Australian Government
Productivity Commission

October 2024

Regulating AI in high-risk settings

Productivity Commission submission

Submission to the Department of Industry, Science and Resources' proposals paper for 'Introducing mandatory guardrails for AI in high-risk settings'

The Productivity Commission acknowledges the Traditional Owners of Country throughout Australia and their continuing connection to land, waters and community. We pay our respects to their Cultures, Country and Elders past and present.

The Productivity Commission

The Productivity Commission is the Australian Government's independent research and advisory body on a range of economic, social and environmental issues affecting the welfare of Australians. Its role, expressed most simply, is to help governments make better policies, in the long term interest of the Australian community.

The Commission's independence is underpinned by an Act of Parliament. Its processes and outputs are open to public scrutiny and are driven by concern for the wellbeing of the community as a whole.

Further information on the Productivity Commission can be obtained from the Commission's website (www.pc.gov.au).

An appropriate reference for this publication is:
Productivity Commission 2024, *Regulating AI in high-risk settings,* Productivity Commission submission to the Department of Industry, Science and Resources' proposals paper for 'Introducing mandatory guardrails for AI in high-risk settings', Canberra.

Publication enquiries:
Phone 03 9653 2244 | email publications@pc.gov.au

Introduction

The Productivity Commission (PC) welcomes the opportunity to make a submission to the Department of Industry, Science and Resources' consultation process on *Safe and responsible AI in Australia: Proposals paper for introducing mandatory guardrails for AI in high-risk settings* (September 2024).

This submission focuses on consultation questions about defining high-risk AI and regulatory options to mandate guardrails.

The PC's previous research on AI in Australia has noted:[1]

- **AI has significant productive potential.** AI will have a substantial impact on productivity and could help to overcome some of Australia's longstanding productivity challenges. While much AI uptake is likely to occur without government intervention, the foundations for digitisation will be important for widespread adoption. Government needs to continue to enable both the rollout of digital infrastructure and uplifting of digital skills.
- **Regulation should enable AI adoption, not stifle it.** The PC has outlined a framework for regulating AI, which focuses on using existing regulation and regulators to manage risks from AI applications wherever possible.
- **Getting data access right will facilitate quality AI use.** The PC has made recommendations to improve data sharing, including extending data sharing arrangements to trusted private entities under the *Data Availability and Transparency Act 2022* (Cth) (the DAT Act) and developing a national strategy for data to facilitate sharing within the public sector, and challenging data excludability in the private sector.

AI regulation that offers the best chance of improved productivity while managing risks will:

- focus on the *net benefit* of regulation – weigh the expected harm from the use of AI with the expected cost of regulating to reduce that expected harm
- be proportionate, effective and risk based – enabling productivity gains from AI use while providing strong safeguards against adverse outcomes
- regulate *outcomes* where possible, rather than using technology-specific approaches that can quickly become outdated
- compare the risks of AI use to *real-world counterfactuals* – not necessarily aim for zero risk
- recognise that Australia will often be an *international regulation-taker* on AI and work with other countries on interoperable and consistent regulatory approaches.

[1] The PC released a series of three papers relating to AI in January 2024 (PC 2024c, 2024d, 2024b). The PC has also examined AI specifically in the healthcare sector in a May 2024 research paper (PC 2024a, chap. 5) and responded to the Senate Select Committee on AI (PC 2024e).

Defining high-risk AI

Consultation questions addressed in this section:

Q. 1 Do the proposed principles adequately capture high-risk AI? Are there any principles we should add or remove?

Q. 3 Do the proposed principles, supported by examples, give enough clarity and certainty on high-risk AI settings and high-risk AI models? Is a more defined approach, with a list of illustrative uses, needed?

If you prefer a list-based approach (similar to the EU and Canada), what use cases should we include? How can this list capture emerging uses of AI?

If you prefer a principles-based approach, what should we address in guidance to give the greatest clarity?

Q. 4 Are there high-risk use cases that government should consider banning in its regulatory response (for example, where there is an unacceptable level of risk)? If so, how should we define these?

Q. 5 Are the proposed principles flexible enough to capture new and emerging forms of high-risk AI, such as general-purpose AI (GPAI)?

Q. 6 Should mandatory guardrails apply to all GPAI models?

In defining high-risk AI, it is crucial to compare risks from AI use with real world counterfactuals, and focus on the net benefit of regulation.

