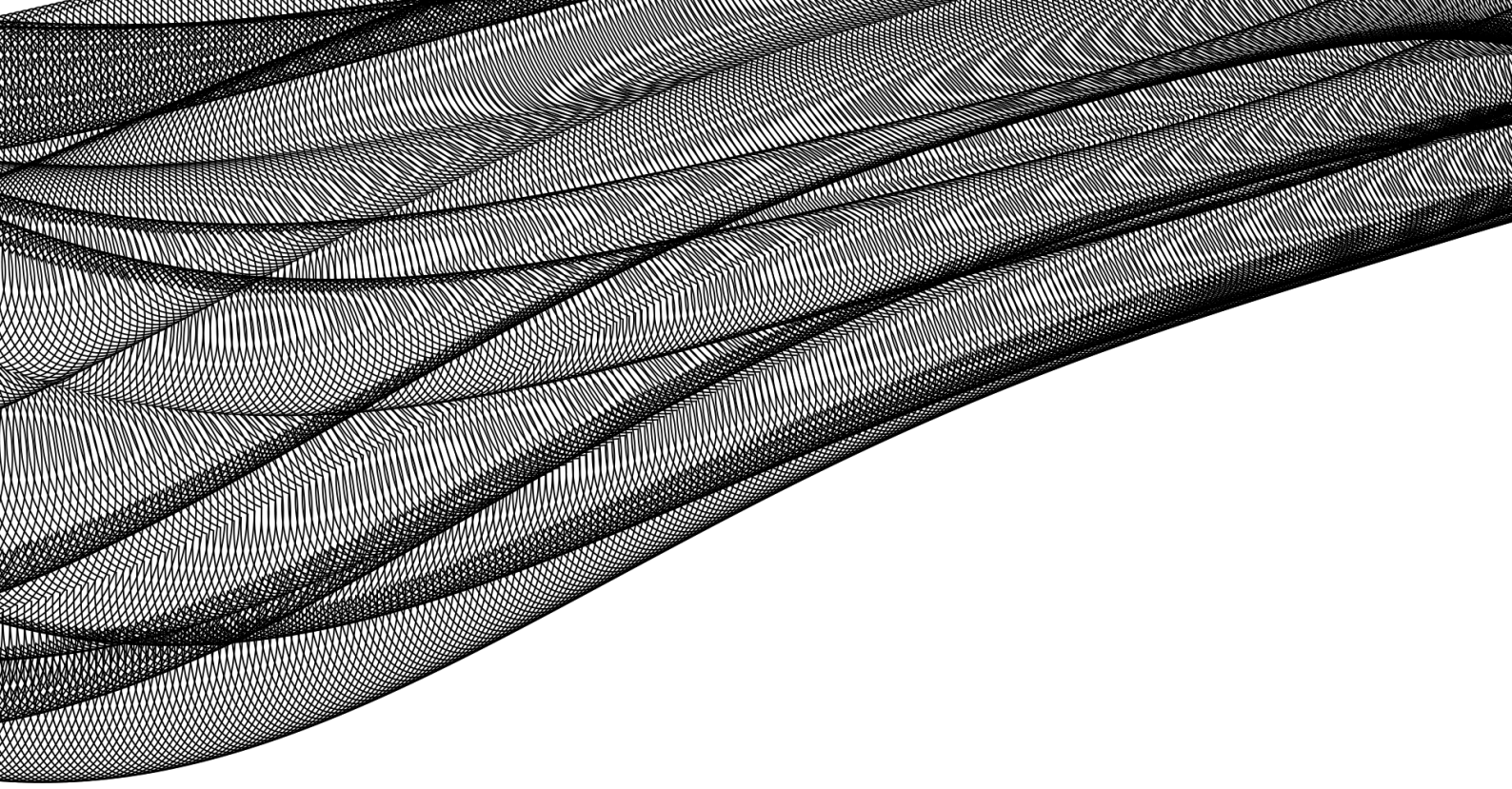


Australian Human Rights Commission and World Economic Forum Project Submission

White Paper on Artificial Intelligence: Governance and Leadership

By: Samuel Walter, Martin Kemka, Noon van der Silk and Faye Tayib

What should be the main goals of government regulation in the area of AI?

