Ethics Matters.
Ethics Matters Pty Ltd
[redacted]

Submission in Response to Australian Human Rights Commission Human Rights and Technology Discussion Paper December 2019

Introduction to my Submission

My compliments to the AHRC on such a comprehensive discussion paper and consultation process.

I, [redacted], have an extensive background as a consultant in governance, risk management and regulatory compliance including as a partner in Big4 firms, as well as in professional development and training. I hold a Masters degree in Applied Ethics from the ANU.

My responses to the questions posed in this submission draws on my extensive background in financial services, which is technology intensive; my work as an independent expert advisor to the Australian Securities and Investment Commission on a complex matter; and providing pro-bono consulting services over several years to a Yolngu indigenous tourism business. I also provide casual research support on technology and governance related matters to Peter Leonard of Data Synergies, a leading business academic and technology lawyer.

I did not participate in the first round of consultations and appreciate the opportunity to comment in this round. I hope my comments are useful.

Responses to AHRC Questions:

Part A: Introduction and Framework

Proposal 1
I support this proposal

Proposal 2
I support this proposal and make the following observations:

A. Ethical reasoning can be complex and situational. Ethical theories are like political and religious theories and cultural norms (Yolngu Rom, or law, for example), each with their own advocates, and each with their own strengths and shortcomings and without universal adoption, as the Commission notes. Trade offs often occur between different moral principals or values both within and between theories. It matters who decides what are acceptable trade-offs, how they are implemented and how impacted stakeholders are informed and consent to accepting these trade offs.

B. As noted in the discussion paper, ethics encompasses the negative obligation not to cause harm. Ethics also encompasses the positive duty to assist where possible. The UN Guiding Principles for Sustainable Development (UN SDG's) is a universal framework that has direct relevance to many human rights issues. As Secretary General of the United Nations Antonio Guterres stated recently:

> *"Human rights permeate the 2030 agenda for sustainable development.*

The vast majority of goals and targets correspond to legally binding Human Rights commitments made by every member.[1]"

Notably, the adoption of the UN SGD's is gaining traction in business in Australia through organisations such as the Global Compact Network[2], which is consistent with the increasing expectations for business to act in a socially responsible manner and consistent with human rights obligations.

I believe the UN SGD's ought to be considered as a broad based ethical framework for how society can work innovatively towards building a sustainable future with human rights at the centre and as a basis for assessing ethical trade offs. They should therefore be included in the AI ethics framework. I note that Data 61 also supports the inclusion of the SGD's in an ethical AI framework.

C. Contemporary applied ethics also takes account of our increased understanding of the human condition and our cognitive decision-making processes. For example:

- Understanding what contributes to wellbeing requires an understanding of utility theory and preference theory (which Utilitarian ethical theory also takes account of)
- Understanding key psychological decision theories such as Kahneman and Tversky's Prospect Theory (risk decision making under uncertainty) and the field of Behavioural Economics (BE) is important to understand how government, business and consumers make policy decisions and act in the markets.

These theories reflect how people actually make decisions and how individual decision-making can be guided for different outcomes. They can be used for good (a BE nudge, for example) or by organisations with asymmetrical bargaining power to exploit others and cause harm (a BE sludge, for example). The Robodebt scheme, for example, has included significant BE sludges.

Prospect theory has been prevalent in a number of areas involving price setting, wages and conditions in supply chains and the gig economy, in the wage theft scandals and in the milk price wars, where dominant market actors have used asymmetrical information and bargaining power to disrupt labour markets and supply chains. Regulatory interventions such as the introduction of the Modern Slavery Act, litigation over wages and conditions in the gig economy such as with Uber, Worksafe investigations at Amazon warehouses (where AI is extensively deployed), and the FairWork Commissions actions on wage underpayments, are often required to remediate the worst effects of this asymmetrical bargaining power related to Prospect Theory.

ASIC has established a behavioural economics unit and the ACCC is concerning itself with these matters in the application of AI as they relate to the principle of fairness and trust.

Clear guidance is therefore required in an AI ethics framework as to the application of these theories for good outcomes.

Part B: Artificial Intelligence

Question A

Yes, it appears appropriate to me.

[1] https://www.un.org/sg/en/content/sg/speeches/2020-02-24/human-rights-council-remarks-the-highest-aspiration

[2] https://www.unglobalcompact.org.au/

Proposals 3, 4, 5, 6, and 7
I support these proposals

Proposal 8
I support this proposal

As Secretary general to the United Nations, Antonio Guterres said in his recent speech:

> *"Human rights are about the dignity and worth of the human person….they expand the horizons of hope, enlarge the boundaries of the possible and unleash the best of ourselves on our world.[3]"*

Any infringement of human rights has the potential to undermine the dignity and worth of the person, and any infringement regardless of scale of impact can lead to a slippery slope where human rights are seriously eroded. This can never be allowed to happen.

Question B
I am not qualified to comment on the notion of rebuttle presumption.

