Introducing mandatory guardrails for AI in high-risk settings: Proposals paper by the Australian Government Department of Industry, Science and Resources (DISR)

RACGP submission October 2024

Background

The Royal Australian College of General Practitioners (RACGP) has prepared a response to the Australian Government Department of Industry, Science and Resources' (DISR's) proposals paper on mandatory guardrails for AI in high-risk settings, released to the public for consultation in September.

The DISR has previously sought the public's views on safe and responsible AI, and the RACGP provided a response to that July 2023 consultation (available on our website here). With that feedback, the DISR concluded that the current regulatory system is not fit for purpose, and it is now seeking views on proposed guardrails for the use of AI in high-risk settings (of which healthcare is considered one), as well as regulatory options for mandating the guardrails.

The proposals paper can be viewed and downloaded via DISR's consultation hub at https://consult.industry.gov.au/ai-mandatory-guardrails.

The RACGP requested feedback on our draft response to the proposals paper. The DISR has posed 16 questions pertaining to the paper, and has directed respondents to make their submissions through an online platform. The discussion questions are replicated below, together with the RACGP's proposed responses highlighted in yellow. When finalised, the RACGP's response will be submitted via the online platform.

Questions

1. Do the proposed principles adequately capture high-risk AI?

Yes

No

Please identify any:

- low-risk use cases that are unintentionally captured
- categories of uses that should be treated separately, such as uses for defence or national security purposes.

Lower-risk use cases of AI in a general practice healthcare setting might include those applications for appointment scheduling and/or billing as these would not be expected to affect patient health and safety.

2. Do you have any suggestions for how the principles could better capture harms to First Nations people, communities and Country?

Yes

No

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?

YES the principles give enough clarity and certainty

NO a more defined list-based approach is 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?

- Biometrics
- Critical infrastructure
- Education/training
- Employment
- Access to essential public services and products
- Access to essential private services
- Products and services affecting individual and public health and safety
- Law enforcement
- Administration of justice and democratic processes
- Other (please specify)]

N/A

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

Clearly defined use cases.

How can this list capture emerging uses of AI?

This list will need to be periodically reviewed and updated.

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)?

Yes

No

If so, how should we define these?

Consideration should be given to banning the use of AI for diagnostic or therapeutic uses where the model or system has not been trained on medical/pharmacological data.

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

Yes

No

6. Should mandatory guardrails apply to all GPAI models?

[✓ Yes]

No

7. What are suitable indicators for defining GPAI models as high-risk?

For example, is it enough to define GPAI as high-risk against the principles, or should it be based on technical capability such as FLOPS (e.g. 10^25 or 10^26 threshold), advice from a scientific panel, government or other indicators?

[✓ Define high-risk against the principles]

Base on technical capability

[**What technical capability should it be based on?**

- FLOPS
- Advice from a scientific panel
- Advice from government
- Other (please specify)]

[✓ N/A]

8. Do the proposed mandatory guardrails appropriately mitigate the risks of AI used in high-risk settings?

[✓ Yes]

No

[What guardrails should we add or remove?

- Establish, implement and publish an accountability process including governance, internal capability and a strategy for regulatory compliance
- Establish and implement a risk management process to identify and mitigate risks
- Protect AI systems, and implement data governance measures to manage data quality and provenance
- Test AI models and systems to evaluate model performance and monitor the system once deployed
- Enable human control or intervention in an AI system to achieve meaningful human oversight
- Inform end-users regarding AI-enabled decisions, interactions with AI and AI-generated content
- Establish processes for people impacted by AI systems to challenge use or outcomes
- Be transparent with other organisations across the AI supply chain about data, models and systems to help them effectively address risks
- Keep and maintain records to allow third parties to assess compliance with guardrails
- Undertake conformity assessments to demonstrate and certify compliance with the guardrails
- Other (please specify)

9. How can the guardrails incorporate First Nations knowledge and cultural protocols to ensure AI systems are culturally appropriate and preserve Indigenous Cultural and Intellectual Property?

Consultation with Aboriginal and Torres Strait Islander groups will be essential to the development of culturally appropriate AI systems. The RACGP is the only medical college to have an Aboriginal and Torres Strait Islander Faculty with 15,000 members. We recommend further consultation with our Faculty on this matter. Please email Vanessa Harris (Manager, RACGP Aboriginal and Torres Strait Islander Health) for further discussion at aboriginalhealth@racgp.org.au.

10. Do the proposed mandatory guardrails distribute responsibility across the AI supply chain and throughout the AI lifecycle appropriately?

For example, are the requirements assigned to developers and deployers appropriate?

Yes

[✓ No]

Please provide additional comments.

Many general practitioners (GPs) are independent practitioners with limited time and means. They would likely be classed as ‘deployers’, yet they do not have the status of an ‘organisation’ to assist them to meet the requirements of a deployer. This is likely to place an additional burden on an already under-resourced and highly regulated general practice sector.

More generally, it is not made clear in the paper how risk/accountability will be apportioned between ‘developers’ and ‘deployers’ for upholding the guardrails, with the body of the paper often referring to ‘organisations’ where greater specificity is required. While Appendix E provides more context, detail is lacking. In many of the situations described, it should be incumbent upon the developer to ensure the guardrails are met, not the deployer. For example, Guardrail 2 (Establish and implement a risk management process to identify and mitigate risks) puts an onus on the deployer to mitigate risks that they may not know exist as a result of the opacity of the AI system. For Guardrail 5 (Enable human control or intervention in an AI system to achieve meaningful human oversight), assessing the quality of the system’s output might also pose a problem for the deployer given the ‘black box’ nature of these systems.

11. Are the proposed mandatory guardrails sufficient to address the risks of GPAI?

[✓ Yes]

No

[How could we adapt the guardrails for different GPAI models, for example low-risk and high-risk GPAI models?]

N/A

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

[✓ Yes]

No

Please provide additional comments.

As above, the greatest burden of compliance must lie with the developer.

13. Which legislative option do you feel will best address the use of AI in high-risk settings?

A domain specific approach – adapting existing regulatory frameworks to include the proposed mandatory guardrails

A framework approach – Introducing new framework legislation to adapt existing regulatory frameworks across the economy

[✓ A whole of economy approach – introducing a new cross-economy AI Act]

What opportunities should the Government take into account in considering each approach?

Alignment with leading international models.

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

Yes

[✓ No]

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

A domain specific approach – Adopting the guardrails within existing regulatory frameworks as needed

A framework approach – Introducing new framework legislation to adapt existing regulatory frameworks across the economy

[✓ A whole of economy approach – introducing a new cross-economy AI Act]

Other (please specify)

Please provide additional comments.

This option acknowledges the transformative power of AI for many sectors, including healthcare, and the need to future-proof the regulatory approach.

16. Where do you see the greatest risks of gaps and inconsistencies with Australia's existing laws for the development and deployment of AI?

A piecemeal approach that involves amending a suite of existing laws might not be a suitable option given the scale and speed of AI adoption.