The focus should be on innovation, empowerment and equity, competition, and minimising harm and maximising community value.

1) Innovation

The main innovation drivers we want to build are: global competitiveness and supporting new entrants of all sizes. By working with industry, we can create legislation that encourages the development of AI systems that have positive impact in the community, in terms of creating jobs instead of replacing jobs, and in terms of supporting human-in-the-loop learning.

New entrants can be supported by rules and leadership relating to the handling of data: who owns it, what rights do they have to it, how can it be used? Regulation in this area can also help to protect and support the consumers.

2) Empowerment, Equity and Minimising Harm

We see a few key opportunities to explore: the Right To Explanation/ Information, and Certification. Being able to receive an explanation or at least information about why a machine reached a conclusion is crucial. In particular, it will become necessary, when a consumer receives any decision, to understand what actions they need to take in order to potentially change that judgment in the future.

Furthermore, this kind of explanation/information needs to be available to all consumers regardless of their technical capability, and must be available across all areas of application, including justice, marketing, finance, insurance and more. From a technical standpoint, this is a very challenging area, and regulation will help to encourage investment in this area of research.

Certification and validation of algorithm is another way to tackle this problem. Certification should assess how a given algorithm will affect society and the consumers. Certification need not be mandatory, but provides a signal to consumers that the algorithm has seen a base-level of scrutiny, and will provide a competitive edge to those who adopt it.

Working in the certification area will also allow us to address the unfair bias and exclusion that AI can often result in. As an example, a certification process for computer-vision applications would feature all the varieties of Australians in equal measure, and so potential computer-vision applications could be judged on how well they perform against people of all colours and appearances.

3) Maximising community value

Any AI system should have a positive impact on the community. This may be in that it makes a business more competitive, or provides a consumer service that was previously unavailable.

Directors of companies and management are obliged to make pursuing profit their highest priority. We recognise that this interest doesn´t always align with the community´s best interests and encourage this regulatory body to prevent companies putting profit above the safety and human rights of the community.

Considering how AI is currently regulated and influenced in Australia:

A) What existing bodies play an important role in this area?

We´re aware of:

ANUs 3Ai Program - https://3ainstitute.cecs.anu.edu.au/

CSIRO - being a partially-government body, the work they do informs government policy.

Standards AI Committee - IT-043 - https://iecetech.org/Technical-Committees/2018-03/First-International-Standards-committee-for-entire-AI-ecosystem

CEDAW - https://en.wikipedia.org/wiki/Convention_on_the_Elimination_of_All_Forms_of_Discrimination_Against_Women

And of course, Human Rights and Technology - https://www.humanrights.gov.au/our-work/rights-and-freedoms/projects/human-rights-and-technology

B) What are the gaps in the current regulatory system?

As AI practitioners, we´re aware of several gaps:

Consistent standards –
What exactly does Ethical AI look like, and how does this align with human rights? What is good and bad, and how do we measure that?

Accountability –
Who is responsible when an AI system makes a decision? How can we understand the chain of provenance for any data that has been used? Who is accountable within the organisation?

Explainability and Transparency –
To what level of detail should different AI systems be held? Should we hold government systems higher than private-sector based ones? What kind of explanations are useful? To whom should they be aimed? What kind of explainability is even possible? When AI is being used, is the general public aware? Is the usage transparent?

Education –
What level of AI literacy should we aim for in our community?

Jurisdiction –
In what realms can AI be applied? Are there any privileged areas where the adoption should be limited, or just face harder constraints to implementation?

Medical –
What are the constraints on building apps for medicine?

Privacy and Data Trading –
What are the limits on how organisations can share and collect data? What is the impact of combining certain datasets? What level of anonymity should we require? How do we build accessible regulation around this area?

Safety –
An overarching concern is the oversight of AI decisions, the checks and balances, reporting, and human-in-the-loop interventions that can be enacted when the AI makes an unsafe decision.

Would there be significant economic and/or social value for Australia in establishing a Responsible Innovation Organisation?

Economic Value
The value that automation represents is huge, and to have this introduced in a fair manner that ensures responsibility and support at all levels of the community will mean that this created value can be enjoyed by all. Establishing a body that provides certification or accreditation that indicates fairness and treatment of bias will increase consumer confidence in organisations that rate highly, and thus confidence in AI will build which will encourage its adoption which leads to economic growth. By addressing from the start potential issues, this will result in less data breachers, biased AI applications, and consumer issues, which will allow the market to focus on productive activities, instead of needing to spend time in court battles, bad press, and consumer backlash.

