' to safeguard the rights, freedoms and legitimate interests of individuals, thus including those contained in EU existing human rights legislation. This provides both a clear human rights-based protection in the context of some AI applications, and a specific avenue for recourse for affected individuals.

4.4. Advisory Bodies

The types of AI advisory bodies which are emerging tend to be independent, and comprise experts across a variety of fields. At this stage, these tend to be purely advisory in nature, as opposed to possessing any regulatory functions. However, the impact of such groups should not be underestimated - the EU High Level Expert Group was instrumental in the development of the EU Communication on AI, and advises on the implementation of EU Charter of Fundamental Rights in the context of AI.[41]

A number of these bodies have been specifically tasked with developing ethical guidelines for AI, or with providing advice to government on doing so. For example, the EU's High-Level Group on Artificial Intelligence is preparing draft ethics guidelines for member states to consider,[42] and Singapore has recently announced a new Advisory Council on the Ethical Use of AI and Data to help the government develop standards and governance frameworks for the ethics of AI.[43]

[41] European Commission, 'High-Level Expert Group on Artificial Intelligence' <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>.
[42] Ibid.
[43] Infocom Media Development Authority, 'Composition of the Advisory Council on the Ethical Use of Artficial Intelligence ("AI") and Data' (30 Aug. 2018) < https://www.imda.gov.sg/about/newsroom/media-releases/2018/composition-of-the-advisory-council-on-the-ethical-use-of-ai-and-data>.

5. Trends and options for regulating the AI sector

In this section we explore a range of options for regulating AI technologies to minimise negative impacts on human rights. We draw on academic research, experiences from other countries, and regulatory approaches in other sectors in Australia.

5.1. Do nothing

According to the Australian Government Best Practice Regulation Handbook the first step when considering policy options is to question whether existing regulation is sufficient.[44] We submit that in the case of AI in Australia, it is not. The first reason for this is the patchy and inconsistent regulation of both AI technologies, and of human rights protections (see section .(4The second reason is that this patchwork of regulation does not provide sufficient certainty to industry or the public .Because of their(primarily) rule-based nature, and the fact that AI applications were generally not envisaged in their drafting, xisting laws lack the flexibility to respond as technology develope. This may create uncertainty about how a legal requirement may be adhered to or enforced. If this uncertainty is not addressed, the applicability of the current legislative framework will increasingly rely upon judicial interpretation. This carries with it the risk of inconsistency, as well as being informed by less specialised knowledge.[45] Therefore a central response to emerging AI technology, which may be adapted at a sector specific level as necessary, will provide greater clarity and consistency.

5.2. Direct regulation of AI by the State

We are not aware of any country or jurisdiction which has yet passed legislation regulating AI technologies generally. However, the question of whether governments ought to - and what any such legislation should look like - dominates discussion of the regulation of AI. A closely-related inquiry is what the powers, functions and structure of any government regulator ought to be.

We believe that there may be cases where clear rules are warranted. For example, an explicit prohibition on the development of lethal autonomous weapon systems would be appropriate. However, a purely 'bright' line' or 'complex or detailed rule' approach to legal regulation would be inappropriate. The inflexibility of these approaches is well-recognised, as is the difficulty in

[44] Australian Government, *Best Practice Regulation Handbook* (2010).
[45] Hon. Michael Kirby, 'The Fundamental Problem of Regulating Technology' (Comments at the Conference on the Ethical Governance of Information and Communication Technology and the Role of Professional Bodies, held 1 May 2008 in Canberra, Australia) <http://www.hcourt.gov.au/assets/publications/speeches/former-justices/kirbyj/kirbyj_1may08.pdf>.

applying them to fast moving new technologies;[46] we have found no serious proposals for a predominantly rule-based legislative response.

Instead, as noted in the Issues Paper, principles-based regulation may offer a way forward. While the concept of principles-based regulation is expansive, in general terms it means moving away from reliance on detailed, prescriptive rules and relying more on high-level, broadly stated rules or principles to set the standards by which regulated entities must conduct themselves.[47] Principles-based approaches have been widely applied in the regulation of privacy and data protection, including in Australia. The *Privacy Act 1988* (Cth), and the Australian Privacy Principles, are an example of this approach. The advantages and drawbacks of principles-based regulation have been thoroughly explored.[48] The main benefit is the flexibility of the approach; a standard can be applied and interpreted in light of new technological and social developments, and will therefore stand the test of time. The main drawbacks are that this flexibility creates uncertainty and unpredictability for regulated entities, and enforcement can be costly.

The problem of uncertainty is often addressed through the establishment of a strong regulator, which possesses both standard-setting and enforcement powers. The regulator will often be empowered to issue guidance on how the standards are to be applied, and to monitor and compel compliance. For example, the Office of the Australian Information Commissioner (OAIC) is granted under the Privacy Act various monitoring (including the power to conduct or compel the conduct of privacy impact assessments), investigative (such as conducting investigations, and resolving consumer complaints), and enforcement powers.

Commentators have begun to consider what an AI regulator ought to look like. In the US, Tutt has called for an 'FDA [Food and Drug Administration] for AI' which would have a role in approving AI applications before they go to market.[49] Scherer has proposed something akin to FDA-lite; an agency which would exercise standard-setting and certification functions, but would not have the pre-market approval role of the FDA. The agency would instead attach limited liability to its certification.[50] While these example are based on the US model of regulatory agencies (which differs from the Australian), they nonetheless demonstrate the central questions in regulator design – just how firm a hand should the regulator be given, and it what ways should it be allowed to wield it.

---

[46] See e.g.: Ruth Carter and Gary Marchant, 'Principles-based Regulation and Emerging Technology', in Gary Marchant, et al (eds.) *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (2011).

[47] Julia Black, et al, 'Making a Success of Principles-based Regulation' (2007) 1(3) *Law and Financial Markets Review,* 191.

[48] See, e.g.: Australian Communications and Media Authority, *Optimal Conditions for Effective Self- and Co-regulatory Arrangements*, Occasional Paper (Jun. 2015); Julia Black, 'The Rise (and Fall?) of Principles Based Regulation', in Kern Alexander and Niamh Moloney (eds.) *Law Reform and Financial Markets* (2011).

[49] Tutt, above n 10.

[50] Scherer, above n 2.