If AI use cases are compared to a situation of *zero* risk, many low-risk use cases will be unintentionally captured under the regulatory regime. It is misleading to measure the risk from a use of AI relative to a fictitious 'perfect world'. Rather, the appropriate benchmark for risk-based regulation is the expected harm from the use of the AI technology relative to the real world counterfactual level of expected harm that would arise if the technology in question was not used. For example, the risk of a self-driving vehicle algorithm should be evaluated against a counterfactual of a competent, licensed human driver, rather than a fictitious world of zero road fatalities. The risk of an AI driven diagnostic tool in health needs to be judged against the alternative of not having such a tool to assist a health practitioner, rather than a false world of perfect diagnosis (PC 2024d, p. 4).

Measuring risk relative to a real world counterfactual avoids harmful regulation that stops technology from improving outcomes. If the counterfactual without the technology entails significant risk, then an AI application can lower risk compared to the counterfactual, even if it does not eliminate risk compared to a fictious 'perfect world'. For example, there are persistent skill gaps in parts of Australia's medical sector, particularly in rural or remote areas. The first-best option may be to fill those gaps with qualified workers over time. However, in the absence of an instant professional workforce, the best alternative could be to employ technologies that can supplement existing expertise (PC 2024d, p. 4).

The assessment of AI risks should also consider non-regulatory measures that can mitigate harms. For business applications, competition between providers and business reputation may mitigate risk adequately. Some applications of AI will create harms that are reversible and compensable, in which case existing laws applying to negligence or consumer safety may be adequate (PC 2024d, p. 4).

A risk-based approach to AI regulation (such as the proposed principles) should weigh the expected harm from the use of the relevant AI with the expected cost of regulating to reduce that expected harm. A risk-based approach to regulation should focus on the expected net benefit of regulation – not on eliminating

a harm. Rather, the aim is to reduce the size and likelihood of harm to acceptable levels without imposing an excessive regulatory burden on society (PC 2024d, p. 4).

For general-purpose AI (GPAI) models, pre-emptive regulations based on hypothetical uses and harms are likely to be ineffective (as harms are unknown) or overly restrictive (costs may outweigh benefits). It is better to wait and see how the technology develops and address any real risks that emerge. As with any new technology, some consequences of AI use will only become apparent as the technology develops further and complementary technologies progress and are taken up. With general purpose technologies in particular, regulation based on 'predicted uses' or 'speculated harms' is likely to be overly broad and limit gains to productivity.

Similarly, prohibitions or bans on general purpose technologies will be generally counterproductive – they may protect against harms but only by also eliminating the benefits. In some circumstances (such as AI being used for applications that are already illegal) better enforcement of existing laws should be considered first. If high risks cannot be appropriately covered by existing laws, a better approach is to define the outcome associated with the risk and see if new, technology neutral regulation is needed.

When it comes to GPAI in particular, idiosyncratic AI-specific rules in Australia would be likely to harm Australia's economy in the long run, unless there is a demonstrated unique need for Australia to depart from international approaches. In general, Australia would do best to be an international regulation-taker in areas such as AI where Australia relies on importing technology or exporting into much larger overseas markets, with active engagement in international forums to design appropriate standards and rules. For example, Australia is a signatory to the first intergovernmental standard on AI – the OECD's Recommendation on Artificial Intelligence – which includes principles for responsible stewardship of trustworthy AI.

Businesses seeking to use AI developed overseas, as well as those intending to sell to international markets, will most likely prosper by meeting recognised international standards without needing to comply with an additional layer of idiosyncratic, local complexity.

In regard to whether a principles-based approach is better than a list-based approach (question 3):

- while a list can provide clarity and certainty, it can also provide false certainty if it is incomplete, or becomes outdated
- a set of principles is a better starting point for assessing the balance of risks and benefits from AI, including analysing the likelihood, severity and extent of potential harms from AI use/misuse
- a set of principles would assist the approach of looking for gaps in existing regulation where risks arising from AI use are not already adequately dealt with
- adopting technology-neutral regulation is more achievable with a principles-based approach.

Nonetheless, list-based examples of high-risk and low-risk AI applications could provide additional guidance to businesses, if regulators are equipped to keep these examples up to date.

