Sustainable Business Australia Limited trading as
Business Council for Sustainable Development Australia
ABN 48 052 135 609
Level 27
20 Bond Street
SYDNEY NSW 2000
+ 61 2 8005 0780
bcsda@bcsda.org.au
www.bcsd.org.au

bcsd australia

Technology Strategy Branch
Department of Industry, Science and Resources
Australian Government

21 July 2023

By email: DigitalEconomy@industry.gov.au

Dear Technology Strategy Branch,

Submission in response to Responsible AI in Australia

On behalf of the Business Council for Sustainable Development (BCSD) Australia, I am pleased to submit our response to the Australian Government's consultation on AI governance. BCSD Australia, a leading business-led sustainability and business peak body, is committed to driving sustainable development in Australia by fostering collaboration among businesses and promoting responsible practices.

BCSD Australia is an affiliate of the World Business Council for Sustainable Development (WBCSD) and a part of its global network. The WBCSD is a CEO-led organization of over 200 leading companies from various sectors, working towards the integration of sustainability into business strategies and practices. Our affiliation with the WBCSD ensures that our submissions and recommendations are informed by global expertise and best practices in sustainable business.

Our submission encapsulates the insights gathered from extensive research and consultation with our member companies, who are at the forefront of sustainable business practices. We have addressed the Australian Government's specific questions on AI governance, highlighting a progressive leading practice perspective and emphasizing the relevance to sustainable development and the achievement of the Sustainable Development Goals (SDGs).

In our response, we have highlighted the importance of robust AI governance mechanisms in ensuring the responsible and ethical use of AI technologies. We underscored the need for clear definitions, the identification of potential risks not covered by existing regulatory approaches, and the importance of both regulatory and non-regulatory initiatives to support responsible AI practices.

Furthermore, our submission emphasizes the significance of coordination of AI governance across government departments to achieve consistency, alignment, and efficient utilization of resources. We have provided examples of governance measures being taken or considered by other countries that are adaptable and relevant to Australia's context.

Specifically, in the context of sustainability and climate risk disclosure, ESG data, metrics, analytics, and reporting, we have outlined how AI and the need for AI Governance intersect. Our submission highlights how AI enhances data collection and analysis, improves ESG performance measurement, identifies emerging risks, mitigates bias, and promotes transparency. We have supported our points with practical business and government examples that demonstrate the application of AI governance principles.

Our Answers to the Specific Questions are in **Appendix A**.

BCSD Australia firmly believes that by incorporating AI governance frameworks and practices, Australia can unlock the full potential of AI while safeguarding ethical considerations, public trust, and sustainable development objectives. Our recommendations aim to enhance the understanding and integration of responsible AI practices across various sectors.

We appreciate the opportunity to contribute to this consultation and look forward to further engagement with the Australian Government to shape effective AI governance policies that foster sustainable business practices and contribute to Australia's transition to a more sustainable and resilient future.

Thank you for your attention to our submission. Should you require any further information or clarification, please do not hesitate to contact us.

Yours faithfully,

Yours faithfully,

Andrew Petersen
CEO I **Business Council for Sustainable Development Australia**
World Business Council for Sustainable Development Australian Partner
0412 545 994 I andrew.petersen@bcsda.org.au

Executive Summary

As BCSD Australia, a leading business-led sustainability and business peak body, we provide the following answers to the 20 questions related to AI risk and governance.

Our responses reflect a progressive leading practice perspective and highlight the relevance of sustainable business practices in contributing to the Sustainable Development Goals (SDGs).

Here is a summary of our key insights:

1. Definitions: We agree with the definitions presented in the discussion paper, as they provide a clear foundation for understanding AI-related concepts and terminology.
2. Potential Risks and Regulatory Action: We identified potential risks from AI that may not be covered by Australia's existing regulatory approaches. We recommended regulatory actions such as mandatory impact assessments, notices to inform users, human oversight assessments, explanations of AI decisions, training programs, and continuous monitoring.
3. Non-Regulatory Initiatives: We highlighted the importance of non-regulatory initiatives to support responsible AI practices, including fostering collaboration, promoting industry standards and best practices, investing in research and development, and facilitating public-private partnerships.
4. Coordination of AI Governance: We suggested that coordination of AI governance across government should aim to ensure consistency, alignment, and efficient use of resources. This coordination can help establish clear guidelines, foster knowledge sharing, and enable effective decision-making regarding AI development and uptake in Australia.
5. International Governance Measures: We highlighted relevant governance measures taken or considered by other countries. Examples include the EU's proposed AI Act, collaborative initiatives like the Partnership on AI, and industry-specific guidelines developed by organizations such as the IEEE.
6. Different Approaches for Public and Private Sectors: We recognized that different approaches may be necessary for public and private sector use of AI technologies, considering factors such as sector-specific risks, resources, and accountability. Tailored guidelines and frameworks can address the unique challenges and responsibilities faced by each sector.
7. Australian Government's Support for Responsible AI: We provided suggestions for how the Australian Government can further support responsible AI practices in its own agencies. These include promoting awareness and understanding, establishing internal guidelines and frameworks, investing in AI skills development, and fostering collaboration with the private sector.
8. Generic and Technology-Specific Solutions: We highlighted that both generic and technology-specific solutions are valuable in addressing AI risks. Generic solutions provide broad frameworks and principles, while technology-specific solutions cater to unique risks associated with specific AI applications. Examples include industry-wide ethical guidelines and sector-specific risk assessments.
9. Importance of Transparency: We emphasized the critical role of transparency in mitigating AI risks and building public trust. Transparency is most valuable during high-stakes decisions affecting individuals and should be mandatorily required across private and public sectors, with clear guidelines for implementation.
10. Banning High-Risk AI Applications: We suggested that high-risk AI applications should be evaluated on a case-by-case basis, considering the potential for harm and societal impact. Criteria for identification and banning should be developed through collaborative efforts involving experts, stakeholders, and regulators.
11. Increasing Public Trust in AI: We recommended initiatives and government actions to increase public trust in AI deployment, such as promoting transparency, fostering inclusive and participatory decision-making, ensuring fairness and accountability, and facilitating public awareness and education.
12. Impact of Banning High-Risk Activities: We acknowledged that banning high-risk activities, such as social scoring or certain uses of facial recognition technology, may have implications for Australia's tech sector, trade, and exports. A balanced approach is needed to ensure responsible use while supporting technological advancements.
13. Changes to Conformity Infrastructure: We recognized the need for potential changes to Australian conformity infrastructure to support assurance processes in mitigating AI risks. These changes may involve enhancing data privacy regulations, developing industry standards, and establishing certification mechanisms.
14. Support for a Risk-Based Approach: We expressed support for a risk-based approach to address potential AI risks.

Relevance of the SDGs to AI Governance

The **Sustainable Development Goals** (SDGs) adopted by the United Nations, and ratified by Australia in 2016, provide a comprehensive framework for addressing global challenges and achieving sustainable development by 2030. Several specific goals within the SDGs are relevant to AI. Here are some examples along with possible targets and guidance:

Goal 3: Good Health and Well-being

Target: Ensure universal access to affordable and quality healthcare services, including through AI-enabled technologies.

Guidance: Develop AI applications for early disease detection, personalized medicine, telemedicine, and healthcare resource optimization while ensuring privacy and ethics.

Goal 4: Quality Education
Target: Enhance inclusive and equitable access to quality education through AI-based tools and platforms.
Guidance: Develop AI-powered educational technologies to provide personalized learning, adaptive assessments, and improved access to education for marginalized communities.

Goal 5: Gender Equality
Target: Eliminate gender disparities in STEM fields and promote gender-inclusive AI development and use.
Guidance: Encourage equal representation of women in AI research and development, address biases in AI algorithms, and promote gender-sensitive AI applications.

Goal 8: Decent Work and Economic Growth
Target: Promote inclusive and sustainable economic growth by fostering AI innovation and job creation.
Guidance: Invest in AI research and development, support AI start-ups, and provide re-skilling and up-skilling programs to ensure a smooth transition in the workforce affected by AI automation.

Goal 9: Industry, Innovation, and Infrastructure
Target: Increase access to affordable and sustainable technologies, including AI, in developing countries.
Guidance: Foster international cooperation and technology transfer to bridge the digital divide and provide capacity-building support for developing nations to harness the benefits of AI.

Goal 16: Peace, Justice, and Strong Institutions
Target: Enhance transparency, accountability, and ethical use of AI technologies to ensure responsible governance.
Guidance: Establish regulations and guidelines for AI development, deployment, and use, addressing issues such as bias, privacy, and algorithmic transparency. Promote multi-stakeholder engagement and ethical AI frameworks.

Goal 17: Partnerships for the Goals
Target: Strengthen global partnerships to support AI research, development, and capacity-building efforts.
Guidance: Foster international collaborations, knowledge sharing, and resource mobilization to facilitate AI-driven solutions for sustainable development.