Case Study: Playing in the regulatory sandbox

As governments seek to balance the rewards of private-sector innovation with the need to protect the public through rigorous regulation, some have turned to 'regulatory sandboxes'. A regulatory sandbox is an experimental space, in which companies are allowed to test innovative products under relaxed regulatory conditions (such as the waiver of certain rules), but under close supervision of the regulator. Regulatory sandboxes are common in the fintech industry,[51] and Australia's regulator ASIC has recently jumped on the trend, launching its own fintech sandbox.[52] AI-related financial products already feature heavily in fintech regulatory sandboxes,[53] and a number of commentators have suggested that they might play a similar role in encouraging innovation in a regulated AI sector.[54]

[51] For example, the UK's Financial Conduct Authority runs a regulatory sandbox (https://www.fca.org.uk/firms/regulatory-sandbox) as does the Monetary Authority of Singapore (http://www.mas.gov.sg/Singapore-Financial-Centre/Smart-Financial-Centre/FinTech-Regulatory-Sandbox.aspx).

[52] See: ASIC, 'Regulatory Sandbox' <https://asic.gov.au/for-business/your-business/innovation-hub/regulatory-sandbox/>.

[53] See, e.g.: PR Newswire, 'Blockchain and Artificial Intelligence Innovation Dominate Theme of Recent FCA Regulatory Sandbox New Successful Companies' (14 Aug. 2018) <https://www.prnewswire.com/news-releases/blockchain-and-artificial-intelligence-innovation-dominate-theme-of-recent-fca-regulatory-sandbox-new-successful-companies-300696718.html>.

[54] William Eggers and Mike Turley*, The Future of Regulation: Principles for Regulating Emerging Technologies*, Deloitte Center for Government Insights (2018). See also: Wolf-Georg Ring and Christopher Ruof, 'A Regulatory Sandbox for Robo Advice' *European Banking Institute,* Working Paper Series n. 26 (2018).

5.2.1. Defining the principles in principles-based regulation

We submit that in any future principles-based legislative regime, both the principles and the process for their elaboration should reflect a human rights-based approach. In practice, this means that the principles should be developed through a process which:

- *is participatory*: the public – and particularly those whose rights are likely to be most affected by AI technologies – should have the opportunity to participate in the process. Participation must be meaningful, that is, it should feed into the final principles, and not merely be cosmetic.
- *is not discriminatory*: the participation of people from historically marginalised groups should be invited and facilitated, including people with disability, indigenous Australians, and people from culturally and linguistically diverse backgrounds. In the Australian context, this might also include older people, people living in regional and rural Australia, and people experiencing economic disadvantage. Appropriate supports should be put in place to facilitate the free and meaningful participation of these groups.
- *is transparent*: information on the process (including information on how to participate) should be made publicly available, along with information about how decisions have been reached.
- *integrates human rights law and standards*: international human rights law must underpin the process and its outcomes.

More substantively, the principles themselves should reflect human rights standards and should take into account:

- the importance of identifying and managing risks to human rights throughout the lifecycle of a technology or product, including through the use of human rights impact assessments, live testing and audits;
- enhanced transparency and accountability: for example through the formulation of a 'right to explanation', and full disclosure to the public when an AI system is being used (particularly for decision-making); and
- the right to remedy: this should include clear and accessible options for review of algorithmic decisions. The EU's General Data Protection Regulation (GDPR),[55] which entrenches a right to (human) review for algorithmic decision-making,[56] presents a possible model for this.

Case Study: the 'right to explanation'

A key normative development in the EU's GDPR is the emergence of a novel 'right to explanation'. Under articles 13-15 of the regulation, in certain cases of automated decision-making the data controller will provide, 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.'

[55] (EU) 2016/679.
[56] GDPR, art. 22.

The scope, meaning and consequences of these provisions have been hotly debated.[57] Whether the GDPR provisions constitute a true 'right to explanation', and whether they will in practice increase transparency and empower consumers, remains to be seen. At a minimum, the GDPR provisions represent an example of the sort of regulatory creativity and innovation needed to meet the challenge of emerging technologies.

5.3. Self-regulation

As governments around the world grapple with how to best regulate AI (if at all), industry is taking steps to develop its own ethical guidance and policy. At present, these efforts generally fall into three categories: self-regulation by individual companies; the elaboration of professional standards; and multi-stakeholder initiatives. Self-regulation by industry has an important role to play in an expanded regulatory toolset, and provides an important avenue for the integration of human rights standards and processes into AI development and deployment. There are, however, limits to the efficacy of voluntary regulation - some of which are general to all self-regulation, and some of which are specific to AI - which mean that this approach should be accompanied by government-led measures.

5.3.1. Organisational self-regulation

In very recent years, some AI industry leaders have developed initiatives to align their AI activities with ethical standards.[58] These initiatives have generally involved the elaboration of principles which the company states will guide their work on AI. These principles tend to be fairly general, and based on the notion of maximising benefit and minimising social harms. For example, OpenAI's Charter states that, 'We commit to use any influence we obtain over [artificial general intelligence's (AGI)] deployment to ensure it is used for the benefit of all, and to avoid enabling uses of AI or AGI that harm humanity or unduly concentrate power.'[59] In a similar formulation, DeepMind's Ethics and Society Principles states that 'We believe AI should be developed in ways that serve the global social and environmental good, helping to build fairer and more equal societies. Our research will focus directly on ways in

[57] See e.g.: Sandra Wachter, et al, 'Why a Right to Explanation of Automated Decision-making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) *International Data Privacy Law*, 76–99; Andrew Selbst and Julia Powles ‚Meaningful information and the Right to Explanation' (2017) 7(4) *International Data Privacy Law,* 233-242.

[58] See e.g.: Deepmind, 'DeepMind Ethics & Society Principles' < https://deepmind.com/applied/deepmind-ethics-society/principles/>; Microsoft, Microsoft AI Principles' < https://www.microsoft.com/en-us/ai/our-approach-to-ai >; Google, 'AI at Google: Our Principles' < https://www.blog.google/technology/ai/ai-principles/>.

[59] OpenAI, 'OpenAI Charter' <https://blog.openai.com/openai-charter/>.

which AI can be used to improve people's lives, placing their rights and well-being at its very heart.'[60]

The principles promulgated by these corporations exhibit a number of shortcomings. The first is that they rely on vague, poorly defined concepts such as 'fairness', 'social good' and 'social harm'. These terms do not have clear, commonly-accepted meanings, and brush over many thorny questions (such as, who determines whether an outcome represents a social good or a harm? What if a technology has applications which may be beneficial as well as harmful? How will the corporation weigh potential benefits to one group, and harms to another?)

