

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?

Technical regulation should be a first priority. This would involve drawing on engineering industry models of safety regulation – such as aviation. The primary focus should be on the acquisition, use, manipulation and deployment of data. Most of the human rights issues outlined in the White Paper relate to the collection and interpretation of data. These concerns should be addressed on a technical level rather than at a conceptual level where definitions are inherently contested.

An example of how to do this is in the German government's approach to self-driving cars. They have mandated that in the event of an impending collision, all lives involved must be weighted equally in an AI agent's decision on how to

manage the collision. Any new organisation should be focused on providing technical guidance and ensuring that any AI agent deployed in a real-world scenario conforms to the organisation's technical guidance and specifications.

Such an organisation would need powers to investigate AI deployment failures in the same manner that aviation authorities currently investigate critical failures around the world.

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

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

When it comes to influence, the research organisations and companies creating, maintaining and deploying AI in real-world scenarios have the greatest influence. The tax office, with respect to the R&D tax offset therefore has a large influence in this area. Adjustments to the R&D tax offset could ensure AI dual-use research of concern (DURC) is properly monitored at both public and private research organisations and companies.

When it comes to the export of Australian-developed AI, there is a role for Department of Foreign Affairs and Trade and the Department of Defence to play in ensuring that DURC AI is not sold to foreign actors that might intentionally misuse the technology, or re-deploy it in ways that infringe the human rights (according to international law) of impacted populations.

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

The primary gaps exist within AI deployment. AI agents that operate in real-world scenarios are not registered nor monitored. If deployed agents are acting on data sets and making decisions that could be seen to have significant impacts for human, animal or plant populations, they should be monitored at a national level. As with regulation of the Australian chemical industry (which in part stems from national alignment with the Chemical Weapons Convention), critical failures are required to be reported. Moreover, the location and manufacture of chemicals on the CWC schedule are known and can be randomly inspected at any time. Such a process should also exist with deployed AI agents. This could work by placing certain AI capabilities on a schedule that would have follow-on reporting obligations. Not only would this increase transparency around AI areas of concern, it would allow for national oversight of AI DURC deployments.

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

Given how low Australian investment in AI technology is right now, there is a risk that if an RIO was implemented incorrectly, it would further hamstring this investment, causing some domestic AI work to migrate to other regulatory environments. If done in the right way, with an envelope of funding available for developing a national AI research roadmap, this could be mitigated. To have substantive positive economic value would require the RIO to be a net contributor to AI research, rather than purely a policy and regulatory organisation.

While there is great social value in establishing an RIO, the extent of that value predominantly depends on how the RIO develops guidance notes to wicked policy problems. If there is not widespread community involvement in the ongoing work of an RIO, it would risk becoming technocratically divorced from community concerns. This means that any RIO would need to undertake a large program of science communication in order to ensure informed citizen input and citizen consent to the regulatory decisions of the RIO. Without such a program of community engagement the RIO would risk losing legitimacy and potentially damaging the emergent Australian AI industry.

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

A well-designed RIO would enable Remi AI to provide technical input into the creation of AI regulatory guidance. It would provide certainty for the Australian AI industry and this would provide stable ground on which to make long-term strategic decisions at Remi AI.

A poorly-designed RIO could create a significant regulatory burden for Remi AI and raise the cost of doing business in Australia. This would have flow-on impacts for the size and scale of emerging businesses that could comply with any new regulation. If the regulatory burden is too high, it will lock out start-up businesses and entrench AI R&D within a well-established cohort of mature companies. This will have flow-on impacts for AI competition in the Australian market and private Australian investment in AI R&D.

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

The business case should not be measured in a timeframe of less than a decade. The return on investment for this kind of an organisation is mostly going to be

intangible and will take generations to mature. No business case will do justice to such a proposition. Most of the impact a future RIO may have will be in negating accidental and intentional misuse of AI technologies. By its very nature, if the RIO is successful, its success will be impossible to measure.

Once an RIO was established, however, it would be possible to measure its success by certain quantitative and qualitative indicators. For example, the number of registered AI agents deployed in real-world scenarios, the number of AI DURC permits issued for firms to export AI solutions, the number of schedule 1 AI capabilities currently under development in Australia. All of these indicators could be accompanied by regular horizon scanning exercises to which community consultation could be integrated. Once an RIO existed, it would be able to justify its existence, and like all organisations, be institutionally motivated by a need for self-perpetuation.

6. If Australia had a Responsible Innovation Organisation:

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

An RIO (AI focused) should predominantly be occupied with the technical monitoring and regulation of real-world AI deployments, to ensure such deployments reflect and are endorsed by informed community consent.

One of the key issues with the idea of an RIO is that it is inherently broad. If the organisation is to be piloted, then it should be limited in scope in order for its success to be measurable in the short-term. This means ensuring that the organisation is ONLY focused on AI real-world applications, and not emerging technology as a generality.

(b) What powers and functions should it have?

It should have the power to require registration of certain AI agents in order for those agents to be deployed in real-world scenarios. It should have the function of monitoring those deployments, this should involve consulting with communities of practice and communities of citizens to ensure those deployments align with the evolving values and concerns of all.

It could have the power and function to sponsor a long-term research roadmap for Australian AI. Such a power and function would necessarily be proactive rather than reactive. It would require commensurably more resources than a reactive approach.

(c) How should it be structured?

The RIO should be completely independent of government and a legislatively established body that cannot be dismantled without consent of the parliament. It should operate in a way that draws from the Civil Aviation

Authority (drawing on aircraft certification and design processes) and from the Australian National Audit Office. This would enable it to have independent commonwealth powers to audit government, corporate and non-government AI accident or misuse.

(d) What internal and external expertise should it have at its disposal?

Initially the RIO should maintain a lean structure and rely on external expertise. This will allow the RIO time to map what its key personnel requirements are, and who best in the country are suited for those positions. The RIO should draw on an AI practitioner community, and could copy the methods used by the FBI and the Alfred P. Sloan Foundation to create a DURC-focused synthetic biology practitioner community in the USA.

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

The RIO should be part of an inter-governmental working group on AI research and deployment. Such a body could have its secretariat embedded within the RIO.

(f) How should its activities be resourced? Would it be jointly funded by government and industry? How would its independence be secured?

The RIO should be funded through the establishment of an AI Future Fund, to which initial Commonwealth and State Government contributions would be required. The AI Future Fund could be periodically topped up by levies on advanced technology industries operating in Australia with revenue above \$200 million.

(g) How should it be evaluated and monitored? How should it report its activities?

The RIO should annually table a report in the Australian Parliament and be subject to normal Commonwealth oversight to ensure it meets its strategic priorities. However, in order for the RIO to maintain legitimacy, it should model organisational transparency and use AI applications in doing this. The RIO should have a live feed of its activities and findings accessible online, curated by an AI. All of the RIO's research, data and work should be openly accessible through commonly used code libraries and data repositories. The only exception to this should be AI DURC code. That said, the RIO should have an open and transparent DURC Committee that considers whether new AI agents should be deployed and what controls should exist on such deployments. A register of DURC findings should be available so that AI R&D firms can identify Committee Precedents and strategically allocate their R&D resources appropriately.