Response to the Commission's Discussion Paper

Will Bateman
Australian National University

Julia Powles
University of Western Australia

We are heartened by the Commission's conscientious and diligent approach to the unique challenges presented by artificial intelligence (AI) to human rights, and commend the Commission on its preparation of the Discussion Paper and its continued efforts around community consultation and engagement.

Our comments focus on three areas of concern:

1. The appropriate policy relationship between economic development, innovation, ethics and human rights law;
2. Legal frameworks governing the use of AI; and
3. The proposed AI Safety Commissioner.

1. Economic development, innovation, ethics and human rights law

In our view, the overriding focus in all the Commission's inquiries in this field should be the protection and promotion of human rights through law, as prescribed by s 10A of the *Human Rights Commission Act 1986* (Cth). This legislative provision imposes a duty on the Commission to exercise its functions with regard to the 'indivisibility…of human rights', without reference to economic development and innovation.

Proposals 1 and 6 indicate that the potential threat of AI to human rights should be balanced against economic development and innovation. This introduces a divisibility to human rights, contrary to the statutory mandate of the Australian Human Rights Commission.

Proposal 1 sets the first objective of the proposed *National Strategy on New and Emerging Technologies* as 'promoting responsible innovation and protecting human rights'. Protecting human rights through 'law' does not appear until the third objective, and is placed on the same policy level as 'self regulation'. That approach, which qualifies human rights by reference to 'responsible innovation', subordinates legal protections to a third-order concern, places those protections alongside 'self regulation' and flouts the statutory duty imposed on the Commission by s 10A.

We see a similarly inappropriate divisibility of human rights in Proposal 6, which proposes that the Australian Government should 'undertake a cost-benefit analysis of the use of AI, with specific reference to the protection of human rights and ensuring accountability'. If human rights are a variable in a cost-benefit analysis, then they are not indivisible: by their nature, cost-benefit analyses assume that there can be benefits (speed, economic efficiency, potential output gains) which outweigh the costs (impacts on human rights). We acknowledge the qualification in Proposal 6 that AI should only be deployed if there 'are adequate human rights protections in place'. As it is currently framed, however, that rider is too weak to ensure that human rights will never be a cost which can be outweighed by other benefits.

The Commission's legislative mandate does not permit it to recommend that human rights can be traded off against perceived economic gains. To the extent that Proposal 6 urges that trade-off, it should be withdrawn and reconceived.

The promotion of ethical frameworks as a means to address human rights concerns in Proposal 2 distracts from clear, enforceable, democratically accountable legal frameworks.

We are also concerned by Proposal 2's recommendation to establish a body to inquire into 'ethical frameworks for new and emerging technologies'. We find several aspects of that proposal confusing.

First, the Discussion Paper defines 'ethics' as decisions that are 'appropriate, right or good' and refers to philosophical theories such as consequentialism and the deontological approach. As academics, we enjoy thinking about those theories, but we consider that the practical effect of concentrating on 'ethics' where AI is concerned has resulted in a failure to enact clear, enforceable, democratically accountable legal frameworks. We would urge the Commission not to be distracted by theoretical debates in formulating national human rights policy.

Secondly, and relatedly, 'ethics' has been promoted as the default regulatory language of AI by the major tech companies which sponsor many organisations – both private sector and civil society – cited in the report. Given the obvious failings of major platform companies to respect community standards, we consider the turn to 'ethics' to demonstrate a reluctance to confront the social dangers of AI. The Commission should reject any approach which distracts from the critical dangers of unregulated technology.

Thirdly, and most concerning, is the fact that a number of ethical systems are simply inconsistent with human rights law. Utilitarian branches of consequential ethics can judge an action 'good' if it produces net benefits to society, while causing extreme harm to individual people. Human rights law has its own ethical logic, which is enacted into law by the various Commonwealth statutes which the Commission administers. Plainly put, human rights law is inconsistent with consequentialism, and many branches of deontological ethics.

Again, the Commission must be scrupulous in its observation of the duty in s 10A of the *Human Rights Commission Act*: to exercise all its functions with regard to the indivisibility of human rights.

2. Legal frameworks governing the use of AI

We are glad to see the Commission propose the adoption of a number of concrete legal reforms to address the potential harms of human rights.

We agree with the substance of Proposal 4, but urge the Commission to recommend that any crown immunity from liability for serious invasions of privacy be explicitly abrogated.