However, it may be worthwhile considering the 'organisational competence obligation' and the 'responsible managers' regime that exists for financial services licensees under s912A(1)(e) of the Corporations Act, as explained in ASIC Regulatory Guide 105[4], as a framework for achieving the desired ends that this question seeks to address.

Proposal 9
I support this proposal

Proposal 10
See my answer to Question B

Question C
Yes.

Question D
Consideration should be given as to whether the law should require that a proper risk assessment system is in place consistent with ISO 31000:2018 Risk Management and the proposed ethics framework, which should identify when human decision makers ought to intervene in the process of AI decision making.

Precedence for this approach exists Under s912A(1)(9)The Corporations Act 2001 s912A(1)(9) where financial services licensees that are responsible entities are legally obliged to have adequate risk management systems - See ASIC Regulatory Guide 259: Risk management systems of responsible entities[5].

Proposal 11
I support this proposal

I also note that neuroimaging technology is being commercially marketed in Australia. Notable recent coverage was given in the AFR to Emotiv (www.emotiv.com) who produces products to

[3] https://www.un.org/sg/en/content/sg/speeches/2020-02-24/human-rights-council-remarks-the-highest-aspiration
[4] https://asic.gov.au/regulatory-resources/find-a-document/regulatory-guides/rg-105-licensing-organisational-competence/
[5] https://asic.gov.au/regulatory-resources/find-a-document/regulatory-guides/rg-259-risk-management-systems-of-responsible-entities/

gather, analyse and experiment with brain data including for monitoring in the workplace. This is fraught with a myriad of serious human rights and mental health implications as well as workplace surveillance issues.

The Australian Government should also urgently introduce a legal moratorium on the development and use of neuroimaging and brain computer interface technology for non-medical use until their application for non-medical purposes are addressed by the appropriate regulatory authorities.

Proposal 12 and 13

I support these proposals

Proposal 14

I support the premise if this proposal, however it is critically important that the Human Rights impact assessment tool be impartial to political influences.

The Australian Government is not an impartial participant in the field of human rights and the current Liberal government has a history of pushing boundaries that breach international human rights conventions and standards such as:

- The treatment of refugees that arrive by boat and their ongoing detention
- Its proposed Religious Discrimination Bill that will cause harm to women, LGBTQI people and the disabled, as Human Rights Commissioner Edward Santow noted in recent Senate hearings
- Its Robodebt program that has come under significant criticism for the unfair (and illegal) treatment of welfare recipients
- The use of facial recognition technology for security and policing matters without clear regulatory guidance on how, why and when it can be used.

I believe it is therefore more appropriate to have the Australian government commission an independent, non-partisan agency under the auspices of the proposed AI Safety Commissioner to undertake the development of the Human Rights impact assessment tool.

Question E

a) It should be available to AI developers for voluntary deployment at any stage of the development cycle and mandatorily deployed for any AI that is assessed in the risk management process as having a human rights impact before the AI goes 'live', and when there are consumer complaints that human rights are being impinged.
b) It should be mandatory. A mandatory requirement sends a clear signal to developers that they must factor human rights impacts into the development of AI, and this will minimise the potential for 'bad' AI to be developed and promote the responsible development of AI.
c) AI cannot be deployed or commercialised where it has a high risk of human rights impact
d) If the AI developed oversees is subject to and meets a human rights assessment that is consistent with Australian requirements then it ought to be approved for use in Australia. If not it should be subjected to the same assessment as an Australian developer.

Proposal 15 and Question F

My understanding from the discussion paper is that the regulatory sandbox would differ from the Human Rights impact assessment tool in Proposal 14,in the following way:

The human rights impact assessment tool would specifically apply to assessing impacts against various UN standards and conventions for human rights and relevant Australian legislation, while the regulatory sandbox would have a broader application for AI developers that should include testing against the ethical framework and the Sustainable Development Goals (see my response to Proposal 2) as well as consumer protection and other relevant legislation as the discussion notes.

Given the impact that AI can have on the wellbeing of individuals and society more broadly I consider a regulatory sandbox to be an essential feature of the way forward in the development process for AI and supporting regulation.

As with the Human Rights Assessment toolkit I consider it more prudent to have an independent non-partisan agency under the auspices of the AI safety Commissioner develop and operate the regulatory sandbox.

As the Commission notes, The Australian Securities and Investment Commission have been operating a regulatory sandbox for several years for the Fintech sector. My limited answer to Question F is to suggest that as much a possible, for consistency both in the law and in practice, the approach that ASIC has adopted be reflected in the oversight, administration and usage of the human rights regulatory sandbox.

Proposals 16, 17 and 18

I support these proposals

Part C: National Leadership on AI

Proposal 19

I support this proposal and believe it is an essential requirement of the regulatory landscape going forward given the impact that AI is having on all aspects of our lives.

Part D: Accessible Technology

Proposals 20 – 29

In principal I support these proposals. Technology has the potential to greatly improve the individual lives of people with a disability and their participation in society and this must be assured as part of the 4th Industrial revolution.

Yours sincerely

[redacted]

Principal