Regulatory Institute: Safe and responsible AI in Australia

The Regulatory Institute is a non-profit think tank that aims to improve regulation globally so that regulations benefit us all. We do this through research into good lawmaking and regulatory techniques, and pro bono consultancy to governments, legislatures and NGOs. We published the Model Law on Artificial Intelligence (the Model Law), which forms the basis of our comments in this consultation by the Australian Department of Industry, Science and Resources.

We have also referenced our Handbook "How to regulate?", which covers theoretical, methodological, and applied aspects of legal regulation and lawmaking, concentrating on the best practices of more than 40 countries/jurisdictions around the world. Noting the importance of conformity infrastructure for regulating AI we urge you read the two chapters that cover this topic.

The Model Law provides a relatively complete basic pattern for the development of laws or regulation for the control of AI systems. The Model Law does not contain detailed technical provisions, which facilitates its use in all jurisdictions regardless of their resources or capacities. It should be used as a toolbox, a checklist or the basis for the development of an adapted law, and optimised as such. The Model Law is not intended to be used exactly as it is drafted. It points to important decisions to be taken by regulatory practitioners without preempting respective choices. Provisions in the Model Law often present choices, be they alternatives or add-on modules, that can be kept or deleted.

Given the importance of AI regulation, the Regulatory Institute is pleased to present its response to the call for consultation laid down in the June 2023 "Safe and responsible AI in Australia" Discussion paper (the Paper).

Consultation questions

Definitions

1. Do you agree with the definitions in this discussion paper? If not, what definitions do you prefer and why?

The definitions in the Paper look solid and self-sufficient taking into account the aim and scope of the Paper. However, legal regulation, for the sake of both flexibility and formality, is likely to need a more elaborated system of definitions.

The Model Law provides a system of definitions with some optional elements (Section 2). A part of those definitions are directly related to the AI itself. While the others (eg. "users", "clients", "traders") are related to the patterns of legal regulation embodied in the Model Law.

The Model Law's var. 2 of AI definition may be used, in particular, to develop a definition encompassing non-AI ADMs, which can be useful if one decides to regulate those in the same way as AI as mentioned on page 5 of the Paper.

Another of the Model Law's legal techniques worthy of attention is explicitly establishing a list of products and/or products' features which qualifies the product's inclusion as an AI (or as a matter falling under AI regulation) even if it does not meet all the criteria of the general definition of AI.

Potential gaps in approaches

2. What potential risks from AI are not covered by Australia's existing regulatory approaches? Do you have suggestions for possible regulatory action to mitigate these risks?

Some claim that there is a risk of AI systems taking over control over humankind. The Regulatory Institute has no competence to assess whether such risk really exists. However, given that some experts assert such a risk, it would be preferable to address it, following the precautionary principle. If AI systems with such potential were not to be banned, they would merit continuous surveillance by a supervising authority, including a state agent installed at the place of business of the respective operator and permanently having access to all documents. Such special agents have been established in the USA. system for the control of certain companies which have massively infringed the law. They operate also outside the USA.

We also recommend that AI-related environmental (ecological) challenges be considered. The Model Law suggests utilising potential environmental impact in the risk-management system (Section 3) and obliging developers and operators to assess potential environmental (ecological) risks (Section 9).

There are a range of smaller risks covered in the Model Law, but not yet covered by Australia's existing regulatory approaches; see our final remarks at the end and the Model Law. See also our recently published Model Law on cross-border internet activities and virtual worlds which contains additional aspects, particularly the interaction between different actors.

3. Are there any further non-regulatory initiatives the Australian Government could implement to support responsible AI practices in Australia? Please describe these and their benefits or impacts.

In the Handbook "How to regulate?", there is a wide range of implementation measures that do not necessarily need a legislative power, but are still beneficial as non-regulatory initiatives. See in particular:

- Chapter 4, specifically Section 4.2.2 about regulatory measures other than regulation;
- Chapter 12, particularly Sections 12.4 to 12.9.

