



RE: PwC’s Human Rights and Technology Discussion Paper Response

PwC Australia (“PwC”) welcomes the opportunity to provide comments on the discussion paper by the Australian Human Rights Commission (“AHRC”) on Human Rights and Technology. We are committed to respecting and upholding internationally recognised human rights; this is fundamental to our purpose as a firm - to build trust in society and solve important problems.

PwC globally are long standing members of the UN Global Compact. In August 2017, PwC introduced a standalone, [Global Human Rights Statement](#), and in August 2018, we published the [PwC Australia Human Rights Statement](#) to give effect to the Global Statement within Australia. PwC has also been a strategic partner with the World Economic Forum (WEF) since the 1980s, supporting its mission to create a community of leaders who are committed to improving the state of the world through public-private cooperation. In consultation with PwC’s Indigenous Consulting (PIC), PwC has taken an industry leading step to embed Aboriginal and Torres Strait Islander peoples’ human rights within and across our Australian Firm in alignment with the Declaration on the Rights of Indigenous Peoples. We are aligned with the comments in the submission made by PIC to the [Human Rights and Technology Issues Paper](#) (Issues Paper) and acknowledge the unique issues which face Aboriginal and Torres Strait Islander peoples and their interface with technology.

We are witnessing an era of unprecedented technological change, with the ‘Fourth Industrial Revolution’ (4IR) now firmly underway. How we harness this revolution will have profound implications for business, government and society at large. At PwC, we believe in rising to this challenge together, not only as a firm but as a responsible business that has a role to play in helping to shape a more prosperous, equitable and resilient society.

The changes that we are experiencing today are underpinned by rapid advances in artificial intelligence (AI), robotics and the Internet of Things (IoT), amongst other emerging technologies. To meet the many challenges the 4IR brings we believe Australia needs to:

1. ensure that we champion and resource research into different cultural understandings of AI (and other emerging technologies),
2. establish the importance of ethical AI frameworks which can act as practical and constant mechanisms while Australia tackles legislative and regulatory challenges and uncertainty, and
3. bring together the public and private sector, along with civic engagement to ensure a unified approach to the future of AI.

Our detailed feedback is below. We would welcome the opportunity to discuss our views further, please contact [redacted] on [redacted] or [redacted] if you would like to do so.

Sincerely,

[redacted]

[redacted]

[redacted], Partner

[redacted], Partner, Chief Data Scientist

PricewaterhouseCoopers, ABN 52 780 433 757
One International Towers Sydney, Watermans Quay, Barangaroo NSW 2000, GPO BOX 2650 Sydney NSW 2001
T: +61 2 8266 0000, F: +61 2 8266 9999, www.pwc.com.au
Level 11, 1PSQ, 169 Macquarie Street, Parramatta NSW 2150, PO Box 1155 Parramatta NSW 2124
T: +61 2 9659 2476, F: +61 2 8266 9999, www.pwc.com.au

Liability limited by a scheme approved under Professional Standards Legislation.

Proposal 1: The Australian Government should develop a National Strategy on New and Emerging Technologies. This National Strategy should:

(a) set the national aim of promoting responsible innovation and protecting human rights

(b) prioritise and resource national leadership on AI

(c) promote effective regulation—this includes law, co-regulation and self- regulation

(d) resource education and training for government, industry and civil society.

Comments

We support the development of a National Strategy on New and Emerging Technologies. However, PwC believes the National Strategy should include an additional point:

(e) champion and resource research into different cultural understandings of AI (and other emerging technologies)

Artificial Intelligence is a collection of technologies and is conceptually understood differently by each culture. How Western cultures understand AI is different to how Aboriginal and Torres Strait Islander cultures might understand AI.

Without studying differing cultural understandings of AI, we risk developing policies and regulation which disadvantage marginalised communities.

As part of the research, the following aspects should be considered in consultation with experts from relevant fields including; sociotechnical systems, cybernetics, and anthropology.

- Conceptual maps for AI - e.g. what other words/concepts do First Nations peoples, and people from other cultures, associate with AI?
- What metaphors do First Nations peoples, and people from other cultures, use to describe AI?
- How do people from different cultures relate to AI agents - e.g. how do First Nations peoples fit into a structure or hierarchy of relationships with AI agents/artefacts?
- What roles should AI agents play in their cultures and society more generally?

