Response to Safe & Responsible AI Consultation Paper

NotCentralised

26th July 2023

CREATED BY
NotCentralised

REPRESENTATIVES
Nick Bishop
Nathalie Mann
Mark Monfort
Arturo Rodriguez

Overview

The Director, Technology Strategy Branch
Department of Industry, Science & Resources
Email: DigitalEconomy@industry.gov.au

Dear Consultation Committee,

We, at NotCentralised, are pleased to contribute to the ongoing dialogue on the safe and responsible use of Artificial Intelligence (AI) in Australia. As a pioneering organisation in the field of emerging technologies, we believe that our unique insights and experience can add valuable perspectives to this important discussion.

NotCentralised is a dynamic entity that operates at the intersection of blockchain and AI technologies. We are the proud orchestrators of two vibrant communities: the Aus DeFi Association, which explores the potential of blockchain technology, and the Data Science and AI Association of Australia (DSAI), which delves into AI and Machine Learning innovations. Our work extends beyond these communities, as we collaborate with companies utilising blockchain and AI technologies and develop proprietary products within this burgeoning sector.

Our portfolio includes a diverse range of projects, and we have established strong collaborations with Federal and State governments. Notably, we were one of the 15 pilot projects for the Reserve Bank of Australia's Central Bank Digital Currency (CBDC) initiative. We also play a significant role in the New South Wales Government Taskforce on Data, Digital and AI, further solidifying our influence and expertise within the emerging technology sector.

In the healthcare domain, we leverage cutting-edge AI tools to design private GPT models that enhance patient outcomes while upholding data privacy. We also develop business-grade GPT use cases tailored to specific roles, ensuring that access to sensitive business information is strictly controlled.

Our initiatives involve the seamless integration of blockchain and AI technologies. We use blockchain as a verification function for source data for training Large Language Models (LLMs), guaranteeing the integrity of information. We also harness Zero-Knowledge

technology to detect any modifications to the data without the need to disclose sensitive information, reinforcing the security and trustworthiness of our AI applications.

NotCentralised is a Corporate Authorised Representative of BK Consult, an Australian Financial Services License (AFSL) holder. This allows us to assist with raising capital for projects spanning the emerging tech space. This vantage point gives us a comprehensive understanding of the ever-evolving trends in startup investments, both domestically and internationally.

Given our extensive involvement in the emerging tech space, we believe we are well-positioned to provide a comprehensive and nuanced perspective on the future of AI in Australia. In our response, we share high-level details of our projects which showcase how AI and blockchain tech can be used to improve AI safety. Further, we delve into the specific questions raised in the consultation paper, providing our insights on AI regulation and governance.

We look forward to contributing to this important discussion and helping shape the future of AI in Australia.

Best regards,

Mark Monfort, Nick Bishop, Arturo Rodriguez

Directors

NotCentralised

Part One - Background on NotCentralised

As part of our ongoing commitment to advancing technology for the betterment of society, NotCentralised is deeply interested in the potential of blockchain and AI technology to enhance the safety and reduce the risks associated with AI systems.

One crucial aspect of AI safety is the question of data security and privacy. Blockchain technology, with its decentralised nature and cryptographic principles, provides a robust platform for ensuring that data used in AI models remains secure, and its integrity is maintained.

Our project, TradeFlows, is an example of how blockchain can be used to bolster security and transparency. TradeFlows is a blockchain-based payments and invoicing platform, which utilises smart contracts and a collaborative layer to increase trust and efficiency in financial operations. By providing transparency around collateral through the use of escrows and automated payment processes, we are reducing risks and providing an extra layer of safety to businesses.

Zero-Knowledge Proofs (ZKPs), used now in blockchain technology, can be applied also to the domain of AI safety. ZKPs allow for the verification of data changes without revealing the data itself, ensuring the privacy of sensitive information even as it contributes to the function of an AI system. This could be particularly relevant in the context of LLMs and other AI models that rely heavily on large, diverse datasets, some of which might contain sensitive information.

From the AI perspective, machine learning techniques can be employed to monitor the outputs of an AI system for any anomalies or deviations from expected behaviours. Just as AI can be trained to identify the style or 'fingerprint' of a specific input in an LLM's output, it could also be trained to recognise patterns indicative of system malfunction or misuse.