Guardrails ensuring testing, transparency and accountability of AI

Consultation questions addressed in this section:

Q. 12 Do you have suggestions for reducing the regulatory burden on small-to-medium sized businesses applying guardrails?

All businesses can benefit from clear regulatory guidance. Outreach programs, examples (such as examples of what compliance looks like) and case studies can help regulators effectively communicate the obligations that small businesses have regarding high-risk AI. Additionally, regularly updated guidance can foster continuous dialogue between regulators and industry, helping both sides stay informed about emerging risks. It is essential that regulators are equipped and resourced to fulfill this role effectively.

Because small businesses face a disproportionate burden of complying with non-interoperable regulatory systems, adopting international regulations on AI may assist small businesses in particular.

Regulatory options to mandate guardrails

Consultation questions addressed in this section:

Q. 13 Which legislative option do you feel will best address the use of AI in high-risk settings? What opportunities should the government take into account in considering each approach?

Q. 14 Are there any additional limitations of options outlined in this section which the Australian Government should consider?

Q. 15 Which regulatory option/s will best ensure that guardrails for high-risk AI can adapt and respond to step-changes in technology?

Regarding which legislative option will ‘best address’ the use of AI in high-risk settings (question 13), it is crucial for government to consider not only which option would effectively protect Australians from adverse impacts of AI, but also which option strikes the best balance of regulatory protection and the cost of regulation. The expected costs associated with regulation – including the potential loss of innovation from overly restrictive rules – should be examined. The longstanding Regulatory Impact Statement process will be essential to determining the best regulatory option.

The PC has previously proposed a stepped approach to regulating heightened or emerging risks from AI (PC 2024d, p. 7):

- Consider if existing regulatory frameworks (including regulations and regulators) adequately address the identified risks, and whether they do so without unduly constraining AI use or presenting inconsistency with equivalent international approaches. If so, there is no need for new regulation. If not:
- Consider if existing regulation can be clarified or amended to bridge any gaps (in regulation or its enforcement) associated with AI development or deployment. If so, clarify or amend existing regulations, and provide appropriate resourcing and training to regulators rather than introducing new regulations. If not:

- Consider the net benefits of new regulation using a risk-based approach. The assessment would need to take into account the relevant outcome(s) and risk(s) to be covered compared to a real-world counterfactual, any non-regulatory counters to the risk, the relevant point(s) in the supply chain where the regulation will apply, and any relevant existing international regulations that may impact the risk or limit regulatory solutions. New regulation should only be introduced if there is a net benefit from the regulation taking these factors into account.

Figure 1 provides a table of the types of issues that decision-makers need to address to apply this approach.

Figure 1 – Regulating AI use

What are the specific uses of the AI model?	• General purpose AI models (e.g. large language models) serve as the foundation of a wide range of applications. • General purpose models are adapted – or narrow models are built from smaller datasets - to address more specific applications and tasks. Use cases span a wide range of occupations and sectors. • Governments should consider how AI is being used, or how it could be used in the immediate future.		
What risks are associated with AI use?	• Risks – which reflect the likelihood, scale and seriousness of harms – become clear in the context of specific use cases. They include economic and financial risks, privacy, discrimination, health and safety, and broader social risks. • The scale, severity, and likelihood of the risks should be assessed relative to real-world counterfactuals.		
Who can influence or control the risk?	• Parties should be accountable where their decisions or behaviours can effectively manage the risk. For instance, users have little control over the quality of the technology, which is influenced by developers, but strongly influence where the technology is deployed, and whether appropriate safeguards are in place.		
What existing protections apply?	Australian legal framework • Consumer protections • Privacy law • Anti-discrimination law • Liability and negligence • Contract law • Sector-specific regulatory frameworks	International legal framework • International regulations may be built into imported models • Technical and other standards	Industry self-regulation • Voluntary codes of ethics and practice • Regulation of technical environments by larger players • Competitive pressures and other incentives to build trust
Where could there be gaps?	• Consider whether existing regulations are adequate. • The nature of AI may challenge or prevent effective enforcement. • New forms of automation may displace existing regulation that focused on the operator or worker (although liability and negligence laws apply). • New forms of harm could be introduced. • The use of AI could highlight known flaws in existing frameworks.		
Can existing regulation bridge the gaps?	• Consider how regulation could be clarified or amended. • Leveraging existing frameworks, processes, or institutions. • Strengthening regulators' capabilities.		
What new regulation might be needed?	• Consider the net benefits of new regulation compared to real-world counterfactuals. • New regulation should be technology neutral where possible. • Regulatory approaches diverge globally. Australia should minimise additional requirements to those in major economies and actively participate in global fora.		