- How do First Nations peoples, and people from other cultures, feel about AI agents and interacting with them?
- What risks and opportunities are associated with AI and interacting with AI agents?
- What has influenced First Nations peoples, and people from other cultures, views / understandings of AI? E.g. personal experience, movies, news articles
- Whether First Nations peoples, and people from other cultures, are interested in better understanding AI?
- What values or principles do First Nations peoples, and people from other cultures, believe should apply to ensuring the responsible design and implementation of AI agents.

We encourage the AHRC to use its position and influence to champion and resource such cultural research, and to use the findings to inform the rest of the proposals.

Proposal 2: The Australian Government should commission an appropriate independent body to inquire into ethical frameworks for new and emerging technologies to:

(a) assess the efficacy of existing ethical frameworks in protecting and promoting human rights

(b) identify opportunities to improve the operation of ethical frameworks, such as through consolidation or harmonisation of similar frameworks, and by giving special legal status to ethical frameworks that meet certain criteria.

In response to Proposal 2, we believe ethical frameworks can play an important role in protecting and promoting human rights by:

- acting as a practical interpretation of human rights, and
- as the bridge between the rate of innovation and potentially lagging legislation and regulation.

As the pace of technological disruption accelerates, ethical frameworks can serve as a timely and practical mechanism to help guide this fast moving industry.

The Challenge

Governments, organisations and the public are faced with many issues regarding the governance and regulation of new and emerging technologies.

While the AHRC discussion paper identifies ways to modernise the approach to AI regulation, it is likely to be constrained by ‘pacing problem’. This occurs when legislation is not keeping pace with technology innovation¹.

Compounding this issue is the intensifying pace of change, increasing the challenge for traditional legal frameworks and regulatory mechanisms to keep up. While there are many proposed solutions or workarounds to the ‘pacing problem’, significant gaps remain, see Table 1.

Table 1: Pacing problem solutions and gaps²

Solution	Gap/Issue
Government Command-and-control tactics	Ineffective due to the speed of technology evolution
Legislative changes	Innovation arbitrage - where innovators will relocate to more ‘hospitable’ jurisdictions. This can also exacerbate the risk of under-capitalisation of AI opportunities
‘Soft-Law’	Can help fill the governance gap as the pacing problem accelerates although critics might decry soft law as either being too lax (and open to private abuse) or too informal (and open to government abuse). The pacing problem makes both arguments increasingly irrelevant due to technical applicability and enforcement on such a rapidly shifting environment.
Precautionary Principle - take a much more cautious approach allowing legislation to keep up	Impractical in the face of the Collingridge dilemma and the pacing problem

¹ <https://www.mercatus.org/bridge/commentary/pacing-problem-and-future-technology-regulation>

² <https://www.mercatus.org/bridge/commentary/pacing-problem-and-future-technology-regulation>

Regulate early and often	While a potentially appealing option and practical, current legislative mechanisms may not be not conducive to this approach.
Anticipatory upstream governance -more control over the way that potentially harmful technologies are developed or introduced into the larger society	Like the Precautionary Principle more generally, the specific solutions are incoherent or sometimes completely lacking. Critics are quick to issue bold calls-to-action, but rarely provide a detailed blueprint
Adopt some form of policy experimentation which can investigate new agile mechanisms of policy making focusing on technology innovation	At this point of time there does not seem to be sufficient political will or appetite for such an undertaking
Categorisation of types of algorithms is a sound approach where algorithms are assessed and grouped according to risk, scale and industry for example and are subjected to appropriate guidelines and regulation	There are challenges relating to the broad spectrum of code and technical knowledge required to apply this as well as enforcement.
AI certification process is something that is often raised as an approach which has merit	Challenges relating to technical understanding and identification of parameters for technology that is evolving at breakneck speed. ‘The centerpiece of the regulatory framework would be an AI certification process; manufacturers and operators of certified AI systems would enjoy limited tort liability, while those of uncertified AI systems would face strict liability.’

These options may address certain aspects of the pacing problem, but the legislative nature of these solutions means they are at risk of being outpaced.

Efficacy of ethical frameworks

Despite their subjective nature ethical AI frameworks derived from Universal Human Rights can provide a consistent, practical and unchanging moral direction for the development of AI. They can also serve as an intermediary to allow time for the development of appropriate regulation and policy mechanisms.

Human Rights versus Ethical Principles

While adopting a Human Rights Approach (HRA) will 'build human rights into all aspects of policy development and decision making' (notwithstanding legislative constraints already discussed) it may be less effective at the business and operational level. We believe granular, ethical principles aligned to human rights are more easily operationalised.

