AUSTRALASIAN INSTITUTE OF DIGITAL HEALTH

Department of Industry, Science and Resources Consultation on Safe and responsible AI in Australia - Introducing mandatory guardrails for AI in high-risk settings

Submission by the Australasian Institute of Digital Health (AIDH)
October 2024

Via questionnaire on consultation webpage at https://consult.industry.gov.au/ai-mandatory-guardrails

Please note: AIDH provided its feedback via the online questionnaire available on the consultation webpage. For publication, responses have been reformatted, the introduction and About AIDH sections were added.

Anja Nikolic, CEO
Australasian Institute of Digital Health (AIDH)
Level 1, 85 Buckhurst Street
South Melbourne VIC 3205
+61(3) 9326 3311, policy@digitalhealth.org.au
ABN 80 097 598 742

Contents

Introduction

The Australasian Institute of Digital Health (AIDH) welcomes the opportunity to provide feedback to the Department of Industry, Science and Resources (the Department) consultation on Proposals paper for introducing mandatory guardrails for Artificial Intelligence (AI) in high-risk settings.

AIDH notes that for a risk-based approach to capture severity and extent of the adverse impact of AI in high-risk settings, it needs to give special consideration to risks where impacts cannot be repaired nor compensated. As mentioned in the Government interim response to the Safe and responsible AI in Australia consultation, harms caused by AI in healthcare settings can be difficult or impossible to reverse, nor can they be appropriately compensated.

AIDH unequivocally supports the introduction of mandatory guardrails and makes recommendations to strengthen the ones proposed by the Department. Particularly, we recommend strengthening the wording of guardrail #5. This is to ensure human decision making, supervision, control and accountability is embedded in high-risk settings.

AIDH proposes to adopt a whole of economy approach that is supported by health sector-specific regulations which could take into consideration the specific needs of healthcare delivery while ensuring robust consistency across sectors.

AIDH acknowledges and thanks the members who have contributed their time and expertise to this submission.

AIDH welcomes further engagement with the Department on any topic explored in this submission.

Consultation questions

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

No

• Are there any principles we should add or remove?

We believe the absence of a human rights charter or act in Australia weakens the intent of principle a. Reference to Australia's international human rights law obligations is not enough of a guardrail as these are not easily enforced.

Beyond human rights, we recommend that there need to be given regard to the risk of adverse impacts to all rights of a person as recognised in Australian law (eg. Disability Act, Privacy Act, all the Discrimination Acts, etc.).

We recommend adding 'democracy' or 'democratic principles' in the wording of principle e.

Principle f regarding severity and extent of the adverse impact of high-risk AI fails to capture risks where impacts cannot be repaired nor compensated. As mentioned in the Government interim response to the Safe and responsible AI in Australia consultation, harms caused by AI in health setting can be difficult or impossible to reverse.

Please identify any:

• low-risk use cases that are unintentionally captured

Our position is that if a use case is likely to contradict any of the principles, therefore, by definition, it cannot be classified as low-risk.

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

Yes

Yes, engage Indigenous experts and knowledge holders for meaningful collaboration. We believe that reflecting their perspectives can effectively address the potential harms of AI to First Nations peoples and ensure that their rights, cultures, and environments are respected and upheld in the deployment of AI technologies.

We strongly believe that long term partnership with First Nations communities such as Centre of Excellence for Aboriginal Digital in Health (CEADH), National Aboriginal Community Controlled Health Organisation (NACCHO), Lowitja Institute, and Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) is required for fostering ongoing dialogue and collaboration in defining appropriate principles.

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?

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?

The proposed principles, supported by examples, provide a solid framework for addressing high-risk AI systems and models. However, for a more defined approach, we propose including a list of illustrative uses capturing emerging AI.

This could enhance clarity and certainty for stakeholders. By identifying high-risk applications such as emerging AI in healthcare decision making, regulators could help organisations better understand the requirement and risk associated with these technologies. We believe that concrete examples and clear guidelines would promote culture of responsibility among developers and users.

We recommend including the uses cases already identified by the EU and Canada in their respective legislations. Continuous monitoring of emerging technologies needs to be implemented, and the list updated accordingly.

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

AIDH has identified high-risk cases that should be considered. These include the use of AI to create avatars based on images and voice depicting or replicating real people without their consent. The use of what can be considered as 'deep fakes' in clinical settings particularly in any situation where mental health information, advice or treatment is involved, presents an unacceptable level of risk. Digital counterfeits can deceive consumers and cause them to act on unverified and misleading health information. AIDH is aware of 'fabricated celebrity endorsement' in the USA which exploits health consumers and undermines trust and safety in healthcare.

Related unacceptable high risks to consider banning are AI chatbots purporting to provide medical advice or clinical mental health support; and AI algorithms that autonomously diagnose diseases from medical imaging or other data.

Consider banning AI systems that use predictive analytics for patient risk profiling without proper safeguards. These systems can lead to discriminatory practices where specific populations may be unfairly flagged as higher risk, resulting in inadequate care or increased premiums.

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

Yes

6. Should mandatory guardrails apply to all GPAI models?

Yes

By definition, GPAI can be used in any context. We know that GPAI is used in healthcare settings even though those models were not designed to be used in that context, therefore guardrails should apply to all GPAI models. Indeed, those models are trained on non-specific health data and therefore are more prone to errors when confronted with a health specific assignment. This was for example documented in the research paper Using ChatGPT-4 to Create Structured Medical Notes From Audio Recordings of Physician-Patient Encounters: Comparative Study (doi: 10.2196/54419, https://www.jmir.org/2024/1/e54419). Therefore, further sub-sets of guardrails will be needed in healthcare settings especially as it relates to accountability.