It is important to note that the SDGs are interconnected, and AI can contribute to multiple goals simultaneously. By aligning AI strategies and initiatives with the SDGs, governments, organizations, and AI practitioners can work towards achieving sustainable and inclusive development while addressing specific targets and following the guidance provided by the goals.

AI Governance's Impact on Sustainability and Climate Risk Disclosure, ESG Data, Metrics, Analytics, and Reporting

The rise of Artificial Intelligence (AI) brings both opportunities and challenges for sustainability and climate risk disclosure, as well as Environmental, Social, and Governance (ESG) data, metrics, analytics, and reporting. Effective AI Governance is crucial to ensure ethical, transparent, and reliable practices in these domains. Below we have outlined the intersection of AI and sustainability, providing, where possible, current business and government examples.

1. **Enhanced Data Collection and Analysis**: AI-driven platforms are and will enable organizations to collect and analyze vast amounts of ESG-related data efficiently. For instance, Microsoft's AI for Earth program utilizes AI algorithms to process satellite imagery and sensor data, aiding in environmental monitoring and ecosystem management. The resulting insights contribute to better-informed sustainability reporting and decision-making.

   However, it's important to consider the counter challenge of the processing power required by large language and data models. The subsequent energy consumption and potential emissions impact cannot be overlooked as we transition to a world of general use of AI. While AI can significantly contribute to environmental sustainability, the energy requirements of these systems could potentially offset some of the benefits. Therefore, it's crucial to balance the advantages of AI in data collection and analysis with the need for energy-efficient and environmentally friendly computing solutions.
2. **Improved ESG Performance Measurement**: AI-powered analytics can enhance the measurement and monitoring of ESG performance indicators. For example, the asset management company BlackRock employs natural language processing and machine learning algorithms to analyze ESG data from various sources. This enables comprehensive assessment and comparison of companies' ESG performance, leading to more robust ESG reporting.
3. **Identification of Emerging ESG Risks**: AI can assist in identifying and assessing emerging ESG risks, including climate-related risks. The Climate Mind platform, developed by the Swiss Federal Institute of Technology Zurich (ETH Zurich),

employs AI algorithms to analyze social media data and identify public sentiment and perceptions related to climate change. Such insights support proactive risk management and inform climate risk disclosure strategies.

4. **Mitigation of Bias and Enhancing Transparency**: AI Governance plays a crucial role in addressing algorithmic biases and ensuring transparency in ESG data and reporting. For instance, OpenAI's GPT-3 language model incorporates fine-tuning methods to mitigate biases, reducing the risk of biased outputs in ESG-related applications. Transparent reporting on AI models' limitations and data sources helps build stakeholder trust.
5. **Ethical Use of AI for Sustainability**: AI Governance frameworks guide the ethical use of AI in sustainability and climate risk disclosure. The European Investment Bank (EIB) has developed AI-related guidelines, promoting responsible AI adoption across its operations. These guidelines emphasize transparency, explainability, and accountability, ensuring that AI technologies support sustainable practices and ethical decision-making.
6. **Standardization and Consistency**: AI Governance promotes standardization and consistency in ESG data, metrics, analytics, and reporting. The Sustainability Accounting Standards Board (SASB) provides industry-specific standards that incorporate AI considerations. For example, the SASB's framework for the software and IT services sector addresses AI-related risks, including data privacy, algorithmic bias, and responsible data use.
7. **Stakeholder Engagement and Accountability**: AI Governance encourages stakeholder engagement and accountability in sustainability reporting. The Global Reporting Initiative (GRI) actively involves stakeholders in the development and revision of reporting standards. By incorporating diverse perspectives, GRI ensures that AI-driven processes align with stakeholder expectations, improving trust and transparency in ESG reporting.
8. **Continuous Improvement and Adaptability**: AI Governance frameworks facilitate continuous improvement and adaptability in ESG data, metrics, analytics, and reporting. The CDP (formerly Carbon Disclosure Project) utilizes AI-driven algorithms to assess climate-related risks and opportunities for companies. By continuously refining their AI models based on feedback and emerging sustainability challenges, CDP enhances the accuracy and relevance of climate risk disclosure.

AI Governance is essential to harness the benefits of AI while addressing its ethical implications in sustainability and climate risk disclosure. Business and government examples illustrate the practical application of AI Governance principles. By incorporating AI Governance frameworks, organizations can enhance data collection and analysis, measure ESG performance, identify emerging risks, mitigate biases, promote transparency, and ensure stakeholder engagement. This enables more reliable and actionable sustainability and climate risk disclosure, driving the transition toward a more sustainable future.

Business Action on AI

Case Study: Fujitsu's Approach to Ethical AI
Fujitsu, a global information and communication technology (ICT) company, and a BCSD Australia Member, is taking significant strides in the ethical AI space. The company's approach, much like Japan's more general stance, is centred around "human-centricity". This means that the impact of AI on people and society is at the forefront of their considerations.

Fujitsu has developed an AI Ethics Impact Assessment to assess the ethical impact of AI on people and society. This tool enables the company to understand the impact of AI before it is provided, and it can be used for trustworthy AI design and audit. The AI Ethics Impact Assessment is available free for public use, demonstrating Fujitsu's commitment to promoting ethical AI practices beyond its own operations.

The company has also released a variety of resources to support the implementation of ethical AI, including an AI Ethics Impact Assessment White Paper, an AI Ethics Impact Assessment Casebook, and an AI Ethics Impact Assessment Practice Guide. These resources provide guidance for developers, providers, and customers of AI systems, helping them to understand and address ethical issues related to AI.

Fujitsu's approach to ethical AI is a practical example of how businesses can take proactive steps to address the ethical implications of AI. By providing tools and resources for ethical AI development and use, Fujitsu is not only ensuring responsible practices within its own operations but also contributing to the broader effort to promote ethical AI practices. This case study highlights the potential for businesses to play a significant role in promoting ethical AI practices. It demonstrates that with the right tools and resources, businesses can effectively address the ethical implications of AI and contribute to the development of responsible AI practices.

Read more

Sustainable Business Australia Limited trading as
Business Council for Sustainable Development Australia
ABN 48 052 135 609
Level 27
20 Bond Street
SYDNEY NSW 2000
+ 61 2 8005 0780
bcsda@bcsda.org.au
www.bcsd.org.au

bcsd australia

Appendix A

Answers to the specific questions	
Definitions	
1. Do you agree with the definitions in this discussion paper? If not, what definitions do you prefer and why?	We agree with the definitions provided in the discussion paper. These definitions align with commonly accepted industry standards and terminology, enabling a shared understanding of key concepts related to AI governance. Consistency in definitions is crucial for effective communication and collaboration among stakeholders, including businesses, policymakers, and the public. Practical Example: In our discussions with our member companies, we have found that these definitions resonate with their understanding of AI and its governance. For instance, when developing their AI strategies, companies often refer to similar definitions to ensure clarity and alignment across their operations. This common understanding enables them to assess the potential risks and opportunities associated with AI technologies and make informed decisions regarding their implementation. Additionally, we have observed that these definitions are consistent with international frameworks and guidelines on AI governance, allowing Australian businesses to align their practices with global best practices. This alignment is essential for fostering international collaboration and ensuring interoperability of AI systems, particularly in cross-border operations and collaborations. By embracing these definitions, we can establish a robust foundation for addressing AI governance challenges, promoting responsible AI practices, and ultimately contributing to sustainable development goals, such as SDG 9 (Industry, Innovation, and Infrastructure) and SDG 17 (Partnerships for the Goals).
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?	Australia's existing regulatory approaches have made significant progress in addressing AI-related risks. However, some potential risks may not be adequately covered, and additional regulatory actions can be considered to mitigate these risks. Here are a few key areas where regulatory action could be beneficial: 1. **Lack of mandatory human intervention**: One potential risk from AI that may not be fully covered by Australia's existing regulatory approaches is the lack of mandatory human intervention throughout the design and operation of AI systems. This is particularly significant when it comes to Automated Decision-Making (ADM) systems, where decisions are made without human involvement. The absence of human oversight in these systems can lead to a variety of issues, including biased decision-making, lack of transparency, and potential misuse of AI. For instance, an AI