Secondly, human rights tend to be left out entirely. The exception is Google's AI principles, released in June 2018. In addition to more broad, ethics-based statements, Google has outlined the AI applications it will *not* pursue, including 'technologies whose purpose contravenes widely accepted principles of international law and human rights.'[61] While Google's principles have generally been received as a positive step forward, a number of commentators have expressed concerns that the principles remain overly vague, fail to incorporate a human rights-based approach throughout and do not capture many of the potential human rights impacts of AI.[62] Concerns have also been raised 'that by relying on "widely accepted principles of international law and human rights" for the purposes that Google will not pursue, the company is potentially sidestepping some harder questions.[63]

Lastly, none of these initiatives has, to date, included any independent, transparent process to ensure the principles are being applied. This includes an absence of information on how broad principles will be operationalised, and their application monitored.[64] Without any accountability mechanisms,

---

[60] DeepMind, 'DeepMind Ethics and Society Principles' <https://deepmind.com/applied/deepmind-ethics-society/principles/>.

[61] Google, 'AI at Google: Our Principles' < https://www.blog.google/technology/ai/ai-principles/>.

[62]Lorna McGregor and Vivian Ng, 'Google's New Principles on AI Need to be Better at Protecting Human Rights', *The Conversation* <https://theconversation.com/googles-new-principles-on-ai-need-to-be-better-at-protecting-human-rights-98035>; Article 19, 'Google: New Guiding Principles on AI show Progress but Still Fall Short on Human Rights Protections' <https://www.article19.org/resources/google-new-guiding-principles-on-ai-show-progress-but-still-fall-short-on-human-rights-protections/>.

[63] Peter Eckersley, 'How Good Are Google's New AI Ethics Principles?', *Electronic Frontier Foundation* <https://www.eff.org/deeplinks/2018/06/how-good-are-googles-new-ai-ethics-principles>.

[64] A good example of this lack of transparency is DeepMind. When DeepMind was acquired by Google in 2014, Google agreed as part of the acquisition to set up an ethics and safety board. While DeepMind leadership have insisted that the Board has been convened and is operational, it refuses to disclose who is on the board, what it discusses, or publicly confirm whether or not it has even officially met (Alex Hearn, 'Whatever Happened to the DeepMind AI Ethics Board Google Promised?', *The Guardian* (27 Jan. 2017)

and without real transparency, it is impossible to assess the effectiveness of these regulatory measures. This of course raises more fundamental questions about the legitimacy and efficacy of self-regulation *as a form of regulation.* The legitimacy and effectiveness of self-regulation to achieve public goals are highly contested.[65] Critics argue that self-regulation can be little more than companies 'marking their own homework'[66] or 'window dressing'[67]. Of particular concern are self-regulation initiatives which are primarily cosmetic, and do not meaningfully shape the corporations behaviour.

It is also arguable that the very nature of AI technologies requires a more 'married up' approach. A key difficulty associated with AI applications is their diffuse development, meaning that different components of an application may be developed by entirely different people, in different places.[68] There can also be a disconnect between the development of a technology, and its final application. A technology developed with a particular purpose in mind may be applied in entirely unanticipated ways (in turn leading to unanticipated impacts). Consequently, the creation and adoption by some companies of certain ethical principles in their AI work is unlikely to produce - on its own - the regulatory coherence and rigour necessary to prevent (or respond to) negative impacts on human rights.

Nonetheless, a human rights framework provides tools which may be helpful in filling some of these gaps. Under the Guiding Principles on Business and Human Rights, a business should have a human rights due diligence process to identify, prevent, mitigate and account for how they address their impacts on human rights.[69] Private, not-for-profit and academic actors have translated this principle into practical tools for businesses, including human rights impact assessments, social impact audits, and human rights reporting. While these tools are far from perfect,[70] they provide a human rights-based framework which can be adapted and applied to AI technologies.[71] The widespread adoption of these approaches by companies working on AI may help them to translate their ethical principles into action, and may also lead to a more coherent approach across key actors in the field.

<https://www.theguardian.com/technology/2017/jan/26/google-deepmind-ai-ethics-board>).

[65] See: Jodi Short, 'Self-regulation in the Regulatory Void: "Blue Moon" or "Bad Moon"? (2013) 649, *Annals of the American Academy of Political and Social Science,* 22-34.

[66] Frances Bowen, 'Marking Their Own Homework: The Pragmatic and Moral Legitimacy of Industry Self-Regulation' (2017) *Journal of Business Ethics,* 1-16.

[67] Short*,* above n 65, 24.

[68] Scherer, above n 2.

[69] Guiding Principles on Business and Human Rights, principle 15(b).

[70]Caroline Rees, 'The Way Businesses' Social Performance Gets Measured Isn't Working', *Shift* (2018).

[71] AI Now has recently published guidance on what it calls 'algorithmic impact assessments'. While the guidance is designed for public agencies, it may also be relevant to the private sector (Dillon Reisman, et al, 'Algorithmic Impact Assessments: A Practical Framework For Public Agency Accountability', *AI Now* (April 2018) <https://ainowinstitute.org/aiareport2018.pdf>).

Case Study: Microsoft Salient Human Rights Issues Report – FY17

Since 2016, Microsoft has published an annual Salient Human Rights Issues Report, based on the UN Guiding Principles Reporting Framework.[72] In the FY17 report, the company stated:
'We began a major, forward looking Human Rights Impact Assessment (HRIA) at the start of FY17 into Microsoft's growing portfolio and expertise in artificial intelligence (AI). The HRIA broadly considers AI technology in order to:

- Identify potential risks related to the research and development (R&D) and sales of AI products and services;
- Contribute to Microsoft's continuing efforts to meet its responsibility to respect human rights through its products, services and business activities and relationships;
- Inform the public debate about benefits and risks of AI and effective policy recommendations;
- Position the responsible use of AI as a technology in the service of human rights.[73]

The HRIA is due for completion in 2018. Previous impact assessments have not been made publicly available, so it is likely that transparency will remain a problem. However, the application of the Guiding Principles - and tools which operationalise them - to the development of AI technologies represents an important way in which human rights norms can influence AI applications.

5.3.2. Industry-level self-regulation

Another form of self-regulation is currently emerging at the industry level in the form of professional standards. The Institute of Electrical and Electronics Engineers (IEEE), which describes itself as 'the world's largest technical professional organization for the advancement of technology', is a leading propounder of voluntary professional standards for the industry. The IEEE has established the Global Initiative on Ethics of Autonomous and Intelligent Systems. The Executive Committee of the Initiative includes experts from a range of fields, including law, ethics and regulation (making it a multi-stakeholder platform.[74] The Initiative has released a major report, *Ethically*

[72] The Framework is not a UN document, but has instead been developed by the NGO Shift, and the global auditing firm Mazars. See: Shift and Mazars, *UN Guiding Principles Reporting Framework* < https://www.ungpreporting.org/>.
[73] Microsoft, *Salient Human Rights Issues Report – FY17* (2017) 6.
[74] IEEE, 'The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems- Executive Committee Descriptions & Members (As of 12 December 2017)' <https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ec_bios.pdf>.