For example, consider the position of a data scientist working on an emergency small business loan application for a large bank. An ethical principle such as Interpretability (covering explainability, transparency and provability considerations) would be interrogated, contextualised and aligned with the organisation's business values and any relevant risk and cultural considerations. This would then be translated into specific tolerances, guidelines and directions which would be aligned at a strategic level and designed at the level of coding.

The AHRC proposes a HRA with high level human rights concepts such as promoting transparency. While this concept and others in the HRA may provide guidance at a high level they lack the consumable and practical application that ethical principles and frameworks provide.

Ethical principles enable geographical, cultural, political, local and business contextualisation making Human Rights concepts more practical and relevant to individuals. Furthermore, ethical principles enable considerations to be addressed in terms of risk, whereas a risk to human rights is much a more serious prospect.

The application of ethical frameworks is likely to be more effective in embedding human rights inspired ethical design.

Ways to improve ethical frameworks

PwC reviewed literature on ethical AI principles when developing its Responsible AI market offering³. Our review identified 155 principles from reputable sources which were then consolidated down to nine. As expected, they align with the AI industry's current thinking on which are the most important and relevant principles. It's our view that the consolidation or harmonising would not

be too onerous because the differences in reputable sources of ethical principles tend to be semantic.

PwC's nine ethical AI principles are: Interpretability; Reliability, Robustness and Security; Accountability; Beneficial AI; Data Privacy; Fairness; Human agency; Lawfulness and compliance; and Safety. Subsequently they were further compared and aligned with AI relevant Fundamental Human Rights, resulting in a comprehensive coverage of ethical AI considerations. They provide a baseline against which PwC assesses and measures the moral permissibility of AI Solutions.

Ethical Due Diligence

In order for ethical frameworks to function as an intermediary between absent or deficient legislation and regulation they must demonstrate practical ways in which they can help organisations augment current work practices. Specifically, ways to accommodate ethical considerations from corporate strategy through to developers and users. Particular attention should also be paid to how frameworks augment current risk management processes to cater for the unique risks of emerging technology.

An output of ethical frameworks should include a report or guideline which documents the ethical due diligence performed throughout the life cycle of AI development or adoption. The output would be embedded in an organisation's governance approach⁴ and be adopted as part of standard project management practices. For example, they would inform go, no-go decisions and/or act as guidelines depending on an application's associated risk. The expectation is that the output would be a comprehensive repository that, if required, could be used as evidence of ethical due diligence should there be unforeseen negative outcomes of an AI application going live.

Ethical Impact Assessment versus Human Rights Impact Assessment

PwC has engaged with clients on the responsible use of AI and data ethics. This led us to develop an Ethical Impact Assessment (EIA) which mirrors the Human Rights Impact Assessments (HRIA) outlined in the discussion document. After comparing our EIA with the HRIA we have identified several key differences and recommend considering the following:

3

<https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai.html>

4

<https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai.html>

- Encourage the use of HRIA’s by private enterprise prior to adoption international AI applications should be subjected to an HRIA.
- HRIAs should have two applications: “validation” or “by design”. The former being an assessment of already designed applications and the latter to complement and assess the effectiveness of ethical framework from inception to completion of AI technology projects.
- All HRIAs should be published and made accessible to the public. This would feed into the concept of both transparency and accountability.
- While much easier to adopt in Government due to the already mandated PIA process and some state-specific requirements, enforcement of HRIA and any associated ‘infringement penalties’ for private organisations would be a challenge.
- *Special Legal Status within Human Rights legislation* - organisations which develop or implement an AI technology which infringe on human rights, has not adhered to or adopted a certified ethical framework, or is identified as having a high ethical risk without an HRIA and committee approval, should face appropriate infringement penalties.
- Introduce third party ethical/responsible or human rights compliant AI assurance certification practices, which are aligned to or amplifies AI certification concepts. An example is the proposed ‘Turing Stamp’—a proposed voluntary certification scheme for ‘ethical AI’ that would be independently audited.⁵

The expected outcomes of the HRIA process should be one of two things:

1. Expected benefits of implementing an AI solution will substantially outweigh the ethical risks. Appropriate mitigations will need to be in place and AI solutions of a certain risk level will need to be approved by an ethics committee
 - The ethics committee should have representation from industry, relevant public interest groups, government (AHRC) and academia

- The ethics committee will need to ensure a lean, efficient process with KPIs relating to response times.