	system might make decisions based on patterns in the data that are discriminatory or unfair, but without human oversight, these issues might go unnoticed and uncorrected. To mitigate these risks, regulatory action could be taken to require a certain level of human involvement in the design and operation of AI systems, particularly those involved in decision-making processes. This could involve requirements for human review of AI decisions, especially in high-stakes contexts such as healthcare, finance, or criminal justice. Additionally, regulations could mandate transparency in AI systems, requiring explanations of how decisions were made and providing individuals with the opportunity to challenge decisions made by AI. By ensuring human oversight and validation of AI decisions, we can better manage the risks associated with AI and ADM, promoting more ethical and responsible use of these technologies. 2. **Algorithmic Bias and Fairness**: Addressing the potential bias and discrimination embedded in AI systems requires specific attention. Regulatory frameworks can mandate fairness assessments and audits for AI algorithms, ensuring that they do not perpetuate bias based on gender, race, or other protected characteristics. Guidelines on data collection and representation can be implemented to ensure representative and diverse training datasets. 3. **Explainability and Transparency:** AI systems' lack of explainability can limit accountability and public trust. Introducing regulations that require AI systems to provide understandable explanations for their decisions and actions can enhance transparency. Companies could be mandated to document the design, development, and deployment processes of their AI systems, allowing for better scrutiny and evaluation. This point is underscored by the findings of the Australian Community Attitudes to Privacy Survey 2020 prepared for the Office of the Australian Information Commissioner (OAIC). The survey revealed that 84% of respondents believed people should have a right to know if a decision affecting them is made using artificial intelligence technology, and 78% believed individuals should be told what factors and personal information are considered by the algorithm and how these factors are weighted. Further, if individuals are not informed about how they are being impacted by AI decisions, they will not have access to remedies. Therefore, we endorse the Privacy Act reform review recommendation that individuals be given the right to know how decisions are made. This would not only enhance transparency but also empower individuals to challenge decisions and seek remedies when necessary. 4. **Data Privacy and Security**: Strengthening regulations on data privacy and security is crucial to mitigate risks associated with AI. Frameworks like the General Data Protection Regulation (GDPR) can be expanded to cover AI-specific data protection requirements. Stricter controls on data sharing, anonymization, and informed consent can help safeguard individual privacy rights. 5. **Ethical Use of AI**: Regulatory frameworks should encourage businesses to adopt ethical AI practices. Guidelines and codes of conduct can be developed, addressing issues like the use of AI for social scoring, facial recognition, or other high-risk applications. Ethical review boards or committees can be established to evaluate the potential societal impacts of AI technologies. 6. **Ongoing Monitoring and Compliance**: Regular monitoring and auditing of AI systems can ensure ongoing compliance with regulations. This becomes increasingly important as machine learning and deep learning models continue to adapt and evolve to improve accuracy or effectiveness. However, these models can also have the opposite effect, where based on defective data or algorithms, they may produce inaccurate or biased results. Therefore, it would not be sufficient to simply intercept at a single point in the design process. Continuous oversight is necessary to ensure that as AI systems learn and adapt, they continue to operate within ethical and regulatory boundaries. Establishing independent bodies or agencies responsible for auditing AI systems and ensuring adherence to ethical and regulatory standards can be considered. Penalties or fines for non-compliance can incentivize organizations to prioritize responsible AI practices. Practical Example: The European Union's General Data Protection Regulation (GDPR) is an example of regulatory action that addresses data privacy and security risks, including those related to AI. The GDPR provides a comprehensive framework for the protection of personal data, establishing guidelines and requirements for businesses handling data. Australian regulatory authorities can study and adapt relevant aspects of the GDPR to further strengthen data protection in the context of AI. This could include provisions for ongoing monitoring and auditing of AI systems, particularly those that use machine learning and deep learning models.
3. Are there any further non-regulatory initiatives	There are several non-regulatory initiatives that the Australian Government could implement to support responsible AI practices in Australia. These initiatives can complement existing regulations and encourage businesses to adopt ethical and responsible approaches to AI. Here are a few examples:

the Australian Government could implement to support responsible AI practices in Australia? Please describe these and their benefits or impacts.	1. **Diversity Programs in the AI Industry**: One initiative could be the implementation of programs that encourage diversity within the AI industry. Generative AI technology has a high barrier to entry due to considerable educational and training requirements. This could lead to underrepresentation from diverse demographic groups, making it increasingly hard to remove bias and achieve fairness. By promoting diversity in the AI industry, we can ensure a wider range of perspectives and experiences are incorporated into the design and implementation of AI systems, which can help to reduce bias and improve fairness. 2. **Digital Inclusion Programs**: Another initiative could be programs to ensure that access to AI tools is fairly distributed. AI tools require considerable internet bandwidth, power, and suitable devices, which are not available or affordable to everyone. Regional Australians and older Australians particularly experience poorer digital inclusion and could suffer an accelerated digital divide due to this emerging technology. By implementing programs that improve digital inclusion, such as providing affordable internet access and digital devices, or offering training programs to improve digital literacy, the government can ensure that all Australians have the opportunity to benefit from AI technology. 3. **Awareness and Education Programs**: The government can establish awareness and education programs to promote understanding and awareness of responsible AI practices. These programs can target businesses, organizations, and the general public, providing guidance on ethical considerations, best practices, and the potential societal impacts of AI. Such initiatives can enhance AI literacy and encourage the adoption of responsible AI technologies. Benefits: Increased awareness and education would enable businesses to make informed decisions about AI implementation, leading to responsible use and addressing potential risks. It would also foster public trust and confidence in AI technologies. 2. **Voluntary Codes of Conduct**: The government can encourage the development and adoption of voluntary codes of conduct for AI. These codes can provide guidelines and standards for businesses to ensure ethical and responsible AI practices. They can cover areas such as transparency, fairness, accountability, and data privacy. Participating businesses could be recognized for their commitment to responsible AI. Benefits: Voluntary codes of conduct create a collaborative environment where businesses take ownership of responsible AI practices. They enable industry self-regulation, facilitate knowledge sharing, and promote a culture of responsible AI innovation. 3. **Public-Private Partnerships**: The government can establish partnerships with the private sector to foster responsible AI practices. Collaborative initiatives, such as research projects, innovation hubs, or funding programs, can support the development of ethical AI technologies. These partnerships can bring together industry expertise, academic research, and government resources to address societal challenges and ensure responsible AI adoption. Benefits: Public-private partnerships leverage collective knowledge, resources, and networks to drive responsible AI innovation. They enable collaboration on AI governance, ethics, and policy development, leading to shared best practices and enhanced industry-government cooperation. 4. **Certification and Accreditation Programs**: The government can introduce certification or accreditation programs that assess the ethical and responsible use of AI technologies. Businesses could voluntarily undergo assessments to demonstrate their adherence to specific standards and principles. Certifications could be granted to organizations that meet the criteria, providing recognition for their commitment to responsible AI practices. Benefits: Certification programs create incentives for businesses to adopt responsible AI practices, as they can enhance their credibility and reputation. They also enable consumers and stakeholders to make informed choices by identifying organizations that prioritize ethical AI deployment. These initiatives would benefit from the alignment with SDGs, particularly Goal 4 (Quality Education) and Goal 17 (Partnerships for the Goals).
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	We recognize the importance of coordinated AI governance across government agencies to foster responsible AI practices in Australia. Here are our suggestions for the coordination of AI governance: 1. **Establishment of a Central AI Governance Body**: The Australian Government could consider establishing a central body responsible for coordinating AI governance efforts across different government departments and agencies. This body would serve as a hub for knowledge sharing, collaboration, and policy development in the field of AI. It would provide a platform for stakeholders, including businesses, academia, and civil society, to contribute to the development of AI governance frameworks. Example: The United Kingdom's Centre for Data Ethics and Innovation (CDEI) serves as an independent advisory body to the UK government, providing guidance on ethical AI development. The CDEI's multi-