Finally, combining AI with blockchain, for instance in the form of Decentralised Autonomous Organisations (DAOs), can lead to improved AI safety. DAOs can serve as democratic, transparent systems for decision-making, reducing the risk of human error or bias. This would be like a futuristic version of a royalty collection society but one that is not centralised and is transparent in its operations. These structures can help ensure AI is used responsibly, maintaining checks and balances on the AI’s operations and outputs.

In summary, NotCentralised believes that a future-focused approach to AI safety must involve a deep integration of AI with blockchain technologies. By combining the strengths of these technologies, we can create AI systems that are not only powerful and efficient but also safe, transparent, and accountable.

Part Two - Responses to Consultation Paper

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

Firstly, we believe that definitions are of great importance when it comes to discussions around potential regulation or oversight of technologies, given the different experiences and expectations that stakeholders will have in this space.

While we agree generally with the definitions provided in the discussion paper - they are comprehensive and align well with widely accepted definitions in the field of AI - we believe there should be consideration for expansion when it comes to this rapidly evolving field. Having a mechanism whereby definitions can be reviewed with a set frequency will ensure that items becoming part of regulation are in step with the technology.

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?

Attribution of Blame and AI Specific Liability Laws: We are fortunate to hear a wide range of conversations via NotCentralised and in the Associations that we run,which both talk about AI. A significant area of concern and confusion is the attribution of blame when it comes to who is responsible for AI decision making. In the transport sector, humans driving non-autonomous cars are responsible for their actions on the road. However, the semi-automation of vehicles has led to the question of who is responsible for vehicular accidents; human or manufacturer? The same can be said when it comes to generative AI and the outcomes from further use of these tools. In one camp, there are those considering these technologies as nothing more than tools, for which humans are ultimately responsible. In another camp, some extrapolate further into scenarios where corporations, and roles within corporations, are run by AI (CEOs for example). New liability laws could be developed to address this issue, potentially including the concept of "AI personhood", or holding companies strictly liable for their AI's actions.

Transparency and Explainability: AI systems, particularly those based on complex machine learning models, can be "black boxes," making decisions that are difficult to interpret or explain. This lack of transparency can make it challenging to determine how a decision was made and who should be held accountable for it. Regulatory action could include requiring AI developers to provide explanations for their AI systems, especially for high-stakes applications like healthcare or autonomous vehicles. Blockchain plays a role in helping provide audit functions for data sources of AI models. Furthermore, when it comes to systems dealing with sensitive datasets, we believe that the use of zero-knowledge proofs (as we mentioned earlier) can help by enabling auditors to see changes, without sensitive data being shared unnecessarily.

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.

There are a number of non-regulatory initiatives that the Australian government can get involved in to help support responsible AI:

Education and Training: Develop and implement AI education and training programs for various sectors of the economy. This could include workshops, seminars, and online courses that provide knowledge about AI, its uses, and ethical considerations. This would help to increase understanding and responsible use of AI across different sectors.

Public-Private Partnerships: Encourage partnerships between government, industry, and academia to foster innovation in AI while ensuring ethical standards are maintained. These partnerships could lead to the development of new AI technologies that are beneficial to society and the economy, cementing Australia as a leading technology developer.

AI Research Funding: Provide funding for research into AI and its impacts, particularly in areas such as fairness, transparency, and accountability. This could help to advance our understanding of AI and how to use it responsibly.

AI Ethics Guidelines: While not a regulation, the government could develop and promote a set of voluntary guidelines, or best practice, for ethical AI development and use. These

guidelines could provide a framework for organisations to ensure they are using AI in a way that is ethical and responsible.

AI Testbeds and Sandboxes: Create environments where AI systems can be tested and experimented with in a controlled manner. This would allow for the exploration of AI capabilities and limitations, and the identification of potential ethical or societal issues before full-scale deployment. NotCentralised is part of the NSW Government Taskforce on Data, Digital and AI. This was setup by NSW Services Minister Victor Dominello before he left office and is headed up by NSW Chief Scientist, Dr Ian Oppermann. Setting up something at a national level that integrates into what the States are doing would be ideal.

