Safe and responsible AI in Australia

——Suggestions from Dr Yuntian Brian Bai

I have a hopeful vision for AI technology, envisioning it as a catalyst for greater assistance and support to Australian people. In light of this, I offer my personal perspectives and views, which are provided below in response to each question.

Definitions

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

   I agree with the general AI definitions. However, I would suggest listing a few more typical AI applications/products such as

   a) Intelligent robots,

   b) Autonomous vehicles, and

   c) Medical diagnosis.

   These will give people a more comprehensive understanding of AI concepts.

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?

   We should pay more attention to:

   1) AI products' autonomous attacks on humans, such as physical assaults, virus attacks, or attacks arising from malicious instructions.
   2) AI-generated deepfake.
   3) Job Displacement.
   4) Deepfake contents. Around 23 billion pieces of information are produced every day. Eighty percent of the content may be synthesized, exaggerated or deep-faked, overwhelming for anyone to consume, and challenging to distinguish.

   I strongly suggest building and refining a top-level AI guideline (or called complete Australian AI profile and representation).

   It is very important to establish specific top-level AI guidelines in Australia. This entails defining the desired characteristics and qualities that align with intended purpose and Australian values. One good choice is adopting the widely acknowledged values and principles advocated by the United Nations (UN) and other countries as the benchmark for AI's applications in Australia. Other elements include, but are not limited to, the following points:

   1) AI ethics Guidelines, and also establish a dedicated regulatory authority with expertise in AI to oversee and govern AI technologies' ethical use.
   2) Algorithmic impact assessments.
   3) Deepfake detection and labelling.
   4) Data privacy laws.
   5) Cybersecurity standards.
   6) Public awareness campaigns to educate citizens about AI risks, benefits, and the importance of responsible AI use.
   7) International collaboration.

It is critical for AI's success to have a clear profile and representation. For example, while I requested ChatGPT to describe the conflicts between Russia and Ukraine, it explains as "Russia occupying the Crimean region of Ukraine in 2014". This answer represents a particular perspective, although this point of view is not encompassing the voices and viewpoints of all countries and individuals worldwide.

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.

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.

- Develop a comprehensive national AI strategy that sets clear objectives and principles for AI governance across government agencies as an extension of the top-level AI guidelines as descripted in Question 2.
- Create a unified policy framework that addresses AI ethics, data privacy, transparency, and accountability and ensure consistent adherence to ethical AI principles in all government AI initiatives.
- Conduct periodic audits and assessments to ensure compliance with policies and ethical standards and help identify areas for improvement and mitigate risks.

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?

Collaborating with other countries and learning from global experiences will be valuable in crafting a comprehensive and effective AI governance framework across areas such as data protection and privacy laws, algorithmic impact assessments and AI regulation and standards.

Target areas

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

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

Develop comprehensive guidelines and policies specific to AI adoption within government agencies. Also invest in AI education and training programs for government employees. Encourage inter-agency collaboration and knowledge sharing on AI best practices and improved decision-making and resource optimisation.

Implement independent monitoring and auditing processes to evaluate the impact of AI applications.

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.

Generic solutions are valuable for common risks across various AI applications such as ethical AI guidelines (e.g., fairness, transparency, and privacy protection), data governance and regulatory frameworks to different industries or agencies.

Technology-specific solutions are more effective in tackling unique challenges associated with specific AI technologies or domains. For example, bias mitigation and reliability.

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?

Let's look at the following conversations:

A: You know what? Yesterday, while I was driving, I suddenly received a message saying that the road ahead was congested, and I was advised to take an alternate route.

B: That's how AI helps us!

A: AI is so smart! But if we all switched to the other road, wouldn't it also get congested immediately?

B: AI doesn't inform everyone, so only some of us go to the other road, while the rest stick to the original route.

A: But how does AI decide who goes to the other road and who stays?

B: AI makes that decision for you; it just doesn't explicitly reveal it.

A: Doesn't that mean AI takes away our freedom to make our own choices? It's unfair!

This example demonstrates how AI's transparency is important and valuable in mitigating potential AI risks and to improving public trust and confidence. Did AI help us? Yes, it helped us indeed, but at the same time, it made a decision for us without any notification/transparency, and many people opt to place unconditional trust in it due to their increasing reliance day by day. We should always fully respect the public trust.

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

Mandating transparency requirements across the private and public sectors plays a pivotal role in ensuring responsible AI practices and enabling better understanding and scrutiny of AI systems. Transparency can empower individuals to make informed decisions about their interactions with AI-driven technologies. It is crucial to ensure users are aware of AI involvement.

10. Do you have suggestions for:

    a. Whether any high-risk AI applications or technologies should be banned completely?
    No. Only the unacceptable risks should be banned completely. High-risk AI applications should be allowed but with clear and strict regulation and monitoring process.
    b. Criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts?
    Top-level regulation and standards, like a country's constitution, are crucial for AI development and applications. The kind of AI application or technology can only be banned if it clearly violates the top-level standards.
    Then, what is top-level regulation and standards and how to build them, see Question 4.

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

    1) Well-designed top-level regulation and standards (see Questions 2 & 4).
    2) Dissemination and training in AI-related knowledge and technologies.
    3) Embracing both internal and external monitoring.

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?

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

Risk-based approaches

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

Yes. I support to start the implementation from small scale testing and it has to be well evaluated, fully tested to minimise the cost.

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

Benefits: more capable to address critical areas where adverse consequences and mitigate more significant risks effectively.

Limitations: as different people may perceive risks differently, striking the right balance between precaution and innovation can be complex.

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?

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

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?

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

I will focus on one risk-based approach to LLMs in this discussion. Many examples emphasize the importance of conducting a comprehensive risk assessment to identify potential challenges related to LLMs. Such an assessment should consider factors like data biases, misinformation propagation, misuse by malicious actors, and ethical concerns in content generation.

Recently, while researching minimum social welfare for the elderly in a rural area, I observed a significant disparity between the information provided by ChatGPT and the actual situation. The content seemed to reflect a propagandistic tone in line with government agencies rather than accurately portraying reality.

Similarly, when searching for information about the Aboriginal population in Melbourne using ChatGPT, I noticed that the data was neither stable nor accurate. This issue may be attributed to inadequate and incomplete training data.

In response to these examples, it is crucial to adopt timely risk management approaches. For instance, dividing information areas based on geographic location and size (e.g., 30x30 square kilometers, small countries, or states within larger countries) and implementing risk control measures through information filtering, labeling, and other complementary methods. Additionally, government-led initiatives and support are essential to increase the involvement of minority and vulnerable groups in LLMs.

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?

All these uncertainties can be examined and determined by the top-level AI guidelines.