development and uptake of AI in Australia.	stakeholder approach has facilitated coordination and engagement between government, industry, and society, influencing the responsible development and uptake of AI technologies in the UK. 2. **Alignment of AI Strategies and Policies**: The coordination mechanisms should aim to align AI strategies and policies across government departments. This alignment would ensure a coherent approach to AI governance, minimizing inconsistencies and promoting standardized practices. It would facilitate the sharing of best practices, lessons learned, and insights from different sectors, enabling efficient and effective AI implementation. Example: The European Union's European AI Alliance brings together stakeholders from various sectors to exchange knowledge and develop recommendations for AI policy and governance. The collaboration among policymakers, businesses, and researchers has fostered alignment and coordination in AI strategies across EU member states. 3. **Information Sharing and Collaboration**: The coordination mechanisms should promote information sharing and collaboration between government agencies. Regular forums, workshops, and working groups can be organized to facilitate knowledge exchange, share emerging trends, and discuss policy challenges related to AI. Cross-departmental collaboration can lead to a comprehensive understanding of AI governance requirements and the development of holistic approaches. Example: In Canada, the Canadian Institute for Advanced Research (CIFAR) has established the Pan-Canadian Artificial Intelligence Strategy to coordinate AI research and development efforts across federal, provincial, and territorial governments. The collaboration enhances the sharing of expertise, fosters joint initiatives, and accelerates AI innovation in Canada. The goals of these coordination mechanisms would include: • **Consistency and Harmonization**: Coordinated AI governance would aim to establish consistent principles, standards, and guidelines across government agencies. This consistency would help avoid regulatory fragmentation and promote a unified approach to responsible AI practices. • **Knowledge Sharing and Capacity Building**: The mechanisms would facilitate knowledge sharing, best practices, and capacity building among government agencies. They would enable the dissemination of AI expertise, emerging trends, and lessons learned, enhancing the collective understanding of responsible AI governance. • **Effective Policy Development**: Coordinated AI governance would lead to more effective policy development by incorporating diverse perspectives, industry insights, and societal concerns. It would allow for comprehensive policy assessments, identification of regulatory gaps, and the formulation of evidence-based policies. • **Encouraging Industry Collaboration**: Coordinated AI governance would encourage collaboration between government and businesses. Engaging businesses in the policy-making process can ensure that regulations are practical, consider industry dynamics, and align with sustainable development objectives.
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?	We have identified several governance measures being taken or considered by other countries that are relevant, adaptable, and desirable for Australia in the context of AI. These examples showcase innovative approaches to AI governance: 1. **Canada's Algorithmic Impact Assessment (AIA)**: Canada is developing an AIA framework to assess the potential impacts of AI systems on individuals and society. The framework aims to enhance transparency, accountability, and fairness in AI decision-making processes. It requires organizations to conduct impact assessments and take corrective actions to mitigate biases and discrimination. Example: Australia could draw inspiration from Canada's AIA framework to develop a similar mechanism for assessing the impacts of AI systems. By mandating AI impact assessments, businesses would be required to evaluate and address potential risks, promoting responsible and fair AI practices. 2. **Finland's "Trustworthy AI" Approach**: Finland has adopted a holistic approach to AI governance, focusing on "trustworthy AI." Their guidelines emphasize human-centric and ethical AI, addressing issues such as transparency, fairness, and accountability. Finland encourages AI developers and users to adhere to these principles and offers support through a national AI program. Example: Australia could develop similar guidelines under the "trustworthy AI" framework to foster responsible AI practices. These guidelines would serve as a reference for businesses, promoting ethical considerations and reinforcing public trust in AI technologies. 3. **Singapore's Model AI Governance Framework**: Singapore has developed a Model AI Governance Framework that provides practical guidance for organizations to implement responsible AI practices. It covers areas such as fairness, transparency, explainability, and robustness. The framework encourages organizations to develop internal governance structures and processes to ensure ethical AI deployment. Example: Australia could

	adopt a similar model AI governance framework, customized to suit its unique context. This framework would assist organizations in implementing responsible AI practices, aligning with international best practices and contributing to building public trust. 4. **European Union's AI Act**: The European Union's proposed AI Act aims to regulate AI systems' risks and ensure their conformity with ethical principles. It establishes a legal framework that categorizes AI applications based on their risk levels and imposes specific requirements and obligations accordingly. The Act addresses issues like transparency, human oversight, and data governance. Example: Australia could study the European Union's AI Act to explore the feasibility of implementing a similar regulatory framework. This approach would enable a risk-based approach to AI governance, ensuring responsible deployment while considering the potential societal impacts of different AI applications. 5. **Establishing Multi-Regulator Sandboxes**: Both the EU and UK are exploring the concept of multi-regulator sandboxes. These sandboxes allow innovators and entrepreneurs to experiment with new AI products or services under enhanced regulatory supervision without the risk of fines or liability. This approach encourages innovation while ensuring that new technologies are developed and tested within a controlled environment that considers ethical and regulatory standards. Australia could consider establishing similar multi-regulator sandboxes to foster innovation in AI while ensuring responsible practices. The benefits of such an initiative would be manifold. For innovators and entrepreneurs, it provides a safe space to experiment and innovate without fear of regulatory repercussions. For regulators, it offers a way to keep pace with technological advancements and understand their implications in a controlled environment. For the public, it ensures that new AI technologies are developed and tested with regulatory oversight, enhancing trust in these technologies. 6. **Japan**: **Adopting an Agile Governance Approach**: Japan takes a voluntary, risk-based, AI-agnostic approach to the regulation of AI. They believe that regulation can struggle to keep up with the rapidly changing AI landscape, and therefore, an “agile governance” based approach could be more effective and responsive to change. This approach involves continuous monitoring and adaptation of policies to keep pace with technological advancements. Australia could consider a similar approach, which would allow for more flexibility and responsiveness in the face of rapid AI development. **Establishing Social Principles for AI**: The Japanese cabinet office has issued "Social Principles of Human-Centric AI". These principles are based on three basic philosophies: human dignity, diversity and inclusion, and sustainability. The goal is not to restrict the use of AI in order to protect these principles, but rather to realize them through AI. The principles include protective elements as well as guidelines for the active use of AI, such as education, fair competition, and innovation. Australia could consider establishing a similar set of principles to guide the development and use of AI in a way that aligns with societal values and goals. **Promoting Multi-Stakeholder Collaboration**: The Japanese government advocates for multi-stakeholder collaboration in AI governance. Stakeholders include not only experts in technology, law, economics, and management but also individuals and communities as the ultimate beneficiaries of AI governance. Australia could consider promoting similar collaboration, ensuring that a wide range of perspectives are considered in AI governance. In conclusion, the Australian Government could consider implementing non-regulatory initiatives inspired by global practices to support responsible AI practices. Here is a suggestion: **Promoting Regulatory and Disclosure Standardisation**: Much like in the space of Environmental, Social, and Governance (ESG), Australia could encourage regulatory and disclosure standardisation in the field of AI. This could be achieved through collaboration with global standardising bodies such as the International Organization for Standardization (ISO), AI4People, UNESCO, and the Organisation for Economic Co-operation and Development (OECD). These organisations have extensive experience in setting international standards and could provide valuable guidance in the development of AI regulations and disclosure requirements. The benefit of this initiative would be the creation of a consistent and comprehensive framework for AI practices. This would not only facilitate compliance for organisations operating in multiple jurisdictions but also enhance transparency and accountability in the AI industry. Furthermore, aligning with international standards could position Australia as a leader in responsible AI practices, attracting investment and fostering innovation in the sector.
Target areas	
6. Should different approaches apply to public and private sector use of AI	We recognize that different approaches may be warranted for the public and private sector use of AI technologies. While fundamental principles of responsible AI apply to both sectors, there are specific considerations that may necessitate differentiated approaches. Here is our thoughts on how the approaches could differ:

technologies? If so, how should the approaches differ?	1. **Public Sector**: The public sector should prioritize transparency and accountability in AI decision-making processes. It should provide clear explanations for automated decisions, ensuring that citizens understand how AI systems are used in public service delivery. In addition to this, the public sector should consider making public datasets available for the public good of solving societal and environmental issues where appropriate (e.g. Environmental -Economic Accounting Dashboard). The availability of these datasets could enable researchers, innovators, and the broader community to develop AI solutions that address pressing societal and environmental challenges. This would not only foster innovation but also promote the use of AI for public good. Establishing clear lines of accountability is crucial to maintain public trust. Example: The Australian Government's Digital Transformation Agency has developed the Digital Service Standard, which emphasizes transparency, user-centricity, and accountability in the delivery of digital services to citizens. This framework can be extended to incorporate specific guidelines for responsible AI use in the public sector, including the responsible use of public datasets. 2. **Private Sector**: The private sector often operates in a competitive landscape, where protecting intellectual property rights and maintaining a competitive advantage are essential. Balancing responsible AI practices with business confidentiality is crucial. Businesses may have proprietary algorithms and models that need protection, while still adhering to ethical considerations. However, this also needs to be balanced with concerns raised in the NSTC report regarding the anti-competitive nature of the ownership of large, rich datasets by certain entities or corporations. These entities may pose barriers to potential competitors entering or expanding into the market. There is a significant risk that the data and the investment in AI is too tightly controlled, which could stifle innovation and competition in the AI industry. Example: In the financial industry, banks and fintech companies deploy AI algorithms for risk assessment and fraud detection. While ensuring transparency and fairness, they also safeguard proprietary algorithms that give them a competitive edge. Striking a balance between responsible AI practices, protecting proprietary information, and ensuring a competitive and fair AI landscape is essential in this context. This could involve regulatory measures to prevent data monopolies and promote data sharing, while still protecting intellectual property rights. 3. **Collaboration and Standards**: Public-Private Collaboration: Collaboration between the public and private sectors is critical for addressing societal challenges associated with AI. Public-private partnerships can promote responsible AI development, encourage knowledge sharing, and ensure alignment with national priorities, including sustainable development goals. Example: In Australia, the government's Cooperative Research Centres (CRC) program fosters collaboration between industry, academia, and government to address industry challenges. A CRC focused on AI governance could facilitate collaboration between public and private sectors to develop responsible AI frameworks and share best practices. 4. **Risk Assessment and Regulation**: Risk-Based Approach: While both sectors should adopt a risk-based approach to AI governance, the specific risks and their impacts may differ. The public sector often deals with sensitive and high-stakes areas such as law enforcement and social services, warranting closer scrutiny and stricter regulation. The private sector, on the other hand, may face different risks related to data privacy, consumer trust, and market competition. Example: The healthcare sector's use of AI in public hospitals versus private clinics may have different risk profiles. Public hospitals may require additional safeguards to protect patient privacy and ensure fairness in resource allocation, while private clinics may focus on ensuring data security and complying with industry regulations.
7. How can the Australian Government further support responsible AI practices in its own agencies?	We believe the Australian Government can further support responsible AI practices in its own agencies through the following measures: 1. **Establish Clear Ethical and Responsible AI Guidelines**: The Australian Government should develop and implement clear guidelines for responsible AI practices within its agencies. These guidelines should address key principles such as transparency, fairness, accountability, and privacy. They should provide specific guidance on AI system development, deployment, and ongoing monitoring. Example: The United States Office of Management and Budget (OMB) has issued guidelines for federal agencies to ensure transparency, accountability, and public trust in AI. These guidelines provide a framework for responsible and ethical AI adoption within government agencies. 2. **Promote AI Skills and Education**: The government can invest in training programs and initiatives to build AI skills and literacy within its agencies. By equipping employees with the necessary knowledge and skills, agencies can make informed decisions regarding the implementation and oversight of AI systems. Collaborations with academia and industry can help develop AI training programs tailored to government needs. Example: The Singapore government has launched the AI for Everyone initiative, providing training programs to equip government officials with AI