Public Awareness Campaigns: Conduct campaigns to educate the public about AI, its benefits, potential risks, and the importance of responsible use. This could help to build public trust in AI technologies.

Support for AI Competitions and Challenges: Organise competitions and challenges that encourage the development of AI solutions to societal problems, with a focus on ethical and responsible AI development. There are a number of groups including not-for-profits which would be very open to government support to help spread the message of ethical use of AI and which are open for the Government to speak at. This includes the Data Science and AI Association of Australia (DSAI) (which NotCentralised helps to run) that ran an AI-focused hackathon earlier this year in partnership with City of Sydney, UTS and others.

Support for AI Startups: Provide support for startups working on AI, such as grants, mentorship programs, and networking opportunities. This could help to foster a vibrant AI ecosystem in Australia. Already, global startup accelerators like Techstars have come to Australia. We need more involvement from Federal Government to help fund startup innovation and avoid homegrown talent moving overseas.

AI Advisory Council: Establish an advisory council made up of experts from various fields related to AI. This council could provide advice to the government on AI-related issues and help to guide the development of non-regulatory initiatives. It would need to include not just those in industry and vendors from the large end of town, but also SMEs, associations and others involved in the grassroots communities.

International Collaboration: Collaborate with other countries on AI research and policy development. This could help to ensure that Australia's approach to AI is in line with international best practices, and even forge stronger links in South East Asia with nations for which relations are tepid.

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.

We believe that the following are areas of coordination across government that can help with the development and uptake of AI in Australia.

Establish a Central AI Agency: A central agency could be established to oversee all AI-related initiatives across different government departments. This agency - perhaps not christened the CAI - could set standards, provide guidance, and ensure consistency in the application and regulation of AI. The goal would be to create a unified approach to AI across all government sectors, which could help to streamline the development and uptake of AI in Australia.

Interdepartmental AI Committee: Without wishing to wrap Australia in more red tape, an interdepartmental committee could be formed, consisting of representatives from various government departments that are using or planning to use AI. This committee could meet regularly to share experiences, discuss challenges, and coordinate efforts. This could help to ensure that all departments are aligned in their approach to AI and are learning from each other's experiences.

AI Strategy and Roadmap: The government could develop a national AI strategy and roadmap, outlining the goals for AI development and use in Australia, and the steps needed to achieve these goals. This strategy could be used to guide all government departments and PPPs in their AI initiatives.

Shared AI Resources: The government could create a shared pool of AI resources, such as prompt tools, datasets, and expertise, that all departments can access. This could help to

reduce duplication of effort and ensure that all departments have access to the resources they need to implement AI effectively.

Regular AI Audits: Regular audits could be conducted to assess the use of AI across different government departments. These audits could help to identify any issues or inconsistencies, and ensure that all departments are using AI in a way that is ethical, transparent, and accountable.

AI Training for Government Employees: The government could provide AI training for all government employees, to ensure that they have the knowledge and skills needed to use AI effectively and responsibly. This could help to increase the uptake of AI across all government departments.

International Collaboration: The government could collaborate with other countries on AI governance, to learn from their experiences and ensure that Australia's approach is in line with international best practices.

Further Public Consultation and Engagement: Apart from this consultation, the government could engage in regular public consultation and engagement activities, to ensure that the public's views and concerns are taken into account in the development and implementation of AI policies and initiatives.

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?

The following are examples of governance measures we believe are relevant for what Australia is trying to accomplish

European Union: The EU is currently conducting a major experiment in the design of AI governance. It is working on a comprehensive regulatory framework for AI, which includes strict rules for high-risk AI systems and a ban on certain uses of AI that could be harmful. The EU's approach is risk-based, with different levels of regulation for different levels of risk.

Source: Lessons From the World's Two Experiments in AI Governance

China: China has the most comprehensive suite of AI regulations in the world, including newly released draft measures for managing AI. China's approach to AI governance is more centralised and top-down, with the government playing a significant role in directing and regulating AI development and use.