Recent history has clearly shown that the entities which are most likely to commit serious invasions of privacy (especially through mass surveillance programs) are government agencies, particularly law enforcement and intelligence agencies.

We also agree with the thrust of Proposal 5, but consider that it fails to protect core human rights threatened by the use of AI. We also urge a reconsideration of the definition of 'AI-informed decision making', because it inappropriately limits the application of human rights law.

A right to be 'informed' provides a person whose rights have been affected by AI with no meaningful remedy. Articles 13-15 and 22 of the *General Data Protection Regulation* (which appear to have inspired Proposal 5) provide a more appropriate form of regulation, and ensures that (1) completely automated and unilateral decisions are unlawful, (2) there is a right to meaningful information about the logic involved in semi-automated decisions, (3) there are functional rights to contest and rectify both automated and semi-automated decisions.

It is critical that people whose rights are impacted by AI have recourse to a socially responsible person to re-make those decisions. Such recourse should not have to wait until the conclusion of an expensive and time-consuming lawsuit. The disadvantages of that approach can be clearly seen in the 'Robo-debt' scandal.

We also agree with the substance of Proposals 7 and 8, but, again, consider them to be underdeveloped without the provision of clear remedial frameworks for breaches of the obligations to explain how AI was used.

We do not agree with the substance of Proposal 10, and would urge the Commission to reconsider it.

First, current legal rules already provide the rule that 'a legal person who deploys an AI-informed decision-making system is legally liable for the use of the system.' Attribution rules in tort and contract both attach legal liability to a person who uses AI, so do public law attribution rules which attach the same liability to government officials and agencies. In that sense, Proposal 10 is surplus to need and enacting it as an 'avoidance of doubt' provision risks creating the perception (via a legislative precedent) that without such a provision, there are no pre-existing liability rules which achieve the same outcomes, which is clearly not the case.

Secondly, we are particularly concerned about the reference to a 'rebuttable' presumption of liability. The circumstances in which the default liability rules would be 'rebutted' are terminally vague. Consider the case of a government official who authorises the use of AI in exercising a statutory power to detain people with violent tendencies. If the AI system used by the official incorrectly identifies people without those tendencies, then the government official has (without anything more) failed to abide by the law. Section 75(v) of the *Constitution* provides a constitutionally entrenched remedy against that official. Any liability rules which sought to circumvent that constitutional rule would be invalid, yet they seem to the type of rules which are contemplated by the 'rebuttable' aspect of Proposal 10.

3. The Proposed AI Safety Commissioner

One of the most significant evolutions between the Commission's White Paper and the Discussion Paper is the shift in nomenclature between the proposed expert body for providing national AI leadership, from a "Responsible Innovation Organisation" to an "AI Safety Commissioner".

While we strongly welcome the focus on safety, we are concerned that the considerations that have informed the construction of the proposed AI Safety Commissioner are drawn from a consultation on very different grounds; one that conflates economic development and human rights considerations, as we cautioned against in point 1 above. At best this avoids – and at worst, undermines – the essence of what safety regulation for "the prevention of individual and community harm, and the promotion and protection of human rights" in relation to pervasive data-informed computation would truly require.

We encourage a proper consideration of the powers of a genuine safety commissioner or authority for AI, including powers of investigation, prevention, and correction proportionate to Australian bodies in other industries. For example, Australian agencies protect safety risks in the fields of gene technology (53 staff),[1] online speech (54 staff),[2] food safety (108 staff),[3] nuclear safety (130 staff),[4] therapeutic goods safety (750 staff),[5] pesticides safety (182 staff),[6] and aviation safety (859 staff).[7] There is no reason for treating artificial intelligence safety as exceptional, particularly given the reach and known harms of this technology, as the Discussion Paper articulates.

Proposal 19 uses the language of safety, but ultimately proposes a body that promotes economic development of AI as much as human rights protections. This should be revised and strongly informed by safety agencies in other fields of technology, including aviation, gene technology, and therapeutic goods.

[1] Office of the Gene Technology Regulator
[2] Office of the eSafety Commissioner
[3] Food Standards Australia and New Zealand
[4] Australian Radiation Protection and Nuclear Safety Authority
[5] Therapeutic Goods Administration
[6] Australian Pesticides and Veterinary Medicines Authority
[7] Civil Aviation Safety Authority