2. A requirement to complete an HRIA will encourage the use of an approved ethical framework process. Ethical risk(s) will be identified early in an application development life cycle. This will then result in appropriate mitigations or direct changes to the design of the application. This in turn will result in less rework, avoiding sunk costs, universal design and ethical design.

Ethical framework acceptance criteria for special legal status

Based on the concepts presented we have identified several criteria that could be used to certify ethical frameworks and qualify them for special legal status.

Criteria 1: Ethical principles within the framework should be directly aligned or reflect specific universal human rights articles. For example, Table 2 shows the nine fundamental human rights identified in the discussion paper mapped to PwCs best practice ethical AI principles.

Table 2: Mapping Human Rights to PwC’s Best Practice Ethical Principles.

Human Right	PwC’s Best Practice Ethical Principles ⁶
Right to equality and non-discrimination	<ul style="list-style-type: none"> • Beneficial AI - including Cooperation & Openness and Sustainability & Just Transition • Fairness - including Diversity & Inclusion • Interpretability
Freedom of expression	<ul style="list-style-type: none"> • Interpretability • Beneficial AI - including Cooperation & Openness and Sustainability & Just Transition • Data Privacy • Fairness - including Diversity & Inclusion • Safety
Right to benefit from scientific progress	<ul style="list-style-type: none"> • Beneficial AI - including Cooperation & Openness and

⁵ <https://www.smartcompany.com.au/startupsmart/startupsmart-technology/chief-scientist-alan-finkel-we-need-turing-stamp-ai-organisations-products/>

⁶ <https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai.html>

	Sustainability & Just Transition
Freedom from violence, harassment and exploitation	<ul style="list-style-type: none"> • Safety • Beneficial AI - including Cooperation & Openness and Sustainability & Just Transition • Fairness - including Diversity & Inclusion
Accessibility	<ul style="list-style-type: none"> • Beneficial AI - including Cooperation & Openness and Sustainability & Just Transition • Human Agency • Accountability
Protecting the community and national security	<ul style="list-style-type: none"> • Safety • Reliability, Robustness and Security • Data Privacy
Right to privacy	<ul style="list-style-type: none"> • Data Privacy • Lawfulness & Compliance
Right to Education	<ul style="list-style-type: none"> • Beneficial AI, including Cooperation & Openness and Sustainability & Just
Access to information and safety for children	<ul style="list-style-type: none"> • Beneficial AI, including: Cooperation & Openness and Sustainability & Justice • Safety • Lawfulness & Compliance

modifications to them in order to produce desired effects. (Arnold & Wade, 2015).

Criteria 4: Demonstrate consideration of current legal and regulatory requirements relating to the ethical AI principles. This would be achieved by assisting organisations in identifying current and emerging legislation relating to applicable ethical principles.

Closing Comments

Over the past eighteen months, PwC has conducted research and engaged clients around the globe to help develop its Responsible AI offering. There are many synergies between the work PwC has done and that of the AHRC. We appreciate the importance of the work that AHRC continues to undertake and would welcome any opportunity to support the AHRC as we believe there are significant mutual benefits.

References

Arnold, R. D., & Wade, J. P. (2015). A Definition of Systems Thinking: A Systems Approach, *Procedia Computer Science*, 44, 669-678. <https://doi.org/10.1016/j.procs.2015.03.050>

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>

Scherer, M. U. (2015). Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. *Harvard Journal of Law & Technology*, 29(2). <http://dx.doi.org/10.2139/ssrn.2609777>

Criteria 2: Ethical frameworks should clearly demonstrate how they facilitate the practical adoption of ethical considerations, risk identification and mitigation. Methodologies, governance and tools with which ethical considerations will be embedded throughout an organisation must be present and adaptable for different organisations and industries

Criteria 3: Where applicable, ethical frameworks should clearly demonstrate a practical approach to broad stakeholder engagement and consideration of downstream and complex system implications.

An effective approach for mitigating the unique risks of AI, such as reducing the digital divide, discrimination, and unexpected consequences, is the adoption or consideration of a systems thinking approach. ‘Systems thinking is a set of synergistic analytic skills used to improve the capability of identifying and understanding systems, predicting their behaviors, and devising

Proposal 14: The Australian Government should develop a human rights impact assessment tool for AI-informed decision making, and associated guidance for its use, in consultation with regulatory, industry and civil society bodies. Any 'toolkit for ethical AI' endorsed by the Australian Government, and any legislative framework or guidance, should expressly include a human rights impact assessment.

Refer to the 'Ethical Impact Assessment versus Human Rights Impact Assessment' section above.