Source: Developments in the regulation of Artificial Intelligence - KWM

Japan: Japan has adopted an "agile governance" approach to digital governance, including AI. This approach is flexible and adaptable, allowing for quick responses to new developments and challenges in the AI field. Japan's approach to AI regulation is more sector-specific, with different regulations for different sectors of the economy. This contrasts greatly with the Chinese example above.

Source: Japan's Approach to AI Regulation and Its Impact on the 2023 G7 Presidency - CSIS

United States: The US is one of the leading countries in developing AI governance strategies, not least because much of the technology has its origins there. The US approach to AI governance is more laissez-faire, with a focus on promoting innovation and competition. The US has also taken steps to address AI ethics and bias, including efforts to develop standards and guidelines for fair, accountable, and transparent AI.

Source: Catching up with AI: Pushing toward a cohesive governance framework

OECD: The Organisation for Economic Co-operation and Development (OECD) supports governments by measuring and analysing the economic and social impacts of AI technologies and applications, and engaging with all stakeholders to develop policy guidance on AI including tools like the OECD.AI Policy Observatory.

Source: Artificial intelligence - OECD

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

Yes, in part, different approaches should apply to public and private sector use of AI technologies due to the distinct roles, responsibilities, and impacts these sectors have on society. Here's how the approaches could differ:

Public Sector: The public sector has a responsibility to protect citizens' rights, uphold transparency, and ensure fairness. Therefore, the use of AI in the public sector should be held to high standards of accountability, transparency, and fairness. This includes clear disclosure of the use of AI systems, the data they use, and the decision-making processes they follow. The public sector should also prioritise the use of AI in ways that directly benefit the public, such as improving public services or policy decision-making.

Private Sector: The private sector, while also needing to uphold standards of fairness and transparency, often deals with competitive and proprietary information. Therefore, the use of AI in the private sector might require a different approach. For instance, there could be more emphasis on protecting intellectual property and trade secrets, while still ensuring ethical use of AI.

The private sector might benefit from more robust auditing capabilities for the data used to train AI models and the data outputs. Blockchain technology, combined with zero-knowledge proofs, could be a powerful tool for this. It would allow for the verification of data integrity without revealing the data itself, which could be crucial for maintaining privacy and confidentiality. This could also help track the use of AI outputs, ensuring they are used responsibly and ethically.

In addition to these, there are a few other considerations:

Collaboration and Knowledge Sharing: Both sectors could benefit from increased collaboration and knowledge sharing. For instance, the public sector could share research and best practices on ethical AI use, while the private sector could contribute innovative solutions and technical expertise.

Regulation and Oversight: The public sector typically has a role in regulating the use of AI, both within the public sector and in the private sector. This could include setting standards for ethical AI use, conducting audits, and enforcing penalties for misuse. The private sector, on the other hand, is subject to these regulations, while also having a role in shaping them through consultation and engagement processes.

Public Engagement: Both sectors should engage with the public and other stakeholders to understand their views and concerns about AI. However, the ways they do this might differ. For instance, the public sector might hold public consultations or use citizen juries, while the private sector might use surveys or focus groups.

In conclusion, while there are common principles that should guide the use of AI in both sectors, such as fairness, transparency, and accountability, the ways these principles are implemented might differ based on the unique roles and responsibilities of the public and private sectors.

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

Here are ways we believe that the Australian Government can further support responsible AI practices in its own agencies:

Establish AI Governance Frameworks: Each agency should follow a common set of overarching principles and rules (see above). This could include guidelines on data privacy, transparency, accountability, and fairness. However, when it comes to subject matter specific to a given agency, it may develop additional AI governance, outlining how it will use AI responsibly in its field.

Creation of an AI Ethics Board: An AI Ethics Board could be established which would have visibility over each agency, to oversee the use of AI and ensure it aligns with ethical guidelines. This board could include a diverse group of stakeholders, including ethicists, data scientists, and representatives from the public.

Implement AI Auditing Systems: Regular audits of AI systems could be conducted to ensure they are operating as intended and not causing any unintended harm. These audits could also assess the data being used by the AI to ensure it is accurate and unbiased. As mentioned earlier, NotCentralised believes in the use of blockchain technology combined with zero-knowledge proofs as a means of providing a secure, immutable audit trail for data usage. This would allow for trustworthy sources to audit these models (whether that is a manual or an AI-automated audit).

