SUBMISSION

Supporting Responsible AI: discussion paper

Submission to the Department of Industry, Science and Resources

**Prepared by the Centre for Social Impact,
Flinders University**

25th July 2023

Peter Richard McDonald

Executive Summary

The Data for Good Project at the Centre for Social Impact (CSI) – Flinders University is an industry partnership between CSI and Uniting Communities (UC).

The CSI at Flinders University sits within the national CSI network. Our common purpose is to enable system change and lasting social impact for people and communities. We do this through transformational education and research, engaging within and across different sectors. CSI has a vision of a world where everyone can thrive and grow their capabilities, no matter their circumstances. We seek systems-level shifts and lasting social impact for people and communities.

Uniting Communities is a community service organisation that provides social services throughout South Australia. It has near a thousand staff who provide a wide range of services including child and youth support, homelessness support, disability support, drug and alcohol support and specialist support for Aboriginal and Torres Strait Islander peoples. UC also provides community and residential aged care.

The project researches the increasing impact of data, artificial intelligence (AI) and automated decision making (ADM) on isolated and disadvantaged people.

In June 2023 the Department of Industry, Science and Resources released the *Supporting responsible AI: discussion paper* and sought feedback on a proposed risk management approach to AI.

We support a risk management approach to managing AI and have offered six recommendations for consideration.

Our focus is on the importance of safeguarding marginalised people and communities from the effects of AI.

Background

The rise of AI and ADM has the potential for both positive and negative impacts on the lives of all Australians.

There is a complete chapter in the Robodebt Inquiry which describes the impact of ADM on social service recipients. The Inquiry found that even well-intentioned staff were unable to stop Robodebt from automatically issuing erroneous debt notices to recipients based on inaccurate calculations and questionable assumptions. Such ADM wrecked lives. As Commissioner Catherine Holmes summed up in her report of the Royal Commission into the Robodebt Scheme (Commonwealth of Australia, 2023), was a "crude and cruel mechanism that was neither fair nor legal."

The lack of recourse for the people impacted by Robodebt made this failed approach to uncovering overpayments worse. There was no access to a legitimate complaints process for recipients. Alarmingly departmental officers receiving calls to the inquiries line were unable to explain how the debts had been calculated (ibid, p.329). There was little to no access to a human who could explain the information on which automated decisions were made.

Looking forward from Robodebt

Robodebt was not an artificial intelligence (AI) system. It completed data matching and ADM without this new technology. AI now generates content, forecasts and recommendations for us. AI adds a layer of sophistication to ADM, turbo charging the ability of industry and governments to gather, process and make decisions based on very large data sets. The coming together of AI and ADM creates a new and significant risk for marginalised people and communities, on a scale that even Robodebt couldn't achieve.

Often marginalised people have limited ability to understand how AI is assisting or entrapping them and what might be done to challenge this new technology. We are concerned that as AI and ADM merges we will see further significant threats against disadvantaged people unless regulatory reform is urgently and thoroughly considered.

Our response to risk management

The Department of Industry, Science and Resources are proposing that Australia adopt a risk mitigation approach to managing AI, aligning with emerging international trends (Dept Industry, 2023, p.31). The possible elements of a draft based-risk assessment are laid out in Attachment C of

the discussion paper (ibid, p.40). We offer our support to a risk management approach. As acknowledged in the discussion paper (Dept Industry, 2023 p.32.) as risks increase so is there a commensurate increase of expectation that those risks are managed by the owner of AI. That said we have some concerns we raise as part of our submission.

Risks to whom?

Risk tools are employed to manage businesses and projects as well as the risks to patients and individuals (NHMRC, 2023 p.9.). It is not clear as to whether the proposed risk management approach for AI is just for the risks faced by the individual or whether the risks include more general risks of AI, including to the business as a whole? We want to ensure that any risk approach instituted around AI focuses on the impact that AI has on individuals in particular.

Such an approach would be congruent with your commitment to human-centred values (Dept Industry, 2023 p.14., Dept Industry, 2022) where human-centred AI systems should respect human rights, diversity and the autonomy of the individual. Assuming the priority of human centredness is the focus of the Department's proposed risk profiling tool, we seek clarification regarding what type of person is being assumed as the recipient of these risks?

We know from our work with disadvantaged people that when gauging the severity of harm, the choices, experience, values and vulnerabilities of different populations will be relevant (NHMRC p.14). Populations have within them different risk profiles. The s*afe and responsible AI in Australia* discussion paper assumes all people are of one type. Our experience of over 100 years of service delivery means we can quite confidentially state this is not the case.

If the Department wants to hold to a person-centred approach then they will need to consider what risks look like within groups of people, especially those who are at higher risk, those who are isolated or disadvantaged.