	knowledge and skills. This initiative enables government agencies to effectively leverage AI technologies while ensuring responsible and ethical practices. 3. **Encourage Collaboration and Knowledge Sharing**: The Australian Government can foster collaboration and knowledge sharing among its agencies to promote responsible AI practices. Establishing forums, working groups, or communities of practice dedicated to AI governance and ethics can facilitate the exchange of best practices, lessons learned, and emerging trends across different government agencies. Example: The UK Government's Office for Artificial Intelligence and the UK AI Council act as central coordinating bodies, encouraging collaboration and knowledge sharing among government departments. These initiatives facilitate the sharing of expertise, lessons learned, and responsible AI practices across different agencies. 4. **Conduct Ethical Impact Assessments**: The government can integrate ethical impact assessments into the decision-making processes for AI adoption within its agencies. These assessments would evaluate potential ethical, social, and environmental impacts of AI systems, ensuring responsible and sustainable deployment. The findings of such assessments can guide policy development and influence procurement decisions. Example: The Netherlands' Ministry of the Interior and Kingdom Relations developed the "Ethics of AI in Government" toolkit, which includes ethical impact assessments for government agencies. This toolkit assists agencies in identifying and addressing ethical considerations in AI system deployment.
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.	We recognize that both generic solutions and technology-specific solutions play crucial roles in addressing the risks associated with AI. The circumstances in which each approach is most valuable can vary. Here are some examples: 1. **Generic Solutions**: Ethical Frameworks and Guidelines: Generic solutions in the form of ethical frameworks and guidelines are valuable in providing high-level principles and standards for responsible AI across different sectors and applications. These frameworks outline principles such as transparency, fairness, accountability, and privacy that are applicable to various AI technologies. Example: The Institute of Electrical and Electronics Engineers (IEEE) has developed the "Ethically Aligned Design" initiative, which provides a set of ethical guidelines for AI development and deployment. These guidelines offer generic principles applicable to diverse AI applications, promoting responsible practices across industries. 2. **Technology-Specific Solutions**: Algorithmic Bias Mitigation: Technology-specific solutions are valuable when addressing biases that are specific to certain AI applications or algorithms. These solutions focus on developing algorithms and methodologies to mitigate bias and ensure fairness in AI decision-making. Example: ProPublica's "Compas" algorithm used in the criminal justice system was found to exhibit racial bias. Technology-specific solutions were implemented to address this bias by refining the algorithm to minimize disparate impact and enhance fairness in predicting recidivism rates. 3. **Context-Specific Regulations**: Data Privacy and Security: Context-specific regulations provide tailored solutions to address specific risks associated with AI technologies. These regulations focus on areas such as data privacy, security, and consent requirements, recognizing the unique challenges posed by AI systems. Example: The European Union's General Data Protection Regulation (GDPR) incorporates specific provisions addressing the processing of personal data in the context of AI. These regulations mandate transparency, user rights, and data protection practices, ensuring responsible AI use while safeguarding individual privacy. 4. **Sector-Specific Guidelines**: Industry Standards and Best Practices: Sector-specific guidelines provide targeted solutions to address risks and challenges unique to particular industries. These guidelines define industry standards and best practices for responsible AI deployment, tailored to the specific requirements and contexts of those industries. Example: The Financial Stability Board's "Principles for Responsible AI in the Financial Sector" provide guidance to financial institutions on the responsible use of AI. These principles address risks associated with data privacy, algorithmic transparency, and explainability in the financial industry. It is important to note that a combination of generic and technology-specific solutions is often necessary for comprehensive AI risk mitigation. Generic solutions establish overarching ethical frameworks and guidelines, while technology-specific solutions address specific risks and challenges unique to certain AI applications.

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? b. mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.	We recognize the importance of transparency in the AI lifecycle to mitigate potential risks, enhance public trust, and improve confidence in AI. Here are our thoughts on the criticality and value of transparency, as well as implementing transparency requirements across the private and public sectors: **a. Criticality and Value of Transparency:** • **During Data Collection and Processing**: Transparency in data collection and processing is critical to address concerns related to privacy, consent, and potential biases. Making the data collection process transparent helps individuals understand how their data is being used and ensures compliance with privacy regulations. • **Algorithmic Decision-Making**: Transparency in algorithmic decision-making processes is crucial, especially when AI systems impact individuals' rights, opportunities, or access to services. It enables individuals to understand how decisions are made, assess potential biases, and seek recourse if necessary • **High-Stakes Applications**: Transparency is particularly valuable in high-stakes applications such as healthcare, finance, and criminal justice. It provides stakeholders with insights into the underlying processes, enabling independent audits, accountability, and fairness. • **Explainability and Interpretability**: Transparency in AI models and outputs is valuable to facilitate understanding and trust. Explainable AI systems that provide understandable explanations for their decisions can enhance user acceptance, regulatory compliance, and societal trust. Practical Example: Google's Model Cards for Model Transparency and Data Sheets for Data Sets initiative provides documentation on AI models and data sets, including details on model performance, limitations, and potential biases. This transparency initiative aims to enhance understanding, trust, and accountability in AI systems. **b. Mandating Transparency Requirements**: • **Framework for Transparency**: The Australian Government can mandate a framework for transparency, outlining the specific information and disclosures required from both private and public sectors. This framework can include details on data collection practices, algorithmic decision-making processes, and the impacts of AI systems on individuals and society. • **Clear Disclosure Obligations**: Mandated transparency requirements should specify the information that organizations need to disclose, such as data sources, data handling practices, and the use of AI in decision-making. These obligations can extend to the explanation of the reasoning behind automated decisions when they significantly impact individuals. • **Industry-Specific Guidelines**: Transparency requirements can be implemented through industry-specific guidelines that outline best practices and reporting standards. These guidelines can be developed collaboratively with industry stakeholders, ensuring practicality and alignment with sector-specific needs. • **Independent Auditing and Certification**: The government can establish mechanisms for independent auditing and certification of AI systems to verify transparency claims made by organizations. Third-party audits and certifications provide assurance to the public and stakeholders regarding adherence to transparency requirements. Practical Example: The Responsible AI Certification program developed by the World Economic Forum provides a framework for independent auditing and certification of AI systems. It assesses various dimensions, including transparency, and enables organizations to demonstrate their commitment to responsible AI practices.
10. Do you have suggestions for: a. Whether any high-risk AI applications or technologies should be banned completely? b. Criteria or requirements to identify AI applications or	We agree on the importance of carefully evaluating high-risk AI applications and technologies. While a blanket ban may not be the most effective approach, certain circumstances may warrant restrictions or regulatory measures to mitigate potential risks. Here are our suggestions: a. **High-Risk AI Applications or Technologies**: • Instead of advocating for complete bans, we recommend a risk-based approach to determine the level of regulation and oversight required for high-risk AI applications. Prohibitions should be considered for applications that pose severe and irreversible harm to individuals, society, or the environment, where the risks outweigh the potential benefits. Example: Autonomous weapons systems, which have the potential to cause

