

Part B: Submission

You may choose to send a document, a video, a voice recording or picture as your submission. *Please contact the Human Rights and Technology Project Team to send a file larger than 20 MB, such as an Auslan submission.*

This section includes a series of questions developed by the Commission that you may respond to. You do not need to answer every question.

Consultation questions

1. What should be the main goals of government regulation in the area of artificial intelligence?

The main goals of government regulation in the area of artificial intelligence should be to:

- a. Develop proactive measures to understand, guide and implement an ethics, values-based and human-rights-based framework for AI research, development, systems and services, which is based on a framework of risk identification and mitigation,
- b. Develop reactive measures to understand, design and implement countermeasures that operate, should the proactive measures and risk mitigations fail (ie. assigning liability and responsibility),

- c. Be consistent with international initiatives, applied universally and consistently, and
- d. Build public trust by finding the right balance between the benefits and risk minimisation of the AI technologies.

The framework must be able to foresee outcomes that may not be ordinarily foreseeable and must be able to be iteratively improved upon as new scenarios or anticipations are developed. It must be able to anticipate the worst of situations as well as the best, be able to assume manipulation, distortion or falsification of data, people and institutions, as well as the capacity for AI to act independently.

AI developers, manufacturers, purchasers and consumers should be encouraged and incentivised to make use of the regulatory framework, in order to build the public trust.

2. Considering how artificial intelligence is currently regulated and influenced in Australia:

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

Current government regulatory bodies exist in a fragmented way. Those that influence how AI is currently regulated and influenced, and affect its current development and deployment in Australia include:

- Commonwealth Privacy Act, 1988 and updates
- National Transportation Commission, and state-based regulations covering deployment of autonomous vehicles on public roads
- Civil Aviation Safety Authority for autonomous drone use
- Australian Human Rights Commission
- Equal Opportunity Commission
- Law Council of Australia
- CSIRO

Other bodies that may potentially be involved in the regulation of AI in Australia:

- Australian Competition and Consumer Commission (ACCC)
- Standards Australia
- IEEE (ie. IEEE Standards Association and associated certification) (3),(4)
- Engineers Australia, including Systems Engineering Society of Australia
- Australian Computer Society

(b) What are the gaps in the current regulatory system?

The gaps in the current regulatory system are due to the application of new technology within an existing paradigm, current legislation or current governmental departments. It is necessary to take a unified approach to the emerging technology and its application in many domains and industries, looking at both the general and the specific uses, benefits and risks in each domain.

6. If Australia had a Responsible Innovation Organisation:

(a) What should be its overarching vision and core aims?

The vision of the Responsible Innovation Organisation (RIO) must reflect a growth in collective societal wisdom that reflects wise use and control of technology for the good of society.

A RIO could provide a direct regulatory regime for AI systems developed and used in Australia, certifying and monitoring the safety and use of AI. The RIO would build and maintain public trust in the responsible development and use of AI technologies, and would reflect the vision and core aims of the society that it serves.

(b) What powers and functions should it have?

The Responsible Innovation Organisation (RIO) could comprise a policy making arm as well as a certification arm that would:

- Develop and refine a regulatory framework for responsible AI innovation, which includes incorporation of transparent ethical and human-rights design based criteria and risk management processes, and processes for monitoring of these criteria and risks during and after deployment
- Provide a certification process for safety of AI systems based on the regulatory framework
- Provide advice to developers, investors, purchasers and other stakeholders on the regulatory framework and certification process
- Permit AI developers to test their designs against the framework and certification criteria within secure environments
- Provide safety certification of AI systems that have demonstrated commitment to the criteria and risk management processes
- Monitor outcomes of certifications for compliance to ethical standards and consequential societal effects

- Monitor outcomes of companies not successful in certification for potential liabilities under the law, should these products be brought to market within Australia
- Regulate Government sponsorship of AI research and development to those whose aim is to achieve certification and work within the regulatory framework
- Coordinate and advise various regulatory government bodies and NGOs for consistency of regulations, certifications and use of AI systems within the domains and industries that they are responsible for
- Education and social outreach amongst the public and private sectors on regulatory frameworks, certification, societal and individual rights and advocacy methods, and
- Education and outreach to the technical, scientific and engineering community of the importance of ethical and human rights criteria within design and management of AI systems.