A good example is the NHMRC (2023) standards for research which acknowledge that respect for humans means that we seek to provide protection to those with diminished or no autonomy, as well as empowering them where possible and protecting and helping them wherever it would be wrong not to do so (ibid, p.9). A resource to consider is Section 4 of the NHMRC (ibid, pp63-82) standards which acknowledges the diversity of our populations and the way different risks impact on those groups.

> **Recommendation 1**
>
> The risk management approach finally determined for regulation of AI must be targeted to assisting the individuals impacted most unfairly by the evolving power of AI;
>
> **Recommendation 2**
>
> Human centredness be confirmed as the central focus of the AI risk management approach.

Medium risk category too broad

The medium risk category (Dept Industry, Science and Resources 2023, p.33.) contains too broad a range of risks. Medium risk is described as "High impacts that are ongoing and difficult to reverse." Describing the medium risk category as holding "high impacts" is confusing and unclear. The *EU AI Act risk level* has many of what the Australian Department suggests are medium risks in the proposed tool as high risk in their classifications (European Commission 2021b and Dept Industry, Science and Resources 2023, p.39).

High impacts which are ongoing and difficult to reverse should be listed in the high risk category and not in the medium risk category.

The EU AI Act risk level provides a much more logical hierarchy of risk classifications and we feel should be adopted as the framing for describing and classifying potential risks (ibid, p.39.).

> **Recommendations 3**
>
> Removal of high impacts in the definition of medium risk as presented in Box 4 (Dept Industry, Science and Resources 2023, p.33) so that it does not contain high risk matters.
>
> **Recommendation 4**
>
> Adopt the *EU AI Act risk level framework.*

High risk should be under legislative control

While we agree that low risk AI should be self-managed it is our view that high risk management requires legislation. The full protection of the law should be utilised to shield people from the potential damage AI is capable of inflicting. In particular, we seek that people who are discriminated against because of racist or biased AI tools have the opportunity to seek a legal remedy. Governments should consider creating AI specific legislation for this purpose. Such regulator action would be consistent with the legislative direction the EU is moving in (2021a, 2021b), and evident in the outcomes of the Robodebt Royal Commission.

> **Recommendations 5**
>
> Consideration must be given to drafting of legislation for high risk AI to provide legal recourse for people damaged by AI.

Complaint process

When a high risk ADM goes wrong individuals should have access to an external redress process that understands the complexity of AI. Creating a complaint process which can investigate when governments and businesses fail to provide transparency to AI and ADM is critical in managing legitimate redress. There needs to be a technical skill base in any independent entity established to host and resolve the complaint process. The function needs to be accessible to marginalised people.

> **Recommendation 6**
>
> An independent body needs to be created to consider complaints about high risk AI when not satisfactorily dealt with by the AI owner.

Specialist Ethics Advisory Panel

The EU has put together a multi-pronged plan to build trust in AI, recognising that this is essential to its uptake (2021a). One of the prongs is the establishment of a High Level Expert Group on AI to focus on the fundamental questions of ethics and AI technologies. The group has established ethical guidelines and created an Assessment for Trustworthy AI (2021a p.31).

Australia should create a specialist multidisciplinary ethics panel/forum to discuss and consider AI and its impact on the wider community. We ask that the department consider the NHMRC or approaching the existing CSIRO / National AI Centre to create and host a specialist multidisciplinary ethics panel to act as a forum for discussion regarding the developments of researchers and the community.

> **Recommendation 7**
>
> A specialist multidisciplinary ethics panel/forum needs to be created and resourced as a key body for discussing and considering responsible and ethical applications of AI.

Reference List

Commonwealth of Australia (2023) *Royal Commission into the Robodebt Scheme,* Commonwealth of Australia.

Department of Industry, Science and Resources (2022) *Australia's AI Ethics Principles,* Department of Industry, Science and Resources, accessed 20th July 2023.

Department of Industry, Science and Resources (2023) *Safe and responsible AI in Australia – discussion paper*, DISR, Australian Government, https://consult.industry.gov.au/supporting-responsible-aiaccessed 20th July 2023**.**

European Commission (2021a) *Coordinated plan on artificial intelligence 2021 review,* EC, Brussels, accessed 20th July 2023.

European Commission (2021b) *Proposal for a regulation of the European Parliament and of the Council, Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts*, EC Brussels, accessed 20th July 2023.

NHMRC (National Health and Medical Research Council) Australian Research Council (2023) *National statement on ethical conduct in human research,* NHMRC accessed 20th July 2023.

NSW Government (2022) *Artificial Intelligence assurance framework*, NSW Government, accessed 20th July 2023.