Aligned Design (v1 and v2), which it hopes will 'facilitate the emergence of national and global policies that align with these principles.'[75]

Interestingly, human rights feature prominently in *Ethically Aligned Design.* While the report takes a broad ethics based approach (drawing from over two thousand years' worth of classical ethics traditions')[76], it is littered with references to human rights standards, and defines 'ethical AI' in reference to its compliance with these. In fact, in articulating General Principles, Principle 1 is 'Human Rights', accompanied by the guiding question, 'how can we ensure that A/IS do not infringe upon human rights?'[77] The analysis which follows is, from a human rights law perspective, underdeveloped. This may be due to the fact that the Committee does not include expertise in human rights. Nonetheless, the Committee clearly sees value in adopting a human rights based approach.

The IEEE has not yet released standards relating to AI, however it has begun the process of developing them. It has established a suite of Standards Working Groups (P7000) on a number of key issues.[78] The process is expected to take several years before standards are ready for adoption. While IEEE standards are voluntary, they can play an important role in shaping industry behaviour. Involvement of the human rights community in these, and similar, projects may be a relatively simple way of increasing the alignment of industry self-regulation initiatives with human rights standards.

5.3.3. Multi-stakeholder initiatives

Another possibility for regulating AI technologies is the use of multi-stakeholder initiatives (MSIs). MSIs are collaborations between businesses, civil society and other stakeholders that seek to address issues of mutual

[75] IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, *Ethically Aligned Design – Version II* (undated) <https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf>.
[76] Ibid, 8.
[77] Ibid, 22.
[78] The Standards Projects are: IEEE P7000 - Model Process for Addressing Ethical Concerns During System Design; IEEE P7001 - Transparency of Autonomous Systems; IEEE P7002 - Data Privacy Process; IEEE P7002 - Data Privacy Processes; IEEE P7003 - Algorithmic Bias Considerations; IEEE P7004 - Standard on Child and Student Data Governance; IEEE P7005 - Standard on Employer Data Governance; IEEE P7006 - Standard on Personal Data AI Agent Working Group; IEEE P7007 - Ontological Standard for Ethically driven Robotics and Automation Systems; IEEE P7008 - Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems; IEEE P7009 - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems; IEEE P7010 - Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems. See: IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 'Background, Mission and Activities of The IEEE Global Initiative' <https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ec_about_us.pdf>.

concern, including human rights and ethics.[79] Critically, an MSI is characterised by multi-stakeholder representation at the *decision-making level.* It is this which distinguishes an MSI from, for example, an advisory body with representatives from various sectors. While the functions of some MSIs are limited to promoting learning and exchange, others have a more explicitly regulatory role, which they achieve through forms of standard-setting, monitoring and accountability for compliance with these standards, and certification.

MSIs have traditionally featured heavily in the mining and energy, agriculture, forestry and fishing, and consumer goods sectors, and have more often than not been industry specific.[80] However exceptions exist; the Global Network Initiative operates in the information and communications technology industry,[81] and a number of MSIs operate across industries (such as the Ethical Trading Initiative and the UN Global Compact).

At present, several MSIs operate in the AI space, although none have a standard-setting function. These include the Council on Extended Intelligence,[82] and the IEEE's Global Initiative on Ethics of Autonomous and Intelligent Systems. The leading MSI is arguably the Partnership on AI. The Partnership was founded in 2016, and now includes over 50 member organisations from industry and civil society. This includes representatives of the human rights movement; Amnesty International, Human Rights Watch, the Electronic Frontier Foundation, and Article 19 are all members. The Partnership also has representation from a range of academic institutions, as well as UNDP and UNICEF.[83] While the Partnership's membership is booming, its aims remain modest. It does not have an explicit intention to develop standards, let alone act as an accountability mechanism for their implementation. Instead, the Partnership describes its goal in this regard as to 'develop and share best-practice methods and approaches in the research, development, testing, and fielding of AI technologies'.[84] This work is undertaken under six thematic pillars, including: safety-critical AI; Fair, Transparent, and Accountable AI; AI, labor, and the economy; and AI and social good.

At present, therefore, the Partnership for AI lacks the standard-setting function to enable it to play a genuine regulatory role. However, its broad membership makes it the most likely entity to take on this role in the future. The involvement of human rights organisations makes it more likely that any outputs from the Partnership – best practices, or eventual standards – will be grounded in a human rights based approach.

---

[79] MSI Integrity, 'What are MSIs?' <http://www.msi-integrity.org/what-are-msis/>.
[80] MSI Integrity and the Duke Human Rights Center at the Kenan Institute for Ethics, *The New Regulators?: Assessing the Landscape of Multi-Stakeholder Initiatives* (2017).
[81] See: Global Network Initiative, <https://globalnetworkinitiative.org/>.
[82] See: Council on Extended Intelligence, < https://globalcxi.org/>
[83] For a full list of members, see: Partnership on AI, 'Meet the Partners' <https://www.partnershiponai.org/partners/>.
[84] Partnership on AI, 'Our Work' <https://www.partnershiponai.org/about/#our-work>.

An AI certification scheme?: Dr Finkel's Turing Stamp

As noted in the Issues Paper, Chief Scientist of Australia, Dr Alan Finkel, has proposed a voluntary certification scheme for 'ethical AI', tentatively named the Turing Stamp. Under the scheme, companies could voluntarily apply for certification which would be granted on the basis of their compliance with ethical standards, and would be independently audited. Dr Finkel draws a parallel to the Fairtrade mark, which is the certification component of the Fairtrade multi-stakeholder initiative.[85] Other MSIs also include certification and compliance functions, including the Forest Stewardship Council (FSC) and Marine Stewardship Council.

While some multi-stakeholder certification schemes have been shown to produce positive outcomes[86] (although this is contested),[87] AI development and deployment demonstrate certain features which make a voluntary certification scheme ill-advised. These include:

- AI is not industry-specific

Certification schemes rely on the elaboration of standards by the MSI which industry actors will commit to being bound by. These standards must be sufficiently specific to support meaningful audit. Consequently, the scope of certification schemes tend to be limited. For example, MSC certifies only wild marine and freshwater fisheries. Fairtrade is a broader scheme, and has several sets of standards covering different producers and processes. However, the overall scheme remains limited to agricultural products and a small number of manufactured products (such as sports balls).

The potential applications of AI are immense, and cross sectors, and include everything from autonomous cars, robotics, algorithmic decision-making in legal settings, interpretation of medical images, social media applications, and beyond. This would present real challenges to standard-setting in an MSI.