Promotion of Interagency Collaboration: Encourage collaboration between agencies on AI projects. This could help to share knowledge, reduce duplication of effort, and ensure a consistent approach to AI across the government (see further above).

Public Transparency and Engagement: Agencies could engage with the public and other stakeholders to understand their views and concerns about the use of AI. This could include public consultations, citizen juries, or other forms of public engagement.

Investment in AI Research and Development: Government agencies could invest in research and development to advance the state of AI and its responsible use. This could include partnerships with universities or research institutions.

Develop AI Skills within the Workforce: In addition to continuous training, agencies could also focus on developing AI skills within their workforce. This could include hiring data scientists and AI specialists, or providing opportunities for existing staff to retrain or upskill in AI.

Use of Explainable AI: Wherever possible, agencies should use AI systems that provide clear explanations for their decisions. This can help to ensure transparency and trust in the AI systems used by the government.

Pilot Projects: Before fully implementing an AI system, agencies could conduct pilot projects to test the system and identify any potential issues or risks. This can help to ensure that the system is fit for purpose and will not cause any unintended harm.

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.

N/A

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

9a - where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI?

Transparency is indeed a crucial factor in mitigating potential risks associated with AI and in bolstering public trust and confidence in these technologies. Your points about the importance of transparency at both the input and output stages of a generative AI model are well-taken. Here's a more elaborated version of your response:

Transparency is of utmost importance at two key stages in the lifecycle of a generative AI model: the data input stage and the results output stage.

At the data input stage, transparency is critical to ensure that the data used to train the AI model is unbiased, representative, and ethically sourced. This is where the model learns from and forms its decision-making basis. Any bias or error at this stage could lead to skewed or unfair outcomes. The classic GIGO problem - garbage in, garbage out.

At the results output stage, transparency helps users understand how the AI model arrived at its decisions or predictions. This is particularly important when AI is used in high-stakes domains, such as healthcare or criminal justice, where decisions can have significant impacts on individuals' lives.

In the context of our work in the web3 space, we see blockchain technology as a promising tool for enhancing transparency in AI systems. Blockchain's immutable and decentralised nature can provide a reliable record of data used in training AI models and the outcomes they produce, without being “owned” by any single entity.

Specifically, techniques like Zero-Knowledge Proofs (ZKPs) can be utilised to interrogate aspects of datasets, without revealing the underlying data involved. This allows for the verification of data used in training the models without revealing the data itself, thereby preserving privacy. Similarly, ZKPs can be used to cryptographically verify that a given

output has been generated by an AI model. This can help in situations where it's important to distinguish between human-generated and AI-generated content.

By integrating these techniques into AI systems, we can create a more transparent and trustworthy AI ecosystem. This, in turn, can help mitigate potential risks and improve public trust and confidence in AI technologies.

9b - mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.

In light of the critical role that transparency plays throughout the AI lifecycle, we propose the following measures:

Mandating Cryptographic Tagging: We suggest the implementation of new laws that mandate the tagging of both input and output data using cryptographic techniques such as Zero-Knowledge proofs. This approach can enhance the traceability and accountability of AI systems by providing verifiable evidence of the data used for training and the outputs generated from that data.

Retroactive Labelling: Recognising that many AI models have already been trained on vast amounts of data, we propose that these data sets be retroactively labelled wherever feasible. This would involve going back to the original data used for training and tagging it appropriately, which could help in understanding the decision-making process of these models.

Output Labelling: In addition to input data, it's crucial to label the outputs generated by AI models. Companies should be required to store this information, which can be used for auditing and accountability purposes. Decentralised storage, with certain common data standards, and immutable audit trails would be useful in this respect.

Future-Proofing AI Models: Looking ahead, we recommend that all generative AI models be designed with this cryptographic tagging as a foundational requirement. This would ensure that all future AI systems are inherently transparent and accountable.

Transparency Across Sectors: These requirements should apply to both the private and public sectors. Transparency in AI is not just a corporate responsibility, but a societal one, and it's crucial that all stakeholders are held to the same basic standards.