Question E: In relation to the proposed human rights impact assessment tool in Proposal 14:

(a) When and how should it be deployed?

An initial high level risk assessment should be made on all AI and emerging tech projects to indicate whether a more detailed HRIA would be required. The proposed parameters for this assessment should include:

- Scale (Financial, population)
- Opacity
- Ethical Risk Severity (financial, physical and mental)
- Data Quality (% of proxy data used)
- Internal, external or both

Assessment of the above criteria for specific AI applications an organisation is planning to develop or implement could provide an indicative ethical risk rating and provide direction on whether or not an HRIA is required.

(b) Should completion of a human rights impact assessment be mandatory, or incentivised in other ways?

Yes - as stated above, depending on the inherent risk rating HRIA should be a legal requirement. Furthermore, HRIA should be made public as this would help address ethical concerns around transparency and accountability.

(c) What should the consequences be if the assessment indicates a high risk of human rights impact?

Those identified as meeting a certain risk threshold would automatically be referred to the Federal Government Technology and Human Rights Committee for review. The expectation is that one of the following things would occur:

1. It would be determined that the expected benefit of implementing the solution substantially outweighs the risk to human rights, and appropriate mitigations have been put in place and are approved by the committee or
2. Following approved ethical framework processes, the risk(s) to human rights (via operationalised ethical principle alignment) would be identified early in the development lifecycle. This would result in the design of appropriate mitigations, or possibly even changes to the design of the technology in question. In turn, this should result in less rework as universal design and ethical design concepts will be embedded within the technology, avoiding implementations not being approved by governing bodies and the associated wasted development efforts.
3. It is determined that unless remediations are made and/or sufficient mitigations are put in place, the risk of harm or violation of human rights is deemed too high and the AI/ human rights special legislation prevents it from being deployed.

The methods of enforcement should an organisation develop or implement an AI technology which either infringes on human rights, or is identified as having a high risk - without an HRIA and committee approval - will need to be developed, testing and implemented.

(d) How should a human rights impact assessment be applied to AI-informed decision-making systems developed overseas?

The expectation is that systems developed overseas, depending on the jurisdiction, would not only adhere to their respective legislative, regulatory and ethical controls but should also be subjected to and scrutinised in the context of Australia and accordingly the same HRIA processes.

Proposal 18: The Australian Government rules on procurement should require that, where the government procures an AI-informed decision-making system, this system should include adequate human rights protections.

Agreed - as per above, the HRIA could be aligned with the current Privacy Impact Assessment processes.

Appendix 1: Fundamental Human Rights PwC Identified as Relevant to AI Risks

Title	Fundamental Human Right
Article 1	Right to Equality: All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.
Article 2	Freedom from Discrimination: Everyone is entitled to all the rights and freedoms, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status. Furthermore, no distinction shall be made on the basis of the political, jurisdictional or international status of the country or territory to which a person belongs, whether it be independent, Trust, Non-Self-Governing territory, or under any other limitation of sovereignty.
Article 3	Right to Life, Liberty, Personal Security: Everyone has the right to life, liberty and security of person.
Article 4	Freedom from Slavery: No one shall be held in slavery or servitude; slavery and the slave trade shall be prohibited in all their forms.
Article 5	Freedom from Torture and Degrading Treatment: No one shall be subjected to torture or to cruel, inhuman or degrading treatment or punishment.
Article 6	Right to Recognition as a Person before the Law: Everyone has the right to recognition everywhere as a person before the law.
Article 12	Freedom from Interference with Privacy, Family, Home and Correspondence: No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.
Article 19	Freedom of Opinion and Information: Everyone has the right to freedom of opinion and expression; this

	right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.
Article 29	<p>Community Duties Essential to Free and Full Development</p> <p>1. Everyone has duties to the community in which alone the free and full development of his personality is possible.</p> <p>2. In the exercise of his rights and freedoms, everyone shall be subject only to such limitations as are determined by law solely for the purpose of securing due recognition and respect for the rights and freedoms of others and of meeting the just requirements of morality, public order and the general welfare in a democratic society.</p> <p>3. These rights and freedoms may in no case be exercised contrary to the purposes and principles of the United Nations.</p>
Article 30	<p>Freedom from State or Personal Interference in the above Rights:</p> <p>Nothing in this Declaration may be interpreted as implying for any State, group or person any right to engage in any activity or to perform any act aimed at the destruction of any of the rights and freedoms set forth herein.</p>