‘SHAPING THE FUTURE: TOWARDS SAFE AND RESPONSIBLE AI REGULATION IN AUSTRALIA’

PROFESSOR ROCKY SCOPELLITI
FUTUROLOGIST

ABSTRACT

Australia's existing regulatory approaches recognise the risks of AI, but further action is necessary to effectively address these challenges. Customised regulations and guidelines are essential to ensure fairness, transparency, and accountability, thereby safeguarding individuals and society within the rapidly evolving AI landscape. By implementing preventive measures, context-specific responses, and collaborative governance initiatives, Australia can promote the safe and responsible use of AI while mitigating potential risks.

A multifaceted strategy is needed to mitigate the potential risks associated with AI and ADM. The Australian Government should adopt a risk-based approach that tailor’s regulatory scrutiny and oversight based on the potential harm and impact of AI systems. This approach allows for a flexible regulatory environment, ensuring that high-risk AI applications receive the necessary scrutiny and safeguards.

Transparency and accountability are vital for building public trust and confidence in AI systems. The government should mandate transparency requirements for both public and private sectors, ensuring individuals are informed when AI systems materially affect them. Impact assessments should be conducted to evaluate potential risks, and their results should be made publicly available to foster transparency and enable external scrutiny. Peer review of impact assessments can enhance their robustness, particularly in high-risk contexts.

Involving humans in the loop or establishing oversight mechanisms can minimise risks and build public trust. The decision to involve humans should be based on factors such as decision complexity, level of discretion, potential damage from incorrect decisions, and required specialist knowledge. Striking the right balance between human involvement and automated processes is crucial, acknowledging instances where human intervention may not be feasible or desirable due to efficiency, speed, scale, or minimal impact on individuals.

Promoting responsible AI practices extends beyond regulation. The Australian Government should consider non-regulatory initiatives that foster collaboration between academia, industry, and civil society. Developing best practices, sharing knowledge and resources, and promoting ethical AI education and training programs are important steps. Government-led initiatives can increase public awareness and understanding of AI technologies, promoting responsible AI literacy among citizens.

Leveraging existing conformity infrastructure and assessment frameworks is another valuable approach. Integrating AI risk assessments within established frameworks like privacy assessments and risk management processes can streamline assurance processes, reduce duplication, enhance efficiency, and ensure comprehensive coverage of potential risks.

By addressing the questions posed in its AI Discussion Paper through a robust regulatory and governance framework, the Australian Government can effectively mitigate the potential risks associated with AI and ADM while increasing public trust and confidence. Achieving the right balance between regulation and voluntary initiatives, implementing transparency requirements, involving humans appropriately, and embracing non-regulatory measures will contribute to the responsible development and use of AI technologies in Australia.

Authors Note: The views expressed in this submission are those of Professor Rocky Scopelliti in an independent capacity and have no bearing on the views of organisations he is affiliated to.

TABLE OF CONTENTS

FOREWORD 4

Submission to the Discussion Paper on Safe and Responsible AI in Australia 4

EXECUTIVE SUMMARY: ADVANCING SAFE AND RESPONSIBLE AI IN AUSTRALIA 5

DEFINITIONS 6

1. Do you agree with the definitions in this Discussion Paper? If not, what definitions do you prefer and why? 6

POTENTIAL GAPS IN APPROACHES 7

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? 7

3. Are there any further non-regulatory initiatives the Australian Government could implement to support responsible AI practices in Australia? Please describe these and their benefits or impacts. 9

4. Do you have suggestions on coordination of AI governance across government? Please outline the goals that any coordination mechanisms could achieve and how they could influence the development and uptake of AI in Australia. 10

RESPONSES SUITABLE FOR AUSTRALIA 12

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? 12

TARGET AREAS 15

6. Should different approaches apply to public and private sector use of AI technologies? If so, how should the approaches differ? 15

7. How can the Australian Government further support responsible AI practices in its own agencies? 16

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. 17

9. Given the importance of transparency across the AI lifecycle, please share your thoughts on: 18

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? 18

b. mandating transparency requirements across the private and public sectors, including how these requirements could be implemented. 18

10. Do you have suggestions for: 19

a. Whether any high-risk AI applications or technologies should be banned completely? 19

b. Criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts? 19

11. What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI? 20

IMPLICATIONS AND INFRASTRUCTURE 22

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? 22

13. What changes (if any) to Australian conformity infrastructure might be required to support assurance processes to mitigate against potential AI risks? 23

RISK-BASED APPROACHES 25

14. Do you support a risk-based approach for addressing potential AI risks? If not, is there a better approach? 25

15. What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome? 26

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? 28

17. What elements should be in a risk-based approach for addressing potential AI risks? Do you support the elements presented in Attachment C? 29