[85] Fairtrade Australia New Zealand, 'What is the Fairtrade Mark?' <http://fairtrade.com.au/What-is-Fairtrade/What-is-the-Fairtrade-Mark>.

[86] See, e.g.: Michael Warner and Rory Sullivan (eds.), *Putting Partnerships to Work: Strategic Alliances for Development between Government, the Private Sector and Civil Society* (2004); Eddie Rich and Jonas Moberg, *Beyond Governments: Making Collective Governance Work* (2015).

[87] See, e.g.: Elizabeth Fortin, 'Transnational Multi-Stakeholder Sustainability Standards and Biofuels: Understanding Standards Processes' (2013) 40(3) *Journal of Peasant Studies,* 563-587; Sandra Moog, et al, 'The Politics of Multi-Stakeholder Initiatives: The Crisis of the Forest Stewardship Council' (2015) 128(3) *Journal of Business Ethics,* 469-493; Luc Fransen and Ans Kolk, 'Global Rule-Setting for Business: A Critical Analysis of Multi-Stakeholder Standards' (2007) 14(5) *Organization,* 667-684; Karin Bäckstrand, 'Multi-Stakeholder Partnerships for Sustainable Development: Rethinking Legitimacy, Accountability and Effectiveness' (2006) 16(5) *Environmental Policy and Governance,* 290-306.

- AI applications are complex

Certification schemes tend to be applied to commodities, or products composed of a single commodity (think coffee, or chocolate, or paper). AI applications tend to be complex - with components or engineering from a range of sources. This raises problems akin to the value chain challenges seen in, for example, textile and clothing certification processes.[88] While efforts to increase accountability along value chains has met with some success, it remains entirely unclear how - from technical and commercial perspectives - the 'black box' of AI might be opened up. It is also unclear how a certification scheme could account for the diffuse nature of AI product development, including for example the use of Open Source code - which is prevalent, and makes tracing the genesis of a piece of code virtually impossible.

Moreover, human rights issues can arise at any number of stages in an AI application's journey from concept to implementation. For example, even if an algorithm is certified 'fair', its application may produce discriminatory results if the data being fed into it contains biases. This raises two issues. The first is, what is to be certified? The 'mechanics' of the AI (such as an algorithm), or its results when applied in the real world? And if it is the 'mechanics' which are to be certified, will doing so potentially lend legitimacy to a product which may still be applied in ways which produce outcomes which do not comport with human rights?

- Certification cannot capture the systemic human rights impacts of AI

Many of the concerns regarding AI relate to the cumulative impact of multiple AI technologies. For example, AI technologies are expected to dramatically change the nature of work and the labour market. These changes may have negative impacts on a range of human rights, and threaten to disproportionately impact lower-skilled workers, who already experience disadvantage. These sorts of systemic impacts cannot be mitigated through a certification scheme.

- Certification relies on consumer choice - which isn't always available

Dr Finkel envisages a certification scheme being effective because 'consumers and governments could use their purchasing power to reward and
encourage ethical AI'.[89] This is the model which is applied to other certification schemes, which deal with products from agricultural products, to fish, wood and paper, and more recently beef. However, many AI applications are invisible to many end-users. It is unlikely that consumer

[88] See, e.g.: Niklas Egels-Zandén and Henrik Lindholm, *'Do Codes of Conduct Improve Worker Rights in Supply Chains? A Study of Fair Wear Foundation'* (2014) 107 *Journal of Cleaner Production,* 31-40.

[89] Dr Alan Finkel, 'Artificial Intelligence – A Matter of Trust' (Keynote address at a Committee for Economic Development of Australia event titled 'Artificial Intelligence: Potential, Impact and Regulation' in Sydney, Australia, 18 May 2018) < https://www.chiefscientist.gov.au/2018/05/speech-artificial-intelligence-a-matter-of-trust/>.

action would be able to sufficiently motivate either participation in a voluntary certification process, or compliance with its standards.

It is therefore our view that certification through a multi-stakeholder process is unlikely to achieve the regulatory goal of protecting human rights at risk from the application of AI technologies. Where a government or co-regulatory agency acts as the certifier, and the scheme sits within a broader principles-based or co-regulatory framework, a certification scheme may however play a useful role.[90]

5.3.4. Human rights by design

At the company level, technical solutions may also be employed to enhance the protection and respect for human rights in the context of AI technologies. The business and human rights consultancy BSR has proposed the concept of 'human rights by design' to refer to processes that integrate human rights considerations during key milestones in product development and deployment.[91]

'Human rights based design' is based on the well-established practice known as 'privacy by design' (PbD).[92] PbD refers in essence to data protection through technology design, and is based on the assumption that privacy interests are best served when protections are proactively embedded into the design and operation of relevant systems. BSR argues that 'there are opportunities to integrate a broader range of human rights considerations—such as non-discrimination, freedom of expression, and labor rights—into existing [privacy protection] processes.'[93] They also propose that best practices in human rights due diligence could be applied within a human rights based design process.[94]

Although the concept of 'human rights by design' is promising, it is important to acknowledge particular features of privacy regulation which might render 'piggybacking' of human rights principles problematic, and complicate efforts to entrench human rights by design. While PbD is concerned with technical solutions, it applies principles which have a firm foundation in law. Over the course of decades, privacy principles have been distilled into Fair Information

[90] See, for example, the use of certification under the GDPR (see case study on co-regulation under the GDPR).
[91] Dunstan Allison-Hope and Mark Hodge, 'Artificial Intelligence: A Rights-Based Blueprint for Business', *BSR*, Working Paper 3, (2018).
[92] See: Ann Cavoukian, 'Privacy By Design: The 7 Foundational Principles' (2011) <www.privacybydesign.ca/content/uploads/2009/08/7foundationalprinciples.pdf>; Ira Rubinstein, 'Regulating Privacy by Design' (2011) 26(3) *Berkeley Technology Law Journal,* 1409-1456; Commissioner for Privacy and Data Protection, 'Privacy by Design: Effective Privacy Management in the Victorian Public Sector', Background Paper (undated).
[93] Allison-Hope and Hodge, above n 91, 5.
[94] Ibid, 14.