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

Our position is that the risk associated with AI are more related to the potential applications of these models rather than their computational abilities. Therefore, basing indicators on technical capabilities would fail to capture risks in specific settings. For example, in the healthcare sector, potential harm won't be proportionate to technical capability.

Additionally, technical capabilities are evolving so quickly and, in some cases, in ways we cannot predict, it is likely that any legislation / regulation based on technical capabilities would fail to provide long-term or ongoing safeguards.

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

No

Although the guardrails cover all relevant aspects of risk mitigation, some can be strengthened.

We recommend amending some of the guardrails as follows:

Guardrail 2. Establish and implement a risk management process to identify and mitigate risks

We recommend rewording the guardrail to read “Establish, implement and maintain a continuous risk management process to identify and mitigate risks as they emerge”.

Guardrail 4. Test AI models and systems to evaluate model performance and monitor the system once deployed

We recommend rewording the guardrail to read “Test AI models and systems to evaluate model performance and monitor the system continuously once deployed to ensure adherence to expected outcomes”.

Guardrail 5. Enable human control or intervention in an AI system to achieve meaningful human oversight

Our position is the wording of guardrail #5 is too weak to ensure human decision making, supervision, control and accountability in high-risk settings, especially as it relates to healthcare. We recommend amending the guardrail to read: “Adopt a ‘human in the loop’ approach and embed human decision making, control or intervention in an AI system to achieve meaningful human oversight.”

Guardrail 7. Establish processes for people impacted by AI systems to challenge use or outcomes

We note that in case those internal processes fail in healthcare settings, it will be critical to have regulatory bodies in place for consumers to escalate their issues. It can be through existing State complaint bodies, the Australian Commission on Safety and Quality in Health Care (ACSQHC), an AI in Healthcare specific body to be created, alongside Ahpra for regulated professions.

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?

As stated above, it is critical to engage Indigenous experts and knowledge holders for meaningful collaboration.

We strongly believe that long term partnership with First Nations communities such as Centre of Excellence for Aboriginal Digital in Health (CEADH), National Aboriginal Community Controlled Health Organisation (NACCHO), Lowitja Institute, and Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) is required for fostering ongoing dialogue and collaboration in defining appropriate guardrails.

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

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

Yes

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

Yes

Many healthcare providers in primary care operate as small to medium size businesses, such as clinics, general medical practices, or allied health practices. They operate in an already highly regulated environment, it is critical to support them when further regulation is introduced as to not create an unnecessary regulatory burden.

Governments can help by providing clear information, free training / education, and resources to support SMBs including health practices / clinics and software / app vendors.

Also, the Australian Government should increase and continue funding for the AI adopt program initiative with a focus on healthcare settings.

Finally, AIDH endorses the recommendations from the National Policy Roadmap for Artificial Intelligence in Healthcare published by the Australian Alliance for Artificial Intelligence in Healthcare (AAAiH). In relation to small-to-medium size businesses, the roadmap recommends to:

Provide support and incentives for local industry (and SMEs in particular):

a. Consider expanding the R&D Tax incentives scheme to cover regulatory compliance costs.

b. Ensure the pathway to reimbursement for AI-based clinical services via Medical Services Advisory Committee (MSAC) is understood.

c. Consider additional funding to support new products to come to market.

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

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

AIDH sees merit in a combination of both a domain specific and a whole of economy approach. A whole of economy approach that is supported by health sector-specific regulations could take into consideration the specific needs of healthcare delivery

while ensuring robust consistency across sectors. For example, this would apply to clinical governance, decision making, accountability, complaint mechanism, and data privacy.
AIDH supports the recommendations outlined by the AAAiH in the National Policy Roadmap for Artificial Intelligence in Healthcare (https://aihealthalliance.org/wp-content/uploads/2023/11/AAAiH_NationalPolicyRoadmap_FINAL.pdf) on establishing a National AI in healthcare council; the development of minimum safety and quality standards governing AI in healthcare; a code of conduct for the safe, responsible and effective use of AI; and the need to develop profession-specific codes of practice for the responsible use of AI.
We encourage the Department to pay particular attention to Professor Coeira and Professor Magrabi's contributions to the Senate Inquiry into artificial intelligence – July 2024.

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

Yes
In any case, it will be critical for the Australian Government to progress the future legislative work at pace and with bi-partisan support. Australia already risks lagging international best practice which puts Australian residents at risk of harm due to current and evolving misuse of AI. This needs to be an absolute priority.

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

A whole of economy approach – introducing a new cross-economy AI Act
As above, a combination of a whole of economy approach with a subset of healthcare specific guardrails to address issues like accountability and ethics.

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

We believe that the greatest risks of gaps and inconsistencies are around data protection and privacy laws, and in relations to the overseas reach of the guardrails. The guardrails have a developer to deployer to user approach, we are unsure how the guardrails would apply to developers based overseas and how much Australian residents' data will be protected. For example, the EU has General Data Protection Regulation (GDPR) which is known for its extraterritorial reach, applying not only to organisations based in the EU but also to those outside the EU that offer goods or services to, or monitor the behaviour of, EU residents.

About AIDH

The Australasian Institute of Digital Health (AIDH) represents a diverse and growing community of professionals at the intersection of healthcare and technology.

The Institute has more than 250 distinguished Fellows who are experts or pioneers in digital health, and has a growing membership of professionals comprising doctors, health informaticians, nurses, midwives, allied health, other clinicians, administrators, and health technology business leaders.

The Institute provides objective, non-partisan, and independent advice on the use of technology and health informatics to improve consumer outcomes and solve the most pressing challenges facing our healthcare system.

The Institute's unique composition and reach brings together an extraordinary network of Australia's leading digital health experts across the private, public and community sectors to advance our nation's transition to a digital health future.