We also recommend the following implementation measures either to be laid down in the law or to be outlined via other means or channels, such as budgetary and administrative measures:

- Minimum resources and minimum control intensity requirements for supervising authorities, as suggested in some of our model laws;
- Alert portals where competitors and employees can drop documents and information pinpointing infringements, if so wished anonymously or with whistle blower protection;
- AI to be used by authorities to check compliance of AI systems ("AI battling AI");
- Playground like test environments for AI systems, either voluntarily or mandatorily to be used;
- Online compliance test for AI systems, either voluntarily or mandatorily to be used;
- Code of conduct, merging the various codes of conduct and AI safety papers currently popping up;
- Voluntary peer review on the application of legal requirements and the code of conduct;

- Voluntary private certification on the application of legal requirements and the code of conduct;
- Voluntary state quality mark based on the state's assessment of the application of legal requirements and of the code of conduct;
- Administrative quality rating by the supervising authority of ai systems, in view of various criteria relevant for users (in some jurisdictions, such rating systems require a legal basis; see the Australia and New Zealand School of Government conference of 31 July 2023 might give you insight into this particular topic);

4. Do you have suggestions on coordination of AI governance across government? Please outline the goals that any coordination mechanisms could achieve and how they could influence the development and uptake of AI in Australia.

The topic of harmonised interpretation between different federal departments or state administration in charge of implementation is relevant for Australia as it is in the EU. The topic has been addressed in the previous question.

For the coordination of positions within the federal government, recommendations could only be developed through a deeper dive into the Australian government system which we are unable to do as part of these comments. Generally speaking, we observe two approaches chosen by governments:

- Several ministries or institutions operate at the same level, but one of them is in a lead / coordinating role; where conflicts cannot be resolved, the issue is moved to the level of the ministers, or cabinet.
- Several ministries or institutions are coordinated by an additional entity which might even have the right to decide in case of conflict; the possibilities for ministries to bring the issue up to the level of the ministers might be limited or not.

As AI policy might require quick decisions, the second (*inter alia* French) model might seem preferable.

Responses suitable for Australia

5. Are there any governance measures being taken or considered by other countries (including any not discussed in this paper) that are relevant, adaptable and desirable for Australia?

For now, the EU AI Act seems to be the most solid and consistent regulatory response to AI issues emerging. With that said, we are eager to share the Regulatory Institute response on the EU AI Act draft we had submitted to the EU authorities (see separate attachment). The Regulatory Institute sees important possibilities for improvement even for the EU AI Act, and this not only for the Large Language Models which were not appropriately covered by the initial proposal. We understand that in the final negotiations (so-called Trilogue) between the European Parliament, Council and Commission, LLMs will play an important role. For this reason it is worthwhile following the development in the EU.

Target areas

6. Should different approaches apply to public and private sector use of AI technologies? If so, how should the approaches differ?

We believe that public-private differentiation is a suitable method for regulating AI. Such an approach is common in those jurisdictions that regulate AI and specifically exclude military use of AI from the scope of general regulation.

The other ways this approach can be implemented are:

1. restricting development and/or distribution and/or usage of certain AI technologies to public bodies only; and
2. laying down extra requirements on AI products (technologies) that are designed to be used or sought to be used for public administration or governance, for example at Section 44 of the Model Law we recommend policy consideration about offering a right to refuse the processing by AI systems.

7. How can the Australian Government further support responsible AI practices in its own agencies?

Private and public organisations alike are obliged to follow the law. In our Model Law on AI the obligations of developers, operators and users of AI are clearly outlined, including a standard for management systems and training requirements of people involved. We believe a regulation with clear requirements and obligations is a good way to support responsible AI practices. We appreciate that with the recent findings of the Royal Commission into Robodebt that trust in responsible AI practices in public administration might be low at the moment and so we highlight the importance of good whistleblowing provisions as well as alert portals.

8. In what circumstances are generic solutions to the risks of AI most valuable? And in what circumstances are technology-specific solutions better? Please provide some examples.

The value of law and legal regulation is that legal rules tend to have a reasonable degree of generalisation and abstraction. Thus legal rules have a great potential to cover new emerging relationships without the need to amend existing regulation.

Considering that, it is rational to evaluate if generic solutions are consistent or they need to be backed or replaced with AI-specific regulations. Excessive use of technology-specific solutions could make for a complex legal system that quickly risks becoming disorganised. For example we suggest that licensing of AI-related activities could be ruled by general licensing legislation (in terms of procedure, general framework etc); and the very licence requirements should be technology-specific.