Principles, which have been enshrined as law in many jurisdictions. In Australia, these principles underpin the *Privacy Act 1988* (Cth), and the Australian Privacy Principles (APPs) which are made pursuant to it. Critically, regulators have in recent years embraced the idea of PbD, and introduced requirements for its adoption by businesses. Australia was a leader in this respect, with APP 1 requiring relevant entities to 'take reasonable steps to implement practices, procedures and systems that will ensure the entity complies with the APPs and any binding registered APP code, and is able to deal with related inquiries and complaints.'[95]

Needless to say, the legislative framework protecting human rights in Australia is considerably weaker. A 'by design' approach concerns the operationalisation of certain principles. BSR seems to envisage that companies themselves will define the principles which will guide their design processes, drawing on human rights standards and due diligence practices. We submit that this would lead to a considerably weaker, more fragmented, less meaningful process than exists in the case of PbD. Instead, the lesson to be learned from PbD is that a 'by design' approach depends on effective standard-setting from government, and is made more effective by deliberate guidance and 'nudging' by authorities.[96]

5.3.5. Is self-regulation enough?

In short - no. While proponents of self-regulation point to its potential to 'build into the social structure of the regulated enterprise a sustained and effective commitment to insecure or precarious values,'[97] others consider self-regulation doomed due to lack of accountability and transparency, free rider issues, weak oversight and enforcement,[98] inadequate public involvement,[99] and the tendency of corporations to put profit before public interest.[100] These drawbacks have led some commentators to conclude that self-regulation is often little more than 'window-dressing'.[101] Empirical research casts doubt over the effectiveness of self-regulation in practice, concluding that

[95] *Privacy Act 1988* (Cth), Sch. 1, Australian Privacy Principle 1. A similar requirement for PbD can be found in the GDRP, art 25.
[96] See, e.g.: Dag Wiese Schwartum, 'Making Privacy by Design Operative' (2016) 24 *International Journal of Law and Information Technology,* 151-175.
[97] Joseph Rees, *Reforming The Workplace: A Study Of Self-Regulation In Occupational Safety* (1988) 10.
[98] See a helpful discussion of the relevant literature in: Ira Rubinstein, 'The Future of Self-Regulation is Co-Regulation', in Evan Selinger, et al (eds.) *The Cambridge Handbook of Consumer Privacy* (2018)
[99] See: Margot Priest, 'The Privatization of Regulation: Five Models of Self-Regulation' (1998) 29 *Ottawa Law Review* 233, 240-41.
[100] Colin Bennett and Charles Raab, *The Governance of Privacy: Policy Instruments in Global Perspective* (2003) 134.
[101] Short, see note 67, 24.

participating companies perform no better (and sometimes perform worse) than their counterparts which do not self-regulate.[102]

More specifically, researchers have identified a number of conditions under which 'blue moon'[103] self-regulation - being that which successfully achieves public regulatory goals - is more likely to occur. Short has summarised these conditions as: first, when government regulators have sufficient resources to monitor and sanction; second, when government regulators refrain from using these resources to force companies to adopt self-regulatory measures; and third, when there is reasonable consensus among regulators and regulated entities about the norms or standards governing behaviour but divergence on the methods of achieving compliance with those norms.[104] It is clear that these conditions are not present in the case of AI technologies. In particular, consensus has not been reached regarding applicable standards, and in the absence of relevant legislation (and a regulator), there are limited external deterrence pressures. Without these, self-regulatory initiatives tend to fail.[105]

More fundamentally, some of the threats to human rights posed by AI technologies relate to the overall, systemic impact of multiple technologies. For example, there are human rights implications to facial recognition software being used to track everyone's every move, even if a specific piece of software has been engineered to remove discriminatory bias. These systemic impacts require a coordinated, 'big picture' view, which self-regulation cannot offer.

This is not to say that self-regulation is pointless or to be discouraged. The initiatives described in this Part can play a critical role in building a culture of compliance, in encouraging a spirit of collaboration within and between sectors, and in facilitating the generation and exchange of knowledge and best practice. Most importantly, these initiatives may advance consensus on key definitions, principles and standards on human rights-compliant AI. They should not, however, be considered the best option to fill the existing regulatory void. Instead, any self-regulation should take place within a regulatory framework which includes more direct intervention by government.

5.4. Co-regulation

Co-regulation presents a third option between direct government regulation and self-regulation. Proponents of co-regulation claim that it represents the

---

[102] Jodi Short and Michael Toffel, 'Making Self-Regulation More Than Merely Symbolic: The Critical Role of the Legal Environment' (2010) 55 *Administrative Science Quarterly,* 364-365. See also: A. Michael Froomkin, 'The Death of Privacy?', (2000) 52 *Stanford Law Review*, 1524-27; Chris Jay Hoofnagle, 'Privacy Self-Regulation: A Decade Of Disappointment', *Electronic Privacy Information Center* (2005) <http://epic.org/reports/decadedisappoint.pdf>.

[103] This term was coined by Jodi L. Short in ' Self-Regulation in the Regulatory Void: "Blue Moon" or "Bad Moon"?' (above n 65) and is used widely in the literature.

[104] Short, above n 65, 24.

[105] Short and Toffel, above n 102.

best of both worlds by offering the flexibility of self-regulation while maintaining the supervision and rigor of government rules.[106] In co-regulatory approaches, industry enjoys considerable flexibility in shaping self-regulatory guidelines, while government sets default requirements and retains general oversight authority to approve and enforce these guidelines.[107] In practice, this often involves industry making its own arrangements, with government providing principles-based regulatory backing, and playing a monitoring and enforcement role.

Because of their capacity to combine technology neutral legislation and technology-specific instruments (such as codes and technical standards), co-regulatory approaches may address some of the challenges in regulating new technologies. The potential of these approaches has been especially well-explored in relation to privacy and data protection.[108] In fact, co-regulatory approaches have been at the heart of the EU's approach to privacy and data protection regulation. The GDPR which came into force this year,[109] endorses a number of co-regulatory mechanisms, including codes of conduct,[110] standardisation, and certification by accredited bodies (see case study below).[111]

Co-regulation also features in the regime established under the Privacy Act. The Office of the Australian Information Commissioner (OIAC) is empowered to request an entity to develop an enforceable code and to apply to the Commissioner for the code to be registered, or to develop a register a code itself.[112] An entity may also develop and register a code of its own initiative. The OIAC keeps a register of these codes, which are binding.