technologies that should be banned, and in which contexts?	significant harm and raise ethical concerns, could be subject to strict regulation or international agreements to prevent their development and use. **b. Criteria or Requirements for Identifying Banned AI Applications**: • **Ethical Considerations**: Applications that violate fundamental ethical principles, such as human rights, privacy, or fairness, should be subject to scrutiny. Criteria could include assessments of potential biases, discriminatory outcomes, or infringements on personal privacy. In addition to these, ethical considerations should also take into account the potential for significant degradation of public trust, widespread misinformation or disinformation, interference with democratic processes, and social discourse. There should also be concern for applications that encourage self-harm or foster a negative sense of self. AI applications that have the potential to cause these types of harm could have profound societal impacts, and it's crucial that these considerations are part of the ethical assessment of AI systems. This would ensure a more comprehensive approach to responsible AI, taking into account not just the direct impacts on individuals, but also the broader societal implications. • **Safety and Security Risks**: Applications that pose significant safety or security risks to individuals, communities, or critical infrastructure should be carefully evaluated. Criteria could include the potential for physical harm, cybersecurity vulnerabilities, or system failures. • **Social and Environmental Impacts**: AI applications that have substantial negative social or environmental impacts should be considered for restrictions. Criteria could include risks of job displacement without adequate social support, exacerbation of societal inequalities, or environmental degradation. In addition to these, it's important to consider that AI applications could also significantly accelerate existing issues or inequalities. For instance, AI systems that are trained on biased data could perpetuate and even amplify these biases, leading to unfair outcomes in areas such as hiring, lending, or law enforcement. Similarly, AI applications could exacerbate environmental issues if they are designed without consideration for their energy consumption or other environmental impacts. Therefore, the assessment of AI applications should take into account not only their direct impacts, but also their potential to accelerate existing societal and environmental issues. This would ensure a more comprehensive approach to responsible AI, taking into account the broader implications of these technologies. • **Public and Stakeholder Engagement**: Transparent processes that include public and stakeholder engagement can help identify applications that raise significant concerns. The input of diverse perspectives, including civil society, academia, and impacted communities, should inform decision-making. In addition to this, we encourage critical reflection earlier in the process. Introspection can be less effective, despite best efforts, those developing algorithms will be prone to bias and intellectual lock-in. People cannot grade their own work, or that of people they know really well. Therefore, the involvement of critical audiences that oppose algorithms and point out their shortcomings will be essential in the development of quality, responsible AI. This could involve independent audits, peer reviews, or public consultations to scrutinize AI systems and identify potential issues. Practical Example: The European Union's proposed AI Act classifies AI systems into different risk categories based on potential harm, such as unacceptable risks, high risks, and limited risks. The Act outlines criteria and requirements for each category, providing a regulatory framework for identifying and addressing high-risk AI applications. This framework, combined with early and ongoing critical reflection, could help ensure that AI applications are developed and used responsibly. It is crucial to strike a balance between enabling innovation and addressing risks associated with AI applications. Regulatory measures should be proportionate, evidence-based, and consider potential unintended consequences.
11. What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?	We recognize that building public trust in AI deployment is essential to encourage broader adoption and utilization of AI technologies. Here are some ideas of initiatives and government actions that can contribute to increasing public trust: 1. **Transparent and Ethical AI Practices**: o Clear Communication: Encourage businesses and government agencies to communicate openly and transparently about their AI systems, including how they are developed, deployed, and used. This includes providing understandable explanations of AI decision-making processes and being transparent about data collection, privacy safeguards, and potential impacts. o Ethical Guidelines: Establish clear ethical guidelines or codes of conduct that encourage responsible AI practices, ensuring fairness, accountability, and privacy protection. These guidelines should be developed collaboratively with stakeholders and should address concerns specific to different industries and applications. Example: The Partnership on AI, a collaborative platform involving leading

	technology companies, has developed ethical guidelines to promote responsible AI deployment. These guidelines emphasize transparency, fairness, and accountability, which can enhance public trust in AI technologies. 2. **Robust Data Privacy and Security Measures**: Strengthened Data Protection: Implement and enforce robust data privacy and security regulations to ensure that individuals' personal data is handled safely and securely throughout the AI lifecycle. Provide clear guidelines and mechanisms for obtaining informed consent, ensuring data anonymization, and enabling individuals to have control over their data. Example: The General Data Protection Regulation (GDPR) in the European Union sets stringent data protection standards and empowers individuals with rights over their personal data. Such regulations inspire public confidence in the responsible use of AI technologies. 3. **Independent Auditing and Certification**: Third-Party Verification: Establish mechanisms for independent auditing and certification of AI systems to provide assurance of adherence to ethical and responsible AI practices. Independent audits can help verify compliance with transparency, fairness, and privacy standards, instilling trust in AI deployments. Example: The Responsible AI Certification program developed by an independent organization or consortium can verify that AI systems and practices meet predefined ethical and responsible criteria. Certification can enhance public trust by providing a recognized standard for responsible AI deployment. 4. **Public Engagement and Collaboration**: Participatory Approach: Engage the public in decision-making processes related to AI deployment through consultations, public hearings, and citizen panels. Seek public input on policies, regulations, and ethical considerations to ensure that AI deployment aligns with societal values and concerns. Example: The Finnish government's "AI4People" initiative engages citizens, experts, and stakeholders in discussions on AI ethics, transparency, and accountability. This participatory approach fosters trust by involving the public in shaping AI policies and practices. 5. **Education and Awareness Programs**: AI Literacy and Education: Develop educational programs to enhance AI literacy and awareness among the general public. Promote understanding of AI capabilities, benefits, risks, and limitations to empower individuals to make informed decisions regarding AI technologies. Example: The Australian government, in collaboration with educational institutions and industry stakeholders, can launch awareness campaigns, workshops, and online resources to promote AI literacy and raise public awareness of its potential and responsible use.
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?	We recognize that banning high-risk activities such as social scoring or facial recognition technology in certain circumstances can have both positive and negative impacts on Australia's tech sector and trade relationships with other countries. Here are our thoughts on some potential impacts to consider: 1. **Impact on the Tech Sector**: ○ **Innovation and Research**: Banning high-risk activities may lead to a temporary setback in specific areas of technological development. However, it can also encourage businesses to shift their focus towards more responsible and ethical alternatives, fostering innovation in areas that align with societal values and needs. ○ **Trust and Reputation**: Banning high-risk activities can help build trust and enhance the reputation of Australia's tech sector. Demonstrating a commitment to responsible practices can attract investments, partnerships, and collaborations with businesses that prioritize ethical considerations. ○ **Market Opportunities**: Proactively addressing high-risk activities can create market opportunities for businesses offering alternative solutions that are aligned with responsible AI practices. By developing and promoting such alternatives, the tech sector can tap into growing global demand for trustworthy and ethical technologies. 2. **Impact on Trade and Exports**: ○ **Compliance with International Standards**: Banning high-risk activities aligns with international standards and regulations focused on protecting individual rights, privacy, and human dignity. This can strengthen Australia's position as a responsible player in the global tech market, enhancing trade relationships with countries that prioritize similar values. ○ **Market Access Challenges**: Some countries may have different regulatory approaches or continue to utilize high-risk activities. This could lead to market access challenges for Australian tech companies that strictly adhere to responsible AI practices. Engaging in bilateral or multilateral discussions to harmonize regulations and ensure fair trade practices becomes crucial in such scenarios.