Implementation Strategies: To implement these requirements, regulatory bodies could provide guidelines and resources to assist organisations. This could include technical guidance on cryptographic tagging, as well as frameworks for how to manage and store data. Regular audits could be conducted to ensure compliance, and penalties could be imposed for non-compliance.

By adopting these measures, we can ensure that AI systems are transparent and accountable, thereby fostering greater trust and confidence in these technologies.

10 - Do you have suggestions for:

10a - whether any high-risk AI applications or technologies should be banned completely?

Rather than implementing outright bans on certain high-risk AI applications or technologies, we suggest a more nuanced approach. Banning specific AI technologies could have several unintended consequences:

International Competitiveness: Prohibiting certain AI technologies could put Australia at a disadvantage compared to other nations, states, and corporations that continue to develop and utilise these technologies. This could lead to a technological gap, potentially impacting Australia's global standing in AI innovation.

Innovation Stifling: An outright ban could stifle domestic innovation. Australian researchers, developers, and companies might be discouraged from exploring new AI applications if they fear potential bans. This could limit the growth of Australia's AI sector and hinder its ability to compete on the global stage.

Missed Opportunities: While certain AI technologies might pose risks, they also often present new opportunities for solving existing problems. The benefits provided by these

technologies could outweigh the introduced risks, especially if appropriate safeguards are put in place.

Instead of banning, we suggest a regulatory framework that includes stringent oversight, rigorous testing, and robust risk management for “high-risk” AI applications - acknowledging this would need to be defined. This approach would allow Australia to harness the benefits of AI while minimising potential risks. It would also encourage responsible innovation and ensure that Australia remains competitive in the global AI landscape.

10b - criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts?

N/A

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

To increase public trust in AI deployment and encourage more people to use AI, the government can undertake a variety of initiatives and actions. One of the key strategies could be to provide more support for associations that are actively working to educate the public and advocate for the responsible use of AI technology. For instance, organisations like the Data Science and AI Association (DSAI) play a crucial role in promoting understanding and acceptance of AI.

In addition to this, the following measures could be considered:

Public Education and Awareness Campaigns: The government could launch campaigns to educate the public about AI, its benefits, and its potential risks. In particular, selected use of this technology in schools and universities would be warranted, given the likely

prevalence across the economy in future. This could involve collaboration with technical educational institutions, community organisations, and the media.

Transparency and Openness: The government could require companies to be transparent about their use of AI, including how decisions are made by AI systems. This could help to demystify AI and make it less intimidating for the average person.

Regulatory Oversight: Establishing clear regulations for AI can help to build public trust by ensuring that AI systems are used responsibly and ethically. This could include regulations around data privacy (see the comments on data tagging above), fairness, and accountability.

Public Participation: The government could involve the public in decision-making processes related to AI, such as the development of AI policies and regulations. This could help to ensure that the public's concerns and interests are taken into account.

Promotion of Ethical AI Practices: The government could promote the development and use of AI that adheres to ethical principles, such as fairness, transparency, and respect for human rights. This could involve providing incentives for companies that follow ethical AI practices.

Investment in AI Research and Development: The government could invest in AI research and development, including research into how to make AI more understandable and controllable. This could help to advance the state of AI technology and make it more trustworthy - less of a black box.

By implementing these initiatives, the government can play a significant role in fostering a more informed and trusting environment for the deployment and use of AI.

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?

Banning high-risk activities such as social scoring or facial recognition technology in certain circumstances could have multifaceted impacts on Australia's tech sector and its trade and exports with other countries and we consider both the bad and the good:

Bad

Potential Negative Perception: As your initial point suggests, such a ban could create a perception that Australia is not fully open for business in the tech sector. This could potentially deter foreign investment and collaboration.

Innovation Constraints: Restricting the use of certain technologies may limit the scope for innovation within Australia's tech sector. It could discourage research and development in these areas, potentially causing Australia to fall behind other countries in the development and application of these technologies.

Trade Implications: If other countries continue to develop and use these technologies, a ban could put Australian companies at a competitive disadvantage in international markets. It could also complicate trade relations with countries that have different regulatory approaches to these technologies.