One concern is that there is a paradox of the potential of AI systems doing things that are beyond the control or anticipation of their original designers, in unplanned or unanticipated ways. Therefore, it is necessary for certification to incorporate continuous monitoring for unanticipated use or outcomes, and this be incorporated into the risk management process.

Use of the term 'ethical' here represents the normative ethical theories of

- Utilitarianism, or consequential ethics, which is based on outcomes and duty,
- Virtue ethics which is based on a person's character, and
- Deontological ethics, based on obligations and duties.

The British Academy and Royal Society recommended objective of 'human flourishing' (1) is based on 'goodness and quality criteria' of a 'Participatory' worldview or paradigm (2). This means a self-reflexive practical knowing of how to choose and act to enhance personal and social fulfillment between the balance of

- Deciding for others (hierarchy),
- Deciding with others (cooperation, collaboration), and
- Deciding for oneself (autonomy).

This is based on participation through a culture of shared values, norms and beliefs, and an agreement about the rules of language and mutual experiential knowing and understanding between people about what is worthwhile.

The key issue is that the technical community of AI developers and system engineering generally subscribes to a 'Positivism' worldview of prediction and

control without a means for self-reflexiveness, as an awareness of their own paradigm, or an awareness of other paradigms or ways of knowing the world. Their worldview does not include human values within its measurement criteria, and only includes 'goodness and quality criteria' such as reliability and objectivity. Hence, there is an important role for the RIO of education and outreach to the technical, scientific and engineering communities in order to shift their worldview from 'Positivism' to 'Participatory' and human flourishing, so that they can understand and incorporate appropriate human values and ethics as 'goodness and quality criteria'.

An example of a shift of paradigm is seen within the 'IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems' (3), where ethics and human values have been incorporated within an 'Ethically Aligned Design', and standard wellbeing metrics relating to human factors directly affected by AI/AS are being developed.

(e) How should it interact with other bodies with similar responsibilities?

The RIO could coordinate various Australian regulatory government bodies and NGOs for consistency of regulations, certifications and use of AI systems within the domains and industries that they are responsible for. In some cases, certification and monitoring could be delegated to the responsible authority (Australian or state-based), in accordance with the criteria of the developed policies.

For example, delegation to National Transportation Commission for specific certification of autonomous vehicles on public roads, or to the Civil Aviation Safety Authority for autonomous drone use in public airspace. These certifications however should include and be within the transparent ethical and human-rights based framework developed and maintained by the RIO, and not just based on functional design criteria and certification (ie. that the technology 'works').

The management and administration of overall safety certification could be delegated to an NGO or non-aligned body that is independent of any technology vendor or development house (similarly to the International Standards Organisation, or the IEEE). International standards associations such as the ISO and IEEE play an important role to ensure harmonisation and congruency with international standards and expectations, and to ensure that Australia is implementing the latest methodologies and frameworks.

Education and outreach to the technical, scientific and engineering community of the importance of ethical and human rights criteria within design and management of AI systems could be accomplished through coordination with

the various engineering and scientific bodies, such as Engineers Australia, the IEEE, Australian Computer Society, CSIRO, etc

References

1. Australian Human Rights Commission and World Economic Forum, 2019, *Artificial Intelligence: governance and leadership White paper*, January 2019, p.15
2. Heron, J & Reason, P 1997, 'A Participatory Inquiry Paradigm', *Qualitative Inquiry*, Sage Publications Inc. vol. 3, no 3 pp. 274-294
3. IEEE Standards Association 2019, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>
4. IEEE Standards Association, 2019, *The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)*, <https://standards.ieee.org/industry-connections/ecpais.html>