The same goes for other traditional legal institutions that have developed decent levels of generalisation and adaptability, like property law, legal liability, consumer protection etc. Thus we believe that using generic regulation backed with AI-specific rules if needed should be favoured where possible. Our Handbook "How to regulate?" covers inter alia the issues of generic and sector-specific regulation balance as well as other issues of regulation architecture (Chapter 2).

9. Given the importance of transparency across the AI lifecycle, please share your thoughts on:

a. where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI?

b. mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.

a. Ensuring transparency is important and valuable at any stage of the AI product life cycle. The focus should be on identifying and strengthening the weak spots in the cycle. We consider the period when an AI system is already in legitimate use, satisfying the regulatory requirements at the development stage and put on the market and/or according

to legal requirements, to be such a potential weak spot. Once the basic permissions are granted, and the basic requirements are met, the regulatory influence and oversight tends to weaken. At this specific stage of the AI product lifespan further development and updates of the product could potentially fail to comply with transparency requirements. Thus, we assume that mandating transparency requirements should be equally strong and demanding throughout the entire AI product lifespan, while paying attention to the specified period of active use. See Section 11 of the Model Law which outlines a set of circumstances where mandatory transparency should be required.

b. Legal mechanisms to be used to mandate transparency requirements could include inter alia: public and civil control, self-control, periodical assessments (self-assessments included), requiring whistleblower mechanisms, alert portals providing whistle-blower protection and permitting anonymous submission of documents or information etc.

10. Do you have suggestions for:
a. Whether any high-risk AI applications or technologies should be banned completely?

Our Model Law (Section 13) recommends the total ban on the development, operation and use of AI systems for the following purposes:

- [Full] societal control;
- Social scoring of individuals [trespassing a concrete context such as behaviour on a trading platform];
- Political profiling and repression;
- Manipulation of democratic elections and political processes;
- Interrupting public services;
- Causing damage to third parties;
- Exploitation of psychological or physical weaknesses or vulnerabilities;
- Manipulation of opinions and preferences using erroneous information;
- Creating psychological dependencies;
- Steering and dissemination of internationally banned arms; and
- Generating “deep fake”.

It might be necessary to assess whether there is a risk of AI systems emerging which could take over control over mankind, in which case a ban and further precautionary measures would be needed. As stated above the Regulatory Institute has no expertise with regard to the question whether such a risk exists, but some experts claim that the precautionary principle calls for a ban and further precautionary measures.

b. Criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts?

We assume that it is reasonable to use the criteria established for risk management purposes and classifying AI products into risk categories.

In general, AI systems should be banned if they jeopardise democracy and state political systems, environment, human life and health, basic human rights. Naturally this should be accompanied by case-specific risk assessment to evaluate the following: probability of harm, gravity of potential harm, (ir)reversibility of consequences, potential risk mitigation, resources to be utilised to reverse consequences etc.

Considering this point 10. you may also be interested in our comments on respective Article 5 of the EU AI Act draft which are attached.

11. What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?

Here are some ways to encourage AI usage:

- providing sound, consistent and transparent regulation and governance in the field;
- providing open access (online included) to the information about exercising that regulation and governance, about approved AI systems (e.g. public registers of approved AI systems);
- encouraging (and in some cases mandating) financial insurance in the field (first of all, financial insurance of liability for damage inflicted by AI);
- mandating transparency of AI systems.

Implications and infrastructure

12. How would banning high-risk activities (like social scoring or facial recognition technology in certain circumstances) impact Australia's tech sector and our trade and exports with other countries?

Any bans on certain economic activities, at the first glance, depress economic activities. But actually reasonable bans pay off. Australia, as any other jurisdiction, needs those bans to secure rule of law, law and order, human rights and freedoms, public order, democratic political system, national security. Moreover, these bans are to be one of the pillars of globalised AI regulation (the emergence of which is inevitable), and following general bans is crucial to be integrated in that globalised regulation.

13. What changes (if any) to Australian conformity infrastructure might be required to support assurance processes to mitigate against potential AI risks?

We are not familiar with Australia's conformity infrastructure to be able to make specific recommendations here. However, our Handbook: "How to regulate?" deals with conformity in advance (Chapter 10) and after, together with enforcement (Chapter 11). These two chapters invite the regulator to consider important elements of good conformity mechanisms, be it in advance or after or through the enforcement pathway.