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? 30

19. How might a risk-based approach apply to general purpose AI systems, such as large language models (LLMs) or multimodal foundation models (MFMs)? 31

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: 32

a. public or private organisations or both? 32

b. developers or deployers or both? 32

RECOMMENDATIONS 35

Creating Regulation and Governance for Safe and Responsible AI in Australia 35

ABOUT THE AUTHOR 44

REFERENCES 45

FOREWORD

SUBMISSION TO THE DISCUSSION PAPER ON SAFE AND RESPONSIBLE AI IN AUSTRALIA

This submission is my response to the Discussion Paper on Safe and Responsible AI in Australia, which seeks to shape the regulatory framework and governance practices surrounding AI in our country. As a concerned citizen/researcher/professional in the field of AI, I commend the Australian government's proactive approach in recognising the importance of addressing the risks associated with AI technologies and ensuring their safe and responsible deployment.

In this submission, my insights and recommendations are based on the comprehensive research I've conducted on emerging technologies for some twenty years now. That research covered many aspects of AI regulation and governance, including the importance of transparency, the need for a risk-based approach, sector-specific considerations, and the role of the government in promoting responsible AI practices. Drawing from this research, I will outline key points that should be considered to create an effective and robust regulatory framework for safe and responsible AI in Australia.

First and foremost, it is crucial to establish a clear set of principles and ethical guidelines that govern the development, deployment, and use of AI technologies. These guidelines should encompass transparency, fairness, accountability, and privacy, ensuring that AI systems are aligned with our societal values and human rights. Transparency should be prioritised, particularly in high-risk AI applications, to mitigate potential risks and foster public trust and confidence.

Additionally, a risk-based approach should be adopted to effectively address the diverse range of AI technologies and applications. This approach recognises that not all AI systems pose the same level of risks and allows for tailored regulations based on the potential harm they may cause. Such an approach will strike a balance between facilitating innovation and safeguarding against unintended consequences, ensuring that regulatory measures are proportional and appropriate for the specific context.

Furthermore, it is imperative to establish mechanisms for ongoing monitoring, evaluation, and accountability of AI systems. Regular audits and impact assessments should be conducted to identify and rectify any biases, errors, or unintended consequences that may arise from AI technologies. This will require collaboration between government agencies, industry stakeholders, and independent experts to ensure comprehensive oversight and meaningful accountability.

Moreover, the Australian government should play an active role in fostering collaboration and knowledge sharing among various stakeholders, including researchers, industry leaders, and civil society organisations. Establishing partnerships and platforms for dialogue will facilitate the exchange of best practices, promote responsible AI innovation, and enable continuous learning and improvement in AI governance.

Lastly, to effectively implement and enforce the regulatory framework, appropriate resources, expertise, and capacity-building initiatives should be provided to both public and private organisations. This will ensure that all stakeholders are equipped with the necessary tools and knowledge to comply with the regulations and adopt responsible AI practices.

I believe that by addressing the key considerations and recommendations outlined in this submission, the Australian government can lay the foundation for a robust and comprehensive regulatory framework that promotes safe and responsible AI in our country. I appreciate the opportunity to contribute to this important discussion and look forward to witnessing the positive impact of these efforts in shaping the future of AI in Australia.

Thank you for your attention to this matter. Should you require any further information or clarification, please do not hesitate to contact me.

Sincerely,

ROCKY SCOPELLITI
FUTUROLOGIST

EXECUTIVE SUMMARY: ADVANCING SAFE AND RESPONSIBLE AI IN AUSTRALIA

This executive summary provides a concise overview of the key insights and recommendations presented in this submission to the Australian government's Discussion Paper Safe and Responsible AI in Australia (Discussion Paper) in June 2023. This submission is a culmination of the critical review, analysis, and research I have conducted. It aims to contribute to the ongoing dialogue on how the government can create the necessary regulation and governance framework to ensure the safe and responsible development, deployment, and use of AI technologies in Australia.