Good

Regulatory Balance: On the other hand, a ban on high-risk activities could also be seen as a strong stance on ethical tech development and use. This could enhance Australia's reputation as a country that prioritises the responsible use of AI and other technologies, which could attract businesses and investors who share these values.

Opportunity for Specialisation: If certain high-risk activities are banned, it could encourage the tech sector to specialise in other areas of AI and technology development. This could lead to the growth of new industries and innovations that align with Australia's regulatory environment.

Public Trust: Such a ban could increase public trust in technology and AI, as it would demonstrate the government's commitment to protecting citizens from potential harms. This could lead to increased adoption of AI and other technologies in sectors where their use is deemed safe and beneficial.

In conclusion, while there are potential negative impacts, there are also opportunities for Australia to differentiate itself as a leader in responsible and ethical technology development and use.

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

N/A

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

We firmly believe that a risk-based approach is a prudent and balanced strategy for addressing potential AI risks. This perspective is shared by many experts in the field, who recognize the varying degrees of risk associated with different AI applications.

Under a risk-based framework, AI technologies that carry a high risk—such as those employed in healthcare diagnostics or autonomous vehicle / aircraft control—would be subject to rigorous regulations and oversight. This is crucial because these applications have a direct and significant impact on human lives, and any malfunction or error could lead to severe consequences.

Conversely, AI applications associated with lower risk—such as recommendation algorithms for online shopping or music streaming—would be subject to less stringent regulations. While these applications can influence consumer behaviour and choices, their potential to cause harm is considerably less than high-risk applications.

However, it's important to note that even 'low-risk' applications can have unintended, pernicious consequences, such as perpetuating bias or infringing on privacy. Therefore, a risk-based approach should not mean an absence of regulation for low-risk AI, but rather a proportional response that balances the need for innovation with the imperative of safety and ethical considerations.

Furthermore, a risk-based approach should be dynamic and adaptable, capable of evolving with the rapid pace of AI development. It should also be complemented by other measures, such as promoting transparency, ensuring accountability, and fostering a culture of ethical AI development and use.

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

We see a risk-based approach having several benefits:

- **Proportionality**: It ensures that the level of regulation is proportional to the level of risk, preventing over-regulation of low-risk AI applications and under-regulation of high-risk applications.
- **Flexibility**: It allows for flexibility in regulation, as the level of oversight can be adjusted as the level of risk changes.
- **Efficiency**: It allows regulatory resources to be focused on the areas where they are most needed, rather than being spread thinly across all AI applications regardless of risk.

However, if this approach is taken, we would caution that there are some drawbacks to address:

- **Assessment of risk**: Determining (even defining) the level of risk associated with a particular AI application can be challenging, particularly given the rapid pace of AI development and the complexity of many AI systems.
- **Dynamic nature of risk**: The level of risk associated with a particular AI application can change quickly as the AI system learns and evolves, making it difficult to maintain an appropriate level of regulation.
- **Unintended consequences**: There is a chance that a risk-based approach could stifle innovation in high-risk areas, where the potential benefits of AI could be greatest.

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?

N/A

17 - What elements should be in a risk-based approach for addressing potential AI risks? Do you support the elements presented in Attachment C? Attachment C is found on page 40 of the discussion paper.

First and foremost, we believe that a risk-based approach could be better suited to some sectors, AI applications, or organisations than others. For instance, sectors that deal with sensitive data or have a high potential for harm (like healthcare, autonomous vehicles, legal use cases, or even in finance) might benefit more from a risk-based approach compared to others where the risk is lower (e.g. online shopping, hospitality). Similarly, AI applications that have a high potential for harm or misuse might also be better suited to a risk-based approach.

Organisations of different sizes and resources might also find a risk-based approach more or less suitable. Larger organisations with more resources might be better equipped to implement a risk-based approach, while smaller organisations might find it more challenging. There is an analogy here to the risk weighting of assets in the banking sector - involving standardised and proprietary approaches. Similarly, organisations with a high level of AI maturity might be better able to implement a risk-based approach than those with a lower level of AI maturity, regardless of scale.