Risk-based approaches

14. Do you support a risk-based approach for addressing potential AI risks? If not, is there a better approach?

We support a risk-based approach. We believe though that AI systems which jeopardise the most basic values, rights and freedoms should be banned regardless of quantified risk-assessment. So a combination of risk-based approach and field-based approach (or goal-based one) is preferable.

15. What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome?

The main benefit of a risk-based approach is its proportionality, fairness and reasonableness. The main limitation is that exercising a risk-based approach could be biased (intentionally or unintentionally), thus failing to provide adequate protection.

This limitation can be overcome by using other approaches alongside the risk-based one.

16. Is a risk-based approach better suited to some sectors, AI applications or organisations than others based on organisation size, AI maturity and resources?

We believe a risk-based approach should be a basic approach accompanied by other approaches which are to assist only. We do not view organisation size, AI maturity or resources as sound indicative criteria. Furthermore we see the stage of AI "maturity" as a potential weak spot as specified above in point 9.

17. What elements should be in a risk-based approach for addressing potential AI risks? Do you support the elements presented in Attachment C?

We support the elements of a draft risk-based approach presented at Attachment C and provide the following specific comments:

a. Impact assessments: our Model Law requires an impact assessment according to specific considerations under each Risk Class. We also provide guidance in the form of ethical rules, to assist the impact assessment. For example at Section 4 Ethical Rules it provides that:

> AI systems shall be developed, operated and used in such a way that the following ethical principles and rules are respected to the extent possible:
> - Where several lives stand against each other, the solution saving the maximum number of lives shall be sought for;
> - The lives of all persons have the same value, in particular regardless of origin and wealth or any of the criteria listed in the definition of "discrimination";
> - The different life expectancy of persons may / may not be taken into account / may only be taken into account where one life stands against another life;

Section 6 concerns how to approach conflict between previous principles.

b. Notices: We agree that users should be informed where automation or AI is used in ways that materially affect them. Section 44 of the Model Law concerns a right to refuse the processing by AI systems in specific areas, we recommend a listing approach. The notices outlined in the Paper are narrow and could be broadened to cover other obligations to inform, which facilitates a system of mutual control. See the Model Law's Section 23:

> Developers shall:
> - Inform operators, also in their commercial contracts, of their respective obligations and the conditions set out in this law;
> - Inform operators, also in their commercial contracts, of ethical problematic aspects mentioned in this law, namely by referring to their own ethics code and respective reports;
> - Keep records of their commercial contacts with operators and inform authorities upon their request; and
> - Inform the supervising authority of infringements they become aware of, regardless whether these are made by competitors, operators, users or conformity assessment bodies.

c. Human in the loop/oversight assessments: Under Section 8 of the Model Law we recommend that Risk Classes 2 and 3 be designed, manufactured and operated in a way that ensures human control of ethical principles as well as parameters and mechanisms of decision-making. This regulatory obligation, and which risk classes it would apply to, will of course depend on a policy decision of the legislators but a balance is required noting that societies fear decisions by AI systems but then full human control takes away the advantage of AI.

d. Explanations: Explainability is an important tool for verifying compliance with obligations. The transparency obligation under Section 11 of the Model Law provides:

> AI systems shall be developed, operated and used in such a way that:
> - Decision-making can be probed, understood and reviewed by authorities, supervisory bodies, common interest third parties, operators, users and their clients;
> - Decisions are explainable [both in technical and non-technical terms], which implies in particular that the processes that extract model parameters from training data and generate labels from testing data can be described and motivated;
> - Inputs and outputs can be verified;
> - Records of design processes, decision-making and other events with external effects or system relevant events are established and kept;
> - The persons steering the processes, decision-making or other operations can be identified, together with the decisions they have been taken during installation or operation of the AI system;
> - Training, validation and testing datasets are accessible; and
> - IT interfaces for full remote authority control (e.g. application programming interfaces) are available and can be operated with commonly available OR freely available software.

Section 18 requires developers and operators to establish, keep up to date and accessible technical documentation which shall, inter alia, include …an explanation of how the ethical principles and others rules set out in Sections 4 and 6 to 14 have been fine-tuned and applied.

e. Training: Training is important for compliance. Section 19 of our Model Law requires that:

> Developers, operators and users with clients shall train their staff (both employees and freelancers) with regard to this law and supplementing decrees, ethics in general and their own ethical code in particular. They shall raise awareness of risks and impacts of the AI systems in question. They shall support their staffs' and freelancers' adherence to professional organisations aiming at the identification and tackling of issues of professional ethics and AI system ethics.

f. Monitoring and documentation: We recommend both a formal monitoring system by the supervising authority particularly for the higher risk classes but also a system of mutual control (Section 23 of the Model Law). The supervising authority could be assisted for the lower risk classes by a third party conformity assessment body (Section 23 of the Model Law).