Social Value
Minority groups will not be disadvantaged because of unfair bias in algorithms. Power will be given back to consumers/individuals if an understand of how decisions are made is provided to them.
Inherent social benefit in decisions made by ethical algorithms - socially minded decisions that consider implications and negative effects.
The RIO should work to encourage involvement from all stakeholders, when an AI system is adopted, such as groups affected by policing activities, or justice reform. Such involvement will lead to better social outcomes for these at-risk groups.

Under what circumstances would a Responsible Innovation Organisation add value to your organisation directly?

For Northraine, as a registered B-Corp with a mission to recondition the human condition, it goes without saying that a RIO would add value. It would provide us with an opportunity to continue doing business how we currently do, with the added benefit of resources and guidelines to ensure we can continue to provide our clients with solutions that have a positive effect on the world. Our focus on education and encouraging participations in the growing machine learning community at all levels would be bolstered by an RIO that encourages these pursuits too.

For Braneshop, we are a teaching organisation; the RIO would add value to this in particular by encouraging more AI education. By strengthening the AI community in Australia, and encouraging global competition this opens significant education opportunities across the full range of address issues: ethical AI, unfair bias, explainability, and general AI skills.

How should the business case for a Responsible Innovation Organisation be measured?

Quantitative:

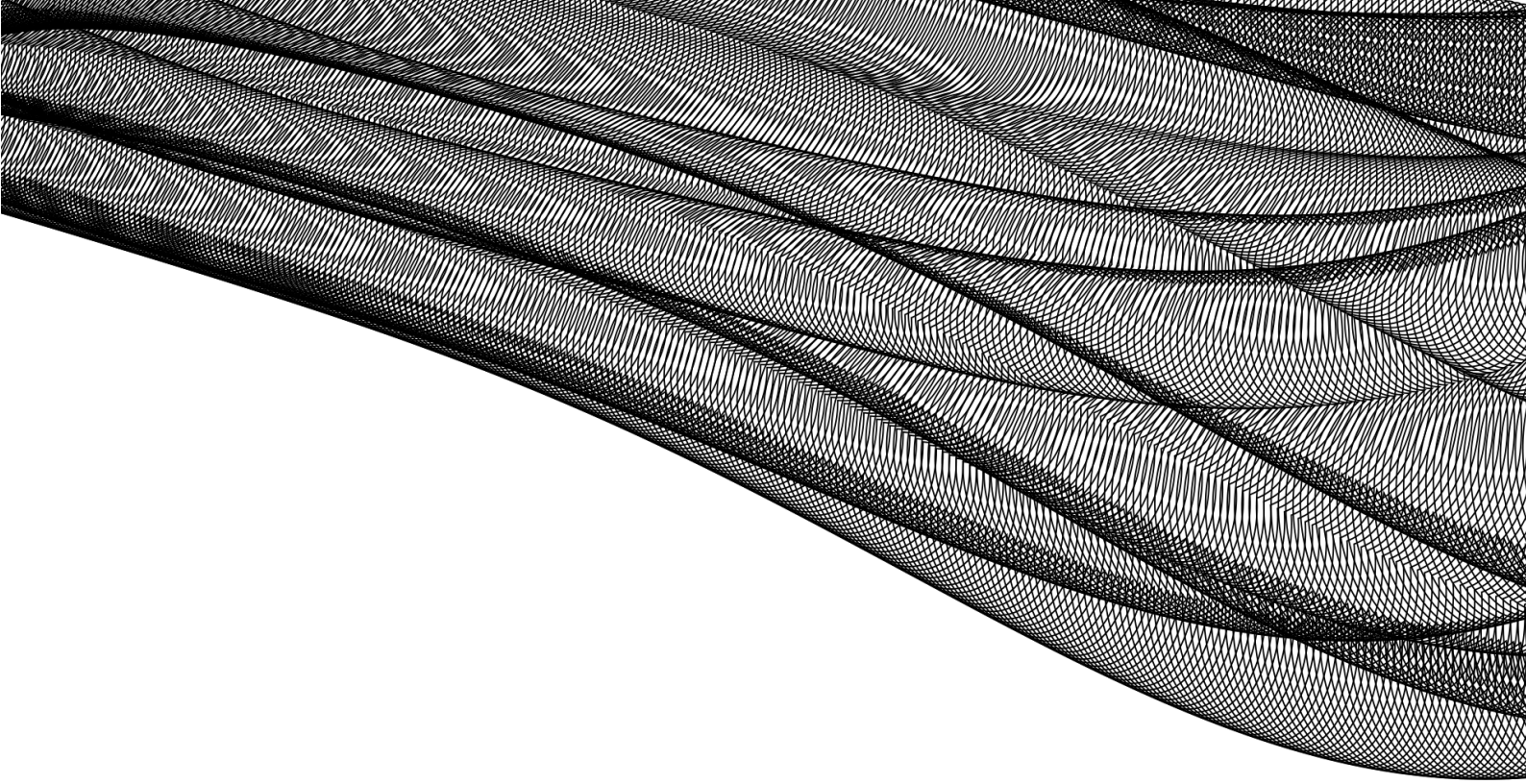
- How much money is being made by AI and what organisations are benefitting the most?
- How many decisions are being made by AI/ML, and how many people are being affected? How many of those decisions positively affect those people?
- How many AI engineers are in Australia?
- How many AI-based startups are launching each year?
- What is their success rate compared to the norm?
- How is the unemployment rate being affected?
- How much explainability-research is coming out of Australia´s institutions?

