

Response to the Australian Human Rights Commission and World Economic Forum White Paper on “Artificial Intelligence: Governance and Leadership”

Julia Powles, Marco Rizzi, Fiona McGaughey, David Glance
University of Western Australia

As scholars of law and computer science, we have been enthusiastic followers of, and participants in, the Human Rights and Technology Project led by the Australian Human Rights Commission. The Project is a crucial one for Australia, and holds considerable promise. In this submission, however, we wish to clearly articulate what we see as considerable risks, based on the products of the Project to date and our participation in the broader consultation process.

The Project is ambitious and time is short, so we have just four points. Our hope is that they can assist in establishing a course correction to the vital and exciting core of what a human rights paradigm can bring to the domain of digital technology.

First, to seize the opportunity for leadership, the Commission should focus on its core mandate and expertise, i.e. effectively monitoring and protecting human rights. Despite its co-authorship by the Commission, we are concerned that the White Paper contains only fleeting references to human rights. Along with the Issues Paper, which took on only the narrow and discrete issues of non-discrimination and accessibility, the Project risks offering only a very limited framing of the issues in play – essentially reducing to a modest analysis of the rights to privacy, non-discrimination, and a fair trial, offering little beyond the state of existing law. It risks failing to reckon with the fundamental challenges that data-driven systems based on classification, prediction, and optimisation (for efficiency and profit, not dignity and autonomy) pose to the very notion and every aspect of human rights. We hope that the Discussion Paper will have worked-through examples of different human rights and how they are challenged by automated systems in Australia, beyond a synthesis of news headlines.ⁱ This is the distinctive contribution that is crying out for more work, and we are grateful for the Commission’s leadership on it.

Second, the Commission should not be concerned with promoting industry, nor creating a light-touch regulatory framework for industry to flourish. These concerns, if and to the extent that they have merit within a nation that is a net importer of technology, are within the power of other parts of government and industry to address. A regulatory model that fails to clearly distinguish the discrete roles of the regulator and the regulated will be destined to fail.ⁱⁱ

Third, an AI or digital technology policy council, led by governance and regulation experts, is a worthy proposition, but the single ‘regulator’ articulated as a ‘responsible innovation organisation’ in the White Paper is fundamentally misconceived and will be ineffective. Properly applied, the Commission’s unique role and expertise as a human rights body, as discussed in point 1 above, would be a great influence on joined-up policy thinking across government. But we otherwise reject the six questions formulated by the White Paper, because we believe they proceed from an unworkable premise. If we wish to talk of regulation and governance, a pluralistic approach that combines a number of effective, sectoral regulators – each of which would indeed be enhanced by a policy council that is minded to the various

challenges posed by AI and related digital technologies – is vastly preferable to a single regulator. This also mitigates against the significant risk of ‘buck passing’, or creating a governance gap, by deferring all power on a cross-cutting subject to a single body, especially one with built-in compromises. Building on point 2 above, there is no rationale for requiring a regulator or policy council that is directed at protecting individuals to make a business case for its own existence or otherwise prove its economic value.ⁱⁱⁱ

Fourth, there is a great opportunity and pressing obligation for effective regulatory design around digital technologies. We believe this opportunity is missed by the White Paper, but hope it can be addressed by the Project overall. In particular, an effective regulator must possess both a preventative and a corrective mandate. Beyond codes of conduct, self-assessed certifications, problematic ‘trust marks’,^{iv} or cautionary fines, there must always be the possibility to stop the use of automated systems with appreciable societal costs, just as there is with every other kind of technology (see, within Australia, the Therapeutic Goods Administration^v and the Gene Technology Regulator). The proposed Regulation must also boldly address the foundational issue, which is identified but not elaborated in the White Paper – namely, the existing antidemocratic state of collection of data, ownership of datasets, and concentration among a small number of private sector entities. This creates a number of opportunities to disincentivise and devalue data hoarding with creative policies, including carefully defined bans, levies, mandated but qualified data sharing, and community benefit policies, all backed up by the brass knuckles of the law. We would be happy to elaborate further on how the Commission can work in each of these areas, complementing other parts of the Australian public sector.

ⁱ To take just a few examples, the White Paper Introduction’s claim preceding fn9 that AI will, on its own, increase the world’s GDP by 40% in 10 years lacks evidence. Similarly, in the Forward, the point about AI saving lives preceding fn1 is grossly overstated, and ignores substantial literature on the risks (see, e.g., Cabitzta, Rasoini & Gensini, ‘Unintended Consequences of Machine Learning in Medicine’, *JAMA*. 2017;318(6):517-518; Adamson & Smith, ‘Machine Learning and Health Care Disparities in Dermatology’, *JAMA Dermatol*. 2018;154(11):1247-1248.) The point about AI entrenching bias in criminal justice preceding fn2 refers to an article on the software COMPAS, but the technology clearly breaches existing equal opportunity / anti-discrimination laws (like others that are within the Commission’s scope, like targeted advertising or social credit scoring).

ⁱⁱ Recent examples include, [on a global scale with repercussions in Australia](#), the ‘Implants Files’ scandal and, as regards Australia specifically, the ground-breaking findings of the [Royal Commission into Misconduct in the Banking Superannuation and Financial Services Industry](#), which revealed frightening levels of systemic [regulatory capture](#) and flawed oversight.

ⁱⁱⁱ The White Paper (and consultations) seems reliant on inappropriate models from the UK, especially the AI Council, an industry body tasked with promoting the AI sector and AI adoption in other sectors of the economy, and the Centre for Data Ethics & Innovation, which is again an advisory body but without a policymaking or regulatory mandate. We would be better served by looking to Australian examples, such as the [ACCC’s proposed algorithmic regulator](#), the UTS Technology Assessment Office cited in the White Paper, the Therapeutic Goods Administration, or Office of the Gene Technology Regulator.

^{iv} A trust mark that is based on a standard such as ISO 9000 would be ineffective, given that this is simply a certification for the implementation of process, and does not make any judgment of actual quality or allow effective measurement of improvements in quality. In terms of other examples of trust marks or seals, these have become almost meaningless on websites and are at great risk of reproduction on phishing sites.

^v The potential magnitude of harm, and a rather grim history of drug disasters, have shaped the international consensus on pharmaceutical regulation to the effect that public regulators are vested with the mission of undertaking tight pre-marketing scrutiny and [registration of newly developed products](#) to ensure these meet the highest levels of safety and efficacy. Once a new product is marketed, the regulator performs a monitoring role over performance through instruments such as the [Early Warning System](#). Should critical information emerge indicating potential (or actual) harm, the TGA has the power to [issue recalls and withdrawals](#).