In this submission and response to the twenty questions listed in the Discussion Paper, I emphasise the importance of adopting a comprehensive approach to address the potential risks associated with AI while promoting innovation and fostering public trust. I highlight several key themes and recommendations:

1. **GOVERNANCE AND REGULATION:** I recommend the establishment of a regulatory framework that combines both legal and ethical considerations. This framework should encompass clear guidelines and standards, while allowing for flexibility and adaptability to accommodate the rapidly evolving AI landscape. It should also involve multi-stakeholder collaboration and oversight to ensure transparency, accountability, and responsible AI practices.
2. **RISK-BASED APPROACH:** I support the adoption of a risk-based approach that assesses AI applications and technologies based on their potential risks and societal impacts. This approach should be context-specific, considering factors such as organisation size, AI maturity, and available resources. I believe that a risk-based approach can effectively mitigate potential harms while avoiding undue burden on low-risk applications.
3. **TRANSPARENCY AND ACCOUNTABILITY:** I emphasise the need for transparency across the AI lifecycle. Transparency initiatives, such as impact assessments, notices, human in the loop/oversight assessments, explanations, and ongoing monitoring, are crucial to building public trust and confidence. Mandating transparency requirements across the public and private sectors can ensure that individuals are informed about the use of AI systems and can seek reviews of decisions when needed.
4. **SECTOR-SPECIFIC CONSIDERATIONS:** I acknowledge that certain sectors may require tailored approaches due to their unique characteristics and risks. For instance, high-risk AI applications or technologies, such as social scoring or facial recognition, may warrant stricter regulations or even complete bans in specific contexts. It is essential to conduct thorough assessments and engage in sector-specific consultations to determine the appropriate regulatory measures.
5. **COLLABORATION AND EDUCATION:** I stress the importance of collaboration between government, industry, academia, and civil society in fostering responsible AI practices. Initiatives such as public-private partnerships, knowledge-sharing platforms, and AI education programs can enhance awareness, capacity building, and best practices dissemination. Moreover, international collaboration and harmonisation of AI governance efforts can facilitate global trust, interoperability, and ethical AI development.

This submission urges the Australian government to adopt a proactive and forward-thinking approach to regulate AI effectively. By incorporating the insights and recommendations presented in this submission, I believe that Australia can establish a robust regulatory and governance framework that promotes the safe and responsible development and deployment of AI technologies while positioning the country as a global leader in the ethical AI landscape.

Please refer to the full submission document for a detailed analysis, case studies, and supporting examples.

DEFINITIONS

1. DO YOU AGREE WITH THE DEFINITIONS IN THIS DISCUSSION PAPER? IF NOT, WHAT DEFINITIONS DO YOU PREFER AND WHY?

The definitions presented in the Discussion Paper provide a basic understanding of the terms used in the context of AI regulation. However, there are certain aspects that can be critically examined and refined for greater clarity and accuracy. In evaluating the definitions provided in the Discussion Paper, it is important to consider their clarity, comprehensiveness, and alignment with current industry practices. While the definitions draw from reputable sources like the International Organization for Standardization (ISO), there are areas where further refinement and specificity could enhance their usefulness.

One aspect that could benefit from additional clarity is the definition of "governance." While the Discussion Paper states that it includes both regulatory and voluntary mechanisms, it would be valuable to provide more concrete examples and case studies to illustrate how these mechanisms operate in practice. For instance, highlighting specific regulations or voluntary initiatives implemented in other countries and their impact on responsible AI practices would provide a clearer understanding of effective governance approaches.

Regarding the definition of AI as an "engineered system" that generates predictive outputs without explicit programming is overly broad and does not capture the nuances of AI technologies. AI systems can encompass a wide range of techniques and approaches beyond prediction, such as classification, clustering, and reinforcement learning. Furthermore, the definition does not explicitly address the concept of learning and adaptation, which are fundamental aspects of AI. For example, considering the specific challenges and risks associated with different AI techniques, such as deep learning or reinforcement learning, and their implications for responsible AI deployment would provide a more comprehensive definition.

The definition of machine learning as patterns derived from training data is accurate to a certain extent, but it fails to mention the iterative nature of the learning process. Machine learning involves the development of models that improve their performance over time by adjusting their parameters based on feedback from the data. This iterative learning process distinguishes machine learning from static patterns derived from training data.

Regarding the definition of generative AI models, it correctly highlights their ability to generate novel content in response to prompts. However, it does not mention the underlying mechanisms that enable this capability, such as neural networks and deep learning architectures. Providing more detail on the technical aspects of generative AI models would enhance the definition.

The definition of automated decision making (ADM) captures the broad scope of its application in various stages of the decision-making process. However, it lacks specificity in distinguishing between different types of automated systems and their levels of autonomy. Differentiating between rules-based systems and advanced technological systems, such as machine learning-based decision-making models, would provide a clearer understanding of the landscape. It would be valuable to provide more detailed examples and case studies that highlight the risks and challenges associated with different types of ADM systems. For instance, discussing instances where biases or discriminatory outcomes have arisen from automated decision-making processes, regardless of AI involvement, would contribute to a more nuanced understanding of governance needs.