Inclusiveness in the industry - Percentage of female or minority founders, cultural diversity and mature-age workers, people who are culturally and linguistically diverse (CALD), increase in mentors across these areas. Uptake in AI education initiatives across all levels, from school, to Universities courses, to offerings to mature-age people and CALD groups.

Qualitative:

- How explainable are the adopted algorithms?
- What is the public discussion around AI focusing on?
- Are consumers feeling safer or more at risk in terms of data?
- What kinds of AI startups are being created?
- How many are community-focused?
- Are other countries following in our footsteps in terms of addressing unfair bias?

If Australia had a Responsible Innovation Organisation, what should be its overarching vision and core aims?	Transparency, fairness, accountability and access to AI education for all members of society.
What powers and functions should it have?	Concrete and objective measures of fairness and responsibility when it comes to organisations and their models that use AI. A rating and approval process to give consumers an understanding of the quality, across various axes relating to ethics, privacy, fairness, accuracy. Either the ability to enforce rules, or report to a body that can regulate and enforce socially responsible and fair AI. Ability to inform regulation across industry-applications of AI, such as in medicine and governmental decision making.
How should it be structured?	With contribution from all stakeholders, particularly those with a demonstrated record of ethical and socially responsible action in the human rights, legal, and technology fields. Create a system that ensures all participants are accountable and transparent about how their models are build. The system should involve peer assessment and promote voluntary accreditation. A combination of government and private-business on the leadership group. A floating panel of diverse members of the public and other stakeholders, brought on in an advisory role to provide guidance and context for overall direction.
What internal and external expertise should it have at its disposal?	Academia Private industry Public sector Legal International guidance and "sister" organisations Government Human Rights Organisations
How should it interact with other bodies with similar responsibilities?	The RIO should collaborate with other similar bodies. Part of the charter should be to meet with, and work collectively to understand the goals and needs of other organisations. The RIO should act in a supportive manner, while also providing leadership on the aspects that are within its domain.
How should its activities be resourced? Would it be jointly funded by government and industry? How would its independence be secured?	Ideally through non-partisan government funding, but with the provision to also be funded by collaborating public bodies, such as Universities and industry groups. Private funding could be considered, but needs to be restricted to not come from bodies whom are to be regulated or significantly impacted by the decisions of the RIO. The RIO may be able to sell services, such as education and guidance in order to obtain certification.
How should it be evaluated and monitored? How should it report its activities?	The RIO shall be required to be completely transparent in its actions. It should report its activities through public documentation and whitepapers, as well as through peer-assessment from other collaborating organisations.



Collaborative submission by:

Nôrthraine

brane shop

Authors

Samuel Walter, Martin Kemka, Noon van der Silk and Faye Tayib (independent)