Another example of the use of co-regulation in Australia is the regulation of radio and television content. Industry groups have developed codes under section 123 of the *Broadcasting Services Act 1992* (Cth), in consultation with the Australian Communications and Media Authority (ACMA). Most aspects of program content are governed by these codes, which include the Commercial Television Industry Code of Practice and the Commercial Radio Australia

[106] Dennis Hirsch, 'The Law and Policy of Online Privacy: Regulation, Self-Regulation, or Co-Regulation?' (2011) 34 *Seattle Law Review,* 441.
[107] Rubinstein, above n 92, 357; Darren Sinclair, 'Self-Regulation Versus Command and Control?: Beyond False Dichotomies' (1997) 19 *Law and Policy,* 529 – 559.
[108] See e.g.: Rubinstein, above n 92; Hirsch, above n 106; Irene Kamara, 'Co-Regulation in EU Personal Data Protection: The Case of Technical Standards and the Privacy by Design Standardisation 'Mandate'' (2017) 8(1) *European Journal of Law and Technology*; Christopher Marsden, 'Internet Co-regulation and Constitutionalism: Towards European Judicial Review' (2012) 26(2-3) *International Review of Law, Computers and Technology,* 211-228.
[109] Directive 95/46/EC, which was replaced by the GDPR, also endorsed codes of conduct.
[110] arts. 40-41.
[111] arts. 42-43.
[112] *Privacy Act*, Pt IIIB.

Code of Practice and Guidelines. Once implemented, the ACMA monitors these codes and deals with unresolved complaints made under them.[113]

While co-regulatory approaches show promise in the field of regulating new technologies, they are not without detractors. Criticisms of co-regulation have focused on the lack of transparency and accountability, and lack of public involvement in the process.[114] Fears have been expressed that the 'backroom' nature of discussions between industry and the regulator can lead to an overly-cosy relationship, and potentially even agency capture.[115]

This problem of regulatory capture is arguably evident in the banking sector in Australia. Several examples emerging from Commissioner Hayne's Interim Report include ASIC's reluctance to litigate misconduct,[116] taking a consultative or advisory (rather than compulsory enforcement) approach,[117] and providing inadequate infringement notices in cases of non-compliance.[118] ASIC's failure to properly exercise its statutory powers is indicative of a regulator captured by its own industry. Hayne argues this soft approach has created a culture in which the major banks are undeterred in their misconduct, thus undermining the efficacy of the regulatory authority.[119]

A human rights based approach could play an important role in mitigating these risks. Application of the norms and practices relating to the participation of rights holders would demand high levels of transparency and meaningful consultation. This could be achieved through any number of processes, including giving civil society and consumer groups a seat at the regulatory table.

Case Study: Co-regulation under the GDPR

The EU's General Data Protection Regulation (EU) 2016/679 (GDPR) came into force on 25 May 2018. The GDPR addresses privacy and other issues arising from the processing of personal data and the free movement of such data.

The GDPR contains a number of interesting co-regulatory mechanisms. The first is the use of codes of practice. Under article 40, industry bodies may prepare codes of conduct for the purpose of specifying the application of the GDPR. Such a code of conduct must contain mechanisms which enable the regulator (or an entity accredited by the regulator) to monitor compliance with

[113] For an overview of media content regulation, see: Australian Law Reform Commission, *National Classification Scheme Review* (2011) 189 – 195.
[114] Hirsch, above n 106, 441.
[115] Neil Gunningham and Darren Sinclair, *Leaders And Laggards: Next-Generation Environmental Regulation* (2002) 105-106.
[116] Commonwealth of Australia, Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Sector, *Interim Report* (2018) 280.
[117] Ibid, 283.
[118] Ibid, 274.
[119] Ibid, 288.

the code. Draft codes must be submitted to the regulator which will confirm the code's compliance with the GDPR, and will approve it.

The second co-regulatory mechanism is a certification process.[120] The arrangements for certification mirror those for codes of conduct: a controller or processor[121] may apply to a certification body[122] for approval of a data protection certification mechanism and a data protection seal or mark, which demonstrates compliance with the GDPR. Importantly, an explicit role of the regulators[123] is to encourage the drawing up of codes of conduct, and the establishment of data protection certification mechanisms and of data protection seals and marks.

It is too soon to assess how these processes will work in practice (although it is worth noting that an equivalent code of practice mechanism was in place under the GDPR's predecessor). The outcomes may provide guidance on possible models of co-regulation for AI.

[120] GDPR, arts 42-43.
[121] This a natural or legal person, public authority, agency or other body which carries out processing of personal data belong to an individual.
[122] Which is, again, accredited by the regulator, or in this case, certain national accreditation bodies.
[123] These are referred to as 'supervising authorities' in the GDPR, and differ across the various EU jurisdictions.

5.5. Non-regulatory measures for public authorities

While much of the development and deployment of AI technologies is undertaken by private corporations, government bodies also play a range of important roles. Some public agencies are involved in R&D activities (for example, Department of Defence and CSIRO).[124] A large, and increasing, number of public agencies use AI applications in their work. These include the use of algorithmic decision-making applications in immigration[125] and social services,[126] and the use of cognitive technologies in health research and diagnosis.[127] The state carries human rights obligations under international law, and public sector agencies should take a number of operational measures to identify and mitigate risks to human rights.

Although part 5.3 of this submission focused on self-regulatory options for private corporations, a number of the measures and tools discussed can also be used by public sector agencies to improve their own practice. AI and human rights policies, human rights impact assessments, live testing, audits, as well as 'human rights by design' approaches can – and should – be implemented by public sector agencies working on, or with, AI technologies.[128]

---

[124] See, e.g.: Dr Larry Marshall, 'Artificial intelligence and Australia's industries of the future' (Speech given at the AFR Innovation Summit, 30 July 2018) < https://blog.csiro.au/artificial-intelligence-and-australias-industries-of-the-future/>.

[125] Justin Hendry, 'Australia's new visa system could use AI to spot dubious applicants' *IT News* (16 Jan. 2018) <https://www.itnews.com.au/news/australias-new-visa-system-could-use-ai-to-spot-dubious-applicants-481148>.

[126] Simon Elvery, 'Do you fit the algorithms' mould? If not, they might 'screw' you' *ABC News* (21 Aug. 2018) < http://www.abc.net.au/news/2018-08-21/algorithmic-decisions-accountability-fears/10139612>; Simon Elvery, 'How algorithms make important government decisions — and how that affects you' *ABC News* (21 Jul. 2017) < http://www.abc.net.au/news/2017-07-21/algorithms-can-make-decisions-on-behalf-of-federal-ministers/8704858>; 'Dominique Hogan-Doran SC, 'Computer says "no": automation, algorithms and artificial intelligence in Government decision-making' (2017) 13(3) *Judicial Review: Selected Conference Papers: Journal of the Judicial Commission of New South Wales*, 345-382.

[127] Beverley Head, 'Why cognitive computing is good for our health' *Financial Review* (12 Aug. 2016) < https://www.afr.com/news/special-reports/the-cognitive-era/why-cognitive-computing-is-good-for-our-health-20160711-gq3cv8>. A major conference on this topic is planned for November 2018 in Melbourne (AI, Machine Learning & Robotics in Health: Demystifying AI, Machine Learning and Robotics in Healthcare, 20-21 Nov. 2018 < https://www.informa.com.au/event/conference/ai-machine-learning-robotics-health/>).