While the definitions presented in the Discussion Paper provide a starting point for understanding the terms related to AI regulation, they can be further refined to improve accuracy and clarity. Through further refinement and context-specific elements it would enhance their effectiveness. By incorporating more precise language and capturing essential technical aspects, the definitions can better reflect the complex nature of AI technologies and their implications for regulation[i].

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 to AI demonstrate a recognition of potential risks and the need for governance mechanisms. The inclusion of general regulations that apply across industries, such as data protection and privacy law, competition law, and consumer protection law, provides a foundation for addressing AI-related harms. Additionally, sector-specific regulations in areas like therapeutic goods, food, and financial services indicate an understanding of the need for tailored regulations in specific domains.

However, despite these efforts, there are still potential risks from AI that are not adequately covered by Australia's existing regulatory approaches. One such risk is the potential for biased or discriminatory outcomes in automated decision-making systems (ADM). While discrimination law is listed as a general regulation, more specific measures may be required to address the unique challenges posed by AI algorithms and their potential for perpetuating bias. Case studies and research have highlighted instances where AI systems have exhibited biased behaviour, such as the discriminatory impact of AI-based hiring tools or biased algorithmic decision-making in the criminal justice system.

Another area where existing regulations may fall short is the transparency and accountability of AI systems. The Privacy Act Review proposals rightly address the need for transparency in ADM. However, there is a broader need for regulations that ensure explainability and auditability of AI systems, enabling individuals and stakeholders to understand the decision-making processes and assess the potential biases or risks involved. This is particularly crucial for critical sectors such as healthcare, finance, and autonomous vehicles, where the impact of AI decisions can have significant consequences on individuals' lives and well-being.

To mitigate these risks and enhance existing regulatory approaches, there are several possible suggestions for regulatory action in Australia. Firstly, specific guidelines or standards can be developed to address algorithmic fairness and non-discrimination, ensuring that AI systems do not perpetuate biases or discriminate against protected groups. This could involve measures such as regular audits of AI systems, transparency in data collection and usage, and independent oversight to ensure compliance.

Secondly, there is a need for regulatory frameworks that mandate explainability and transparency in AI systems. This can be achieved through requirements for clear documentation of AI models, disclosure of training data sources, and the provision of meaningful explanations for AI-generated decisions. Additionally, mechanisms for auditing and validating AI systems can be established to assess their compliance with fairness, transparency, and accountability standards.

Furthermore, collaboration with international bodies and organisations can provide valuable insights and best practices for AI regulation. Learning from the experiences of countries that have already implemented comprehensive AI regulatory frameworks, such as the European Union's General Data Protection Regulation (GDPR) or Canada's Directive on Automated Decision-Making, can help inform Australia's approach to AI governance[ii].

The potential risks associated with AI's impact on consumer rights and product safety are also worth considering. The application of Australian Consumer Law (ACL) to AI is a notable example. While the ACL outlines basic consumer guarantees and remedies for goods and services, it may not explicitly address the specific challenges posed by AI systems. The question of how the ACL applies to consumer-facing uses of AI, particularly generative AI, has not been fully explored by the courts. This raises concerns about consumer protection and the need for more explicit regulations to ensure AI-powered products and services meet the required safety standards and consumer guarantees.

An example where certain risks that may not be adequately covered is the application of the ACL to algorithmic decision-making, as demonstrated in the Trivago vs ACCC case.

The case highlighted how the ACL, originally drafted without AI in mind, was applied to address misleading hotel room recommendations generated by an algorithm. This example illustrates the challenges of applying existing laws to AI systems that may produce unintended or deceptive outcomes.

To mitigate these risks and address potential regulatory gaps, several suggestions for regulatory action can be considered. Firstly, there is a need to develop guidelines or standards that specifically address the application of anti-discrimination laws to AI. This could involve clarifying the responsibilities of AI developers and users in ensuring fairness and non-discrimination in AI systems and establishing mechanisms for auditing and monitoring compliance.

In terms of consumer protection, there is a need for a comprehensive assessment of how the ACL can be applied to AI, including generative AI. This assessment should consider the unique risks associated with AI-powered products and services, such as the potential for bias, inadequate transparency, or unintended consequences. If necessary, specific provisions or guidelines can be introduced to address these risks and ensure that AI-powered goods and services meet the required safety standards and consumer guarantees.

Additionally, regulatory frameworks should emphasise the importance of explainability and transparency in AI systems. This can be achieved through requirements for clear documentation of AI models, disclosure of data sources, and provision of meaningful explanations for AI-generated decisions. Auditing and certification processes can also be established to assess the transparency and fairness of AI systems.

To inform the development of effective regulations, it is crucial to engage with experts and AI practitioners who possess in-depth knowledge of AI technologies and their potential risks