	○ **Export Opportunities**: Banning high-risk activities can present export opportunities for Australian tech companies that specialize in alternative solutions or more responsible technologies. International markets seeking ethical and responsible AI applications may prefer to engage with Australian businesses that align with their values. Practical Example: The European Union's General Data Protection Regulation (GDPR) includes stringent regulations on data privacy and individual rights. While initially causing adjustments and compliance challenges for businesses, it has also presented market opportunities for companies offering privacy-enhancing technologies and services. It is important to balance the potential impacts on the tech sector and trade relationships while upholding responsible AI practices.
13. What changes (if any) to Australian conformity infrastructure might be required to support assurance processes to mitigate against potential AI risks?	We recognize the need for changes to Australian conformity infrastructure to support assurance processes and mitigate potential risks associated with AI. Here are our thoughts on some possible changes that may be required: 1. **Standards and Certification**: ○ **Development of AI-specific Standards**: Australian conformity infrastructure can be enhanced by developing specific standards and guidelines that address the unique risks and challenges of AI technologies. These standards can cover areas such as data governance, algorithmic transparency, fairness, and accountability. Example: The International Organization for Standardization (ISO) has developed ISO/IEC 27018, which provides guidelines for protecting personal data in the cloud computing environment. Similar standards can be developed or adapted to address AI-specific risks and requirements. ○ **Certification Programs**: Introduce certification programs to assess and verify the conformity of AI systems with established standards. Independent certification bodies can evaluate AI technologies and issue certificates or labels to demonstrate compliance with responsible AI practices. Example: The Trustworthy AI Certification by an independent certification body can assess AI systems against predefined criteria, ensuring adherence to ethical and responsible practices. This certification can provide assurance to businesses and consumers regarding the trustworthiness of AI technologies. 2. **Regulatory Framework**: **Regulatory Updates**: Revise existing regulations or introduce new regulations to incorporate specific requirements for AI technologies. These regulations can address issues such as data protection, explainability, bias mitigation, and human oversight. They should be regularly updated to keep pace with technological advancements and emerging risks. Example: The European Union's AI Act proposal aims to regulate AI systems based on their risk levels, imposing specific requirements and obligations. Such regulatory frameworks can provide clear guidelines for AI developers, users, and conformity assessment bodies. 3. **Conformity Assessment Processes**: Third-Party Audits: Strengthen conformity assessment processes by introducing third-party audits and inspections to evaluate the compliance of AI systems with regulations and standards. These audits can ensure that AI technologies meet predefined requirements and mitigate potential risks. Example: The United States Federal Communications Commission (FCC) conducts audits and inspections to ensure compliance with regulations in the telecommunications industry. Similar assessment processes can be established for AI systems to verify conformity with responsible AI practices. 4. **Collaborative Partnerships**: ○ Collaboration with Industry and Research Institutions: Foster collaboration between conformity assessment bodies, industry stakeholders, and research institutions to develop and share best practices for AI risk mitigation. This collaboration can promote knowledge exchange, innovation, and continuous improvement in conformity infrastructure. Example: The European AI Alliance, a platform for collaboration between stakeholders, including conformity assessment bodies, has facilitated discussions and knowledge sharing on AI regulation and conformity assessment practices.
Risk-based approaches	
14. Do you support a risk-based approach for addressing potential AI	We support a risk-based approach for addressing potential AI risks. A risk-based approach enables a nuanced and targeted assessment of AI applications, allowing resources to be allocated based on the level of risk they pose. Here's why a risk-based approach is beneficial: 1. **Tailored Risk Assessment**: A risk-based approach allows for a systematic evaluation of AI applications, considering their potential impacts on individuals, society, and the environment. It enables the identification and prioritization of high-risk areas, ensuring that mitigation efforts are

risks? If not, is there a better approach?	focused where they are most needed. To be clear, this risk-based approach should be applied throughout the entire development lifecycle of the AI, as per Table 1 in the Rapid Response Information Report. This would help to capture different risks as they occur at various stages of the AI development process, from the initial design and data collection stages through to deployment and post-deployment monitoring. By integrating risk assessment and mitigation into every stage of the AI development lifecycle, we can ensure that potential issues are identified and addressed as early as possible, reducing the likelihood of negative impacts and promoting the responsible development and use of AI. 2. **Efficient Resource Allocation**: By prioritizing high-risk applications, a risk-based approach optimizes the allocation of resources, enabling organizations and regulators to concentrate efforts on addressing the most significant potential harms. This approach ensures that limited resources are used effectively to manage and mitigate risks. 3. **Flexibility and Adaptability**: AI technologies and their associated risks evolve rapidly. A risk-based approach provides flexibility to adapt regulatory measures and mitigation strategies in response to changing circumstances. It allows for continuous monitoring, reassessment, and adjustment of risk management practices as new information and technologies emerge. 4. **Proportional Regulation**: A risk-based approach avoids imposing unnecessary burdens on low-risk AI applications and enables the application of proportionate regulatory measures. This allows businesses to innovate and deploy AI technologies responsibly while maintaining a focus on mitigating risks that genuinely require intervention. Practical Example: The European Union's proposed AI Act adopts a risk-based approach, classifying AI systems into different risk categories based on potential harm. The Act imposes stricter requirements for high-risk applications, while proportionate obligations are applied to lower-risk systems. This approach balances risk mitigation and innovation, ensuring responsible AI deployment. While a risk-based approach is generally effective, it is important to consider complementary approaches that address specific concerns or contexts. These may include sector-specific regulations, ethical guidelines, or technology-specific assessments, depending on the nature of the risks involved. However, a risk-based approach provides a foundation for decision-making, enabling the efficient allocation of resources and focusing on the most critical AI risks.
15. What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome?	We recognize both the benefits and limitations of a risk-based approach in addressing AI risks. Here are our observations of the main benefits and limitations, as well as strategies to overcome those limitations: **Benefits of a Risk-Based Approach**: 1. **Targeted Mitigation Efforts**: A risk-based approach allows for the prioritization of resources and efforts on high-risk AI applications, ensuring that mitigation measures are focused where they are most needed. This targeted approach maximizes the effectiveness of risk management efforts. 2. **Proportional Regulation**: A risk-based approach enables the application of proportionate regulatory measures, avoiding unnecessary burdens on low-risk AI applications. This promotes innovation while ensuring that appropriate safeguards are in place for higher-risk applications. 3. **Flexibility and Adaptability**: The risk-based approach provides flexibility to adapt to changing circumstances and evolving technologies. It allows for continuous monitoring, reassessment, and adjustment of risk management practices as new information emerges. **Limitations of a Risk-Based Approach**: 1. **Uncertainty and Subjectivity**: Assessing and quantifying risks associated with AI technologies can be challenging due to uncertainties and subjectivity in risk estimation. Determining risk levels may involve subjective judgments and rely on available data, which may be limited or incomplete. 2. **Rapid Technological Advancements**: The rapid pace of technological advancements in AI can outpace regulatory frameworks based on risk assessment. As a result, emerging risks and novel applications may not be adequately addressed within existing risk-based approaches. **Overcoming Limitations**: 1. **Improved Data and Research**: To address uncertainties and subjectivity, efforts should be made to enhance data collection, research, and collaboration between industry, academia, and government. This can lead to better risk modelling, more accurate assessments, and a stronger evidence base for risk evaluation. 2. **Agile Regulatory Frameworks**: To address the rapid advancement of AI technologies, regulatory frameworks should be designed to be agile and adaptable. Regular updates and iterative improvements can ensure that risk-based approaches keep pace with technological developments.