18. How can an AI risk-based approach be incorporated into existing assessment frameworks (like privacy) or risk management processes to streamline and reduce potential duplication?

No comment.

19. How might a risk-based approach apply to general purpose AI systems, such as large language models (LLMs) or multimodal foundation models (MFMs)?

For general purpose AI systems (eg. LLMs and MFMs) we do not view a risk-based approach to be problematic. Given that a potential use case may fall under the highest category of risk, the regulated assessment should therefore follow the obligations of the highest risk category.

20. Should a risk-based approach for responsible AI be a voluntary or self-regulation tool or be mandated through regulation? And should it apply to:
a. public or private organisations or both?
b. developers or deployers or both?

A risk-based approach is a core element of AI regulation. Thus it should be mandated through regulation. It should be mandatory for both public and private organisations, developers and deployers, as such criteria are not essential when it comes to AI-related risks mitigation.

Final remarks:

There are important aspects of effective regulation not covered by any of the questions above, such as:

- AI platforms enabling the independent development of AI systems ("meta-AI-systems"): these platforms risk to overrun all legal boundaries as the dissemination of AI technology might become uncontrollable, fall into the hands of rogue states or terrorists etc.; we do not know whether such platforms already exist, but we expect them to emerge as we learnt that there are already platforms for the creation of virtual worlds;
- Establishment of basic ethical rules to be followed by AI systems, eg. on trade-offs between different values; see Sections 4 and 6 of our Model Law on Artificial Intelligence;
- Data rights; see Sections 40 to 43 of our Model Law on Artificial Intelligence;
- Accident prevention obligations; see Section 5 of our Model Law on Artificial Intelligence;
- Risk management with precise quantified values; see Section 10 of our Model Law on Artificial Intelligence;
- Obligations on technical documentation and instructions for use; see Section 45 of our Model Law on Artificial Intelligence;
- Prohibition of uncontrolled non-proliferation of AI systems; see Section 14 of our Model Law on Artificial Intelligence;
- Liability and its insurance; see Sections 4 and 6 of our Model Law on Artificial Intelligence.
- Comprehensive empowerments for authorities to act against those who infringe the law and those contributing to infringements or steering from behind infringements, mother and sister companies etc; see Section 48 of our Model Law on Artificial Intelligence and Sections 41 and 45 of our Model Law on cross-border internet activities and virtual worlds.
- Comprehensive empowerments for acting abroad / internationally, also with the help of other states; see Section 44 of our Model Law on cross-border internet activities and virtual worlds.
- Comprehensive empowerments for acting domestically to assist other states when other states wish to enforce their AI law towards actors on the Australian territory ; see Section 56 of our Model Law on Artificial Intelligence and Section 44 of our Model Law on cross-border internet activities and virtual worlds.
- Establishment of systems in which economic actors control each other, refuse to cooperate with infringing actors, are obliged to report to authorities on noted infringements etc; see various Sections of both model laws.
- In case of extremely risky AI systems: obligation of ai system providers to ensure compliance of their trade partners and clients via private law contracts obliging them to respect the AI regulation (law enforcement via a chain of private law contractual

obligations down to the level of the user); possibly to be complemented by an authority licensing mechanism for contracts and business conditions in line with Section 6.f of our Model Law on cross-border internet activities and virtual worlds.

- Comprehensive lists of obligations for all types of natural and legal persons interacting with AI systems, of the style of the lists of obligations of Chapter 2 of our Model Law on cross-border activities and virtual worlds.

See for further complementary aspects our Model Law on Artificial Intelligence, our Model Law on cross-border internet activities and virtual worlds and this innovative product legislation proposal drafted by the European Commission which is partly based on our Handbook "How to regulate?". If you wish a precise analysis of “what is missing” in the “Safe and responsible AI in Australia” Discussion paper, please come back to us. We would be in a position to deliver such analysis in about 10 working days.