The elements presented in Attachment C provide a comprehensive framework for a risk-based approach to addressing potential AI risks. They cover all the key aspects of risk management, from risk assessment to risk monitoring and review, and they emphasise the importance of transparency, accountability, and stakeholder engagement. However, the effectiveness of these elements would depend on their implementation in practice, and they might need to be adapted or supplemented to address specific risks or contexts.

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?

N/A

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

We believe that a risk-based approach to general-purpose AI systems like Large Language Models (LLMs) or Multimodal Foundation Models (MFMs) would involve the same key elements as outlined in Attachment C of the discussion paper: risk assessment, risk management, risk monitoring and review, transparency and accountability, and stakeholder engagement. Here's how these elements might apply:

Risk Assessment: For LLMs and MFMs, the risk assessment would need to consider the wide range of potential uses and misuses of these models. This could include risks related to the generation of inappropriate or harmful content, the amplification of biases in the training data, the potential for misuse in spreading misinformation or disinformation, and the privacy risks associated with the use of personal data in training these models.

Risk Management: The risk management measures for LLMs and MFMs could include technical measures such as robustness and accuracy requirements, bias mitigation techniques, and privacy-preserving methods. Organisational measures could include governance structures and processes for overseeing the use of these models, and regulatory measures could include compliance with existing laws and regulations related to data protection, non-discrimination, and consumer protection.

Risk Monitoring and Review: Given the general-purpose nature of LLMs and MFMs, ongoing monitoring and review of the risks associated with these models would be particularly important. This could involve regular audits or inspections of the models'

performance and impact, and periodic reviews of the risk assessment and risk management measures.

Transparency and Accountability: Transparency about the functioning, use, and risks of LLMs and MFMs would be crucial for building trust and ensuring accountability. This could involve disclosure of information about the training data, the training process, and the performance of these models. Accountability mechanisms could include processes for handling complaints and redress mechanisms for individuals or groups harmed by these models.

Stakeholder Engagement: Given the wide range of potential impacts of LLMs and MFMs, engaging with a broad range of stakeholders would be important. This could include consultation with users, communities, experts, and regulators, and collaboration with other organisations in sharing best practices and lessons learned.

In addition to these elements, a risk-based approach to LLMs and MFMs might also need to consider the dynamic and evolving nature of these models. As these models continue to learn and adapt over time, the risks associated with them might also change, requiring ongoing reassessment and adaptation of the risk management measures.

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:

20a - public or private organisations or both?

The decision to make a risk-based approach for responsible AI a voluntary/self-regulation tool or a mandated regulation depends on several factors, including the potential risks and harms associated with AI, the maturity of the AI sector, and the existing regulatory environment.

Voluntary or Self-Regulation: This approach can be beneficial in rapidly evolving fields like AI, where regulation might struggle to keep pace with technological advancements. It allows for flexibility and innovation, and it can be particularly effective when there's a

strong culture of ethics and responsibility in the AI sector. However, the effectiveness of voluntary or self-regulation depends on the willingness and ability of organisations to implement it properly. Without proper oversight and enforcement, there's a risk that some organisations might not take it seriously or might implement it in a way that's ineffective or superficial.

Mandated Regulation: This approach can provide more certainty and consistency, and it can ensure that all organisations, regardless of their size or resources, are held to the same standards. It can be particularly effective in addressing systemic risks or harms that can't be effectively managed by individual organisations alone. However, mandated regulation can be more rigid and slow to adapt to new developments, and it can potentially stifle innovation if not designed and implemented carefully. It could also create an uneven playing field for those developing AI, because the fixed costs associated with compliance would impact small organisations most heavily.

As for whether a risk-based approach should apply to public or private organisations, and to developers or deployers, it arguably should apply to all of them:

Public or Private Organisations: Both types of organisations can develop or use AI, and both can potentially cause harm if they do so irresponsibly. Therefore, both should be subject to a risk-based approach to responsible AI.

20b - developers or deployers or both?

Developers or Deployers: Both developers and deployers of AI have a role to play in managing the risks associated with AI. Developers can influence the design and functionality of AI systems, while deployers can influence how these systems are used in practice. Therefore, both should be involved in the risk assessment, risk management, and risk monitoring and review processes.