[128] The Toronto Declaration has a helpful summary of steps which states should take to identify and mitigate risks posed by machine learning applications (see above, n 30).

5.5.1. Levers to influence corporate behaviour

Government can also pull a range of non-regulatory levers to nudge private sector developers towards developing rights-compliant AI technologies. The distinction between public and private in AI sectors is extremely blurred. Complex funding and collaboration arrangements; opaque outsourcing of public services; and decisions and actions which result from a mix of public and private ingredients, all complicate traditional public / private distinctions. Public agencies should therefore require any private sector collaborator, grantee, or contractor to operate in accordance with the agency's AI and human rights policies and procedures. This may also involve creating prerequisite steps, such as human rights assessments, in government procurement processes.

A potentially powerful non-regulatory lever is funding. In Australia, a significant proportion of AI R&D funding is provided by the government, which announced in May an additional boost of $29.9 million for AI and machine learning projects over four years.[129] The bulk of this funding will be channelled through the Cooperative Research Centres (CRC) Program round 6.[130] Applications for the round have recently closed. Interestingly, neither the grant opportunity guidelines, factsheet on additional funding for AI, sample application form, nor sample partner agreement for round 6, mention ethics at all (let alone human rights). Funding programmes such as this offer a prime opportunity for the government to encourage the development of ethical AI by requiring grantees to identify, manage and report on human rights risks in their work.

[129] George Nott, 'Budget 2018: Funding boost for AI and machine learning projects' *CIO* (8 May 2018) <https://www.cio.com.au/article/640928/budget-2018-funding-boost-ai-machine-learning-projects/>.
[130] Australian Government, 'Additional funding for CRC projects in artificial Intelligence' (Factsheet) (undated).

More generally, the government should consider broader initiatives which may improve some of the systemic barriers to human-rights compliance AI, for example, the under-representation of women and minorities in technology industries, and the public's limited understanding of AI and its impact on their rights. Ensuring diversity in the development of AI technology has been touted as the most important means of mitigating risk of bias in AI decision making.[131] As these algorithms require human input (at least initially), the existing biases of individuals developing the technology risk being unintentionally encoded into it.[132] Consequently, it is important that developers come from a range of backgrounds. While some companies, such as Google, have acknowledged this,[133] it is also essential to encourage this within education – to ensure there is a diverse pool of developers from which these companies may hire. Specific initiatives which the government should consider include:

- providing incentives for more balanced representation (on the basis of gender, race, ethnicity, and disability) within technology industries;
- cooperating with universities on initiatives to integrate ethics and human rights learning in technical courses, and to encourage greater representation in enrolments; and
- conducting public education campaigns to build AI-literacy among consumers.

6. Recommended approach

While it is too early to make firm recommendations regarding the regulation of AI technologies in Australia, we submit that there are a number of steps which should be taken, and considerations borne in mind, if future regulation is to effectively protect human rights. The first is the need for an 'expanded regulatory toolkit', which draws on a number of direct, co- and self-regulatory mechanisms. The second is the importance of placing human rights at the heart of any regulatory initiatives. This includes not only clear protections for relevant human rights, but also the application of a human rights based approach at all stages of the policy process. Principles of accountability, participation and empowerment are key to this process. We therefore recommend the following:

To the Government of Australia:

1) **Strengthen human rights protections in Australia by enacting a comprehensive, judicially enforceable federal Human Rights Act.**

2) **Develop a national strategy on AI which specifically addresses the human rights implications of AI technologies, and should reflect a human rights based approach throughout.**

[131] Rachel Thomas, 'Diversity Crisis in AI, 2017 Edition', *Medium* (16 Aug. 2017) < https://medium.com/@racheltho/diversity-crisis-in-ai-2017-edition-ce20f11f1230>.
[132] Risse, 'Human Rights and Artificial Intelligence', above n 1, 2.
[133] Google AI, 'Responsible AI Practices' <https://ai.google/education/responsible-ai-practices>.

3) **Establish a new multi-stakeholder (human rights based) advisory body to guide the development of a national AI strategy.**
    - the advisory body should include experts from a range of disciplines and sectors (including private, public and community sectors)
    - the advisory body should include representation from affected rights-holders and communities, and citizens' representatives
    - Among its responsibilities the advisory committee should:
        i) propose a new regulatory framework for AI technologies, which
            (1) identifies instances in which 'bright-line' regulation is necessary to protect fundamental human rights;
            (2) makes principles-based regulation central;
            (3) includes co-regulatory mechanisms, which provide a 'seat at the table' for civil society, specific population groups and representatives of rights holders;
            (4) establishes a regulator with strong investigative, monitoring and enforcement powers; and
            (5) ensures effective access to remedy.
        ii) propose overarching principles for the development and deployment of AI technologies in Australia, which protect relevant human rights, with a focus on:
            (1) prevention of discrimination in algorithmic decision-making;
            (2) protection of due process rights;
            (3) increasing transparency, including through the elaboration of a 'right to explanation'; and
            (4) ensuring the right to review decisions made or assisted by algorithm or machine learning applications.

4) **Establish a human rights committee as part of the advisory committee**, which should be tasked with:
    - conducting meaningful, widespread community consultation on the impacts of AI technologies; and
    - researching and proposing policy responses to the systemic and social implications of AI technologies.

5) **Require all government departments and agencies to take steps to ensure their actions in developing and using AI technologies protect and promote human rights**, including by:
    - adopting policies on the rights-compliant development, acquisition and use of AI technologies;
    - instituting measures to identify human rights risks - such as human rights risk assessments - at regular milestones;
    - taking measures to mitigate these risks, throughout the life cycle of a technology;
    - adopting enhanced transparency measures; and

- providing effective means for redress where AI applications have produced discriminatory results, or otherwise infringed someone's rights.

6) **Require all private sector partners, collaborators, and recipients of public funding to demonstrate the rights-compliance of their activities and products by undertaking the measures outlined in Recommendation 5.**

7) **Take steps to increase diversity within Australia's technology industries, including in particular representation of women, people of colour, indigenous Australians, older Australians, and people with disability.**

To the Australian human rights community, including the AHRC:

1) **Articulate how existing human rights standards do or should apply to AI applications, and identify areas in which additional research or international standard-setting is required.**

2) **participate in existing and future self-regulatory initiatives on AI, including MSIs and industry standard-setting.**

3) **engage with business entities to support the development of 'human rights by design' approaches and the application of human rights due diligence to AI development and deployment.**