Applying the PC's framework to the discussion paper's three options (a domain specific approach; a framework approach; or a whole-of-economy approach) rules out option 3, introducing a new cross-economy AI-specific Act.

Option 3 may create inconsistencies and overlap among different regulators and their regulations. The scope of a new AI Act, including its application in specific circumstances, would ultimately depend on judicial interpretation. It is unlikely that a legal application decided within one domain, such as competition law, would seamlessly translate to other domains such as consumer protection, healthcare, or corporate law. Drafting comprehensive legislation that avoids such ambiguities would be highly challenging and unlikely to be timely.

Option 3 would increase the complexity and the regulatory burden on AI developers, deployers and users. In contrast, options 1 and 2 provide a degree of certainty for businesses by amending existing regulation. Options 1 and 2 are also more likely to build consumer trust in AI as users see they are already protected by existing laws.

There are advantages to the domain specific approach (option 1) identified in the paper (DISR 2024, p. 47), including minimising disruption to business, limiting regulatory duplication and associated compliance burden, enabling an incremental approach to new regulation, and allowing harms to be addressed in their specific contexts.

Options 1 and 2 also ensure that different technologies are held to the same standard regarding their effect on users, consumers, and the public. They ensure that regulation focuses on decisions and actions that cause harm, rather than on a technology itself.

The discussion paper also acknowledges some limitations of option 1, such as the potential for inconsistencies, and a slower pace of reform due to legislative processes (DISR 2024, p. 47). However, any new legislation (such as option 3) would also be subject to legal interpretations and precedent over time. The uncertainty created by new regulations, regardless of how 'tightly' they are drafted, will often be greater than the uncertainty around existing rules that have already been tested in court (PC 2024d, p. 8).

And while regulators may indeed 'decide not to prioritise reforms to address AI issues … because of competing regulatory priorities, lack of resources or lack of technical capability' (DISR 2024, p. 47) these decisions are ultimately within governments' control. Governments and regulators, such as those comprising the Digital Platform Regulators Forum, are already actively engaged in reform processes to address AI concerns within their remits (DP-REG 2024). Before any new AI legislation is reached for, an alternative is to adequately resource regulators and enforcement agencies to keep up with AI risks within their existing mandates.

The framework approach (option 2) also has some advantages, such as providing a consistent set of definitions and measures. However, the uniformity benefits may be overstated, as most businesses deploying AI will already be covered by multiple regulations. The regulatory landscape is already complex for business, and guardrails will make it more so. Nevertheless, the framework approach could be considered if it provides more consistency than the domain specific approach and is easier to amend, while still allowing existing laws to give AI users a degree of certainty.

In terms of which regulatory option best ensures guardrails for high-risk AI can adapt and respond to step-changes in technology (question 15), the key consideration should be to regulate *outcomes*, not specific technologies. Technology-specific regulations quickly become obsolete. To the extent that existing regulations already focus on outcomes, options 1 and 2 would provide a more effective foundation than option 3 for addressing AI risks.

References

DISR (Department of Industry, Science and Resources) 2024, *Safe and responsible AI in Australia: Proposals paper for introducing mandatory guardrails for AI in high-risk settings*, Consultation paper, September.

DP-REG (Digital Platform Regulators Forum) 2024, *Examination of Technology: Multimodal Foundation Models*, Working Paper, August.

PC (Productivity Commission) 2024a, *Leveraging digital technology in healthcare*, Research paper, Canberra.

—— 2024b, *Making the most of the AI opportunity: AI raises the stakes for data policy*, Research paper, no. 3, Canberra.

—— 2024c, *Making the most of the AI opportunity: AI uptake, productivity, and the role of government*, Research paper, no. 1, Canberra.

—— 2024d, *Making the most of the AI opportunity: The challenges of regulating AI*, Research paper, no. 2, Canberra.

—— 2024e, *Senate Select Committee on Adopting Artificial Intelligence (AI)*, Productivity Commission submission, Canberra.