	3. **Stakeholder Engagement**: Engaging stakeholders, including businesses, civil society organizations, and experts, in the risk assessment process can enhance the robustness and credibility of risk-based approaches. Collaborative efforts can help identify emerging risks, ensure diverse perspectives are considered, and foster trust in the regulatory process. Practical Example: The Australian Therapeutic Goods Administration (TGA) adopts a risk-based approach in regulating medical devices. It considers the potential risks associated with different classes of devices, ensuring that higher-risk devices undergo more stringent regulatory scrutiny. By addressing limitations through improved data, agile regulatory frameworks, and stakeholder engagement, the benefits of a risk-based approach can be maximized.
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?	We consider that the suitability of a risk-based approach may vary across sectors, AI applications, and organizations based on factors such as organization size, AI maturity, and available resources. Here is our assessment overview of how these factors can influence the effectiveness of a risk-based approach: 1. Organization Size: ○ **Large Organizations**: Larger organizations typically have more resources and capacity to implement comprehensive risk management practices. A risk-based approach can be well-suited for these organizations as they can allocate dedicated teams or departments to assess and manage AI risks effectively. ○ **Small and Medium-sized Enterprises (SMEs)**: SMEs may have limited resources and expertise to conduct complex risk assessments. In such cases, simplified risk assessment frameworks and guidance tailored to their specific sector or AI application can help SMEs navigate AI risks effectively. Practical Example: The Australian banking sector, comprising both large banks and smaller financial institutions, can benefit from a risk-based approach. Larger banks have dedicated risk management teams to assess and address AI risks comprehensively, while smaller institutions can adopt sector-specific guidelines or frameworks that provide targeted guidance for their risk assessments. 2. **AI Maturity**: ○ **Early-Stage AI Applications**: In the early stages of AI implementation, the focus is often on understanding the technology's capabilities and risks. A risk-based approach can be particularly useful during this phase, as it helps identify and manage the most critical risks while allowing organizations to learn from the deployment process. ○ **Mature AI Applications**: As AI applications become more advanced and integrated into core operations, organizations may have accumulated experience and data on specific risks. In such cases, a risk-based approach can be refined and tailored based on lessons learned and feedback mechanisms from ongoing AI operations. Practical Example: A manufacturing company that implements AI-driven predictive maintenance systems can adopt a risk-based approach during the initial implementation to identify critical risks related to equipment failures. As the system matures, the company can refine the risk assessment process based on historical data and feedback from maintenance operations. 3. **Resources and Expertise**: ○ Adequate Resources: Organizations with sufficient resources and expertise can employ comprehensive risk assessment methodologies. They can invest in dedicated AI risk management teams, conduct in-depth assessments, and develop tailored risk mitigation strategies based on their specific context and AI applications. ○ Limited Resources: Organizations with limited resources may focus on adopting simplified risk assessment frameworks and leveraging industry best practices and guidelines. Collaborative initiatives, knowledge-sharing platforms, and government support can help smaller organizations access the necessary resources and expertise to navigate AI risks effectively. Practical Example: The Australian healthcare sector, comprising both large hospital networks and smaller medical practices, can adopt a risk-based approach to AI implementation. Larger healthcare institutions can invest in dedicated AI risk management teams, while smaller practices can leverage sector-specific guidelines to conduct risk assessments and adopt appropriate mitigation measures. In summary, a risk-based approach can be tailored to suit different sectors, AI applications, and organizations based on factors such as organization size, AI maturity, and available 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?	We support the inclusion of the elements presented in Attachment C as part of a risk-based approach for addressing potential AI risks. These elements are essential in promoting safe and robust practices and increasing community trust and confidence in AI. Here is our assessment of each element: 1. **Impact Assessments**: Conducting impact assessments is crucial to identify and mitigate potential risks associated with AI systems. Peer-reviewed external assessments, particularly for high-risk applications, enhance transparency and provide independent validation of risk management efforts. 2. **Notices**: Informing users when automation or AI is utilized in ways that materially affect them is essential. Providing notices ensures transparency and empowers individuals to seek reviews of decisions or actions made by AI systems, fostering trust and accountability. 3. **Human in the Loop/Oversight Assessments**: In certain circumstances, human involvement or oversight is necessary to minimize potential risks and maintain public trust and confidence. Assessments should consider decision complexity, discretion, potential damage, and required expertise to determine when human intervention is appropriate. 4. **Explanations**: Explanations contribute to transparency and understanding, building public trust. Clear explanations help individuals affected by AI decisions comprehend the factors that influenced outcomes, fostering accountability and mitigating concerns regarding bias or lack of interpretability. 5. **Training**: Adequate training for employees involved in AI design, implementation, and oversight is crucial. Training should encompass understanding potential risks, mitigation strategies, and the ability to explain and supervise AI operations. Increased training depth should align with the level of potential risk. 6. **Monitoring and Documentation**: Ongoing monitoring ensures AI systems operate as intended and identifies any unintended impacts or biases. More frequent and rigorous monitoring is necessary for higher-risk applications. Documentation facilitates better understanding of risks, mitigation measures, and accountability for decision-makers involved in AI deployment. Practical Example: In the financial sector, the Responsible AI Framework developed by a major Australian bank includes impact assessments, notices, human oversight, explanations, training programs, and monitoring and documentation as part of their risk-based approach to AI governance. This framework aims to ensure responsible and ethical AI practices within the organization.
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?	We recognize the importance of incorporating an AI risk-based approach into existing assessment frameworks and risk management processes to streamline operations and reduce potential duplication. Here are some strategies that we consider could achieve this integration effectively: 1. **Alignment with Privacy Assessment Frameworks**: Incorporate AI-specific considerations into existing privacy assessment frameworks to ensure comprehensive evaluation of AI-related privacy risks. This integration helps streamline assessments by leveraging established privacy assessment processes while addressing the unique challenges posed by AI technologies. Example: The Office of the Australian Information Commissioner (OAIC) provides guidance on conducting Privacy Impact Assessments (PIAs). Integrating AI-specific considerations within the existing PIA framework enables organizations to evaluate privacy risks associated with AI applications systematically. 2. **Synergy with Existing Risk Management Processes**: Integrate AI risk assessment within broader enterprise risk management frameworks to avoid duplication of efforts. Aligning AI risk assessment with existing risk management processes allows for a holistic approach to identify, evaluate, and manage risks across various domains. Example: A large retail corporation may have established risk management processes that encompass areas such as operational risks, cybersecurity, and compliance. Integrating AI risk assessment within this framework ensures that AI-related risks are considered alongside other enterprise risks, avoiding duplication of efforts. 3. **Collaboration and Knowledge Sharing**: Foster collaboration and knowledge sharing between different departments or functions within an organization to streamline assessments and reduce duplication. Encourage cross-functional teams, including privacy, legal, compliance, and AI specialists, to work together in conducting risk assessments and identifying common areas of concern. Example: A technology company can establish a multidisciplinary working group comprising privacy officers, legal experts, and AI specialists. This collaborative approach enables the sharing of insights, identification of overlapping risks, and efficient allocation of resources for risk assessments. 4. **Streamlined Documentation and Reporting**: Develop streamlined documentation and reporting processes that integrate AI risk assessments with existing frameworks. This ensures that information related to AI risks, mitigation strategies, and compliance efforts is captured within centralized

	reporting systems, reducing administrative burden and duplication. Example: A financial institution can enhance its risk reporting mechanisms by incorporating specific sections dedicated to AI-related risks and mitigation measures within existing reporting templates. This streamline reporting processes and enables stakeholders to access comprehensive risk information across various domains. By incorporating AI risk-based approaches into existing frameworks, organizations can streamline assessments, reduce duplication, and foster efficient risk management practices.
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 recognize the need to apply a risk-based approach to general purpose AI systems, such as large language models (LLMs) or multimodal foundation models (MFMs). Here's how a risk-based approach could be applied to these systems: 1. **Risk Identification and Categorization**: Identify potential risks associated with general purpose AI systems like LLMs and MFMs. This includes risks such as biased outputs, misinformation propagation, data privacy breaches, and unintended consequences arising from the scale and complexity of these models. Example: In the context of LLMs, potential risks may include the generation of biased or harmful content, spreading misinformation, or amplifying existing societal biases. Identifying these risks allows for targeted mitigation efforts. 2. **Risk Assessment**: Conduct risk assessments to evaluate the likelihood and potential impact of identified risks. Assessments should consider factors such as the scope of deployment, data inputs, model architecture, and potential societal implications. Example: A technology company deploying an LLM for content generation can assess the risk of biased outputs by evaluating the training data sources, potential biases in the training process, and the model's sensitivity to different inputs. 3. **Mitigation Strategies**: Develop and implement mitigation strategies to address identified risks. These may include measures such as data pre-processing to reduce biases, ongoing monitoring of model outputs, user feedback mechanisms, and algorithmic transparency initiatives. Example: A social media platform utilizing an MFM for image recognition can implement mitigation strategies by deploying bias detection algorithms to identify and rectify biases in the model's output. They can also provide clear user guidelines and reporting mechanisms to address any biased or harmful outputs. 4. **Continuous Monitoring and Evaluation**: Regularly monitor and evaluate the performance and impact of general-purpose AI systems. This includes ongoing assessment of risks, feedback loops with users and stakeholders, and adaptation of mitigation strategies based on emerging challenges and societal needs. Example: A research institute developing an MFM for medical diagnosis can continuously monitor the system's performance, evaluate its impact on patient outcomes, and engage healthcare professionals for feedback to ensure ongoing risk mitigation.
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?	We believe that a risk-based approach for responsible AI should involve a combination of voluntary and regulatory measures. Here is our perspective on the voluntary and regulatory aspects and their application to different entities: 1. **Voluntary Measures**: Voluntary measures can encourage proactive and responsible behaviour among organizations and foster innovation. They can take the form of industry-led frameworks, guidelines, and best practices that enable businesses to assess and mitigate AI risks based on their unique circumstances. Example: Collaborative initiatives, such as industry consortia, can develop voluntary frameworks for responsible AI, encouraging organizations to adopt risk-based approaches and share knowledge and experiences. The Partnership on AI, an organization formed by leading technology companies, is an example of such a collaborative effort. 2. **Regulatory Measures**: Regulatory intervention is essential to establish a baseline of responsible AI practices and ensure accountability across the industry. Regulations can provide a level playing field, set minimum standards, and address systemic risks that voluntary measures alone may not adequately address. Example: The European Union's proposed AI Act is an instance of regulatory measures. It aims to establish clear rules and obligations for AI developers and deployers, addressing high-risk AI applications, transparency requirements, and conformity assessments. **Application to Public and Private Organizations**: A risk-based approach should apply to both public and private organizations to ensure a comprehensive and consistent approach to responsible AI. Public sector organizations, as custodians of public interest, have a responsibility to adopt risk-based practices when developing and deploying AI systems. Similarly, private sector organizations must consider the potential risks and societal impacts of AI technologies they develop or deploy.

	Example: The Australian Government's Digital Transformation Agency (DTA) has developed the Digital Service Standard, which provides guidance for both public and private organizations on delivering digital services, including considerations related to AI ethics, privacy, and security. **Application to Developers and Deployers**: Both AI developers and deployers should be subject to a risk-based approach to ensure end-to-end accountability and responsible AI practices. Developers play a crucial role in designing and building AI systems with appropriate risk mitigation measures. Deployers, on the other hand, are responsible for the safe and responsible use of AI systems in real-world applications. Example: The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems provides guidelines for both developers and deployers to promote ethical and responsible AI practices. These guidelines outline principles and recommendations for the entire lifecycle of AI systems. In summary, a risk-based approach for responsible AI should involve a combination of voluntary measures and regulatory interventions. It should apply to both public and private organizations, as well as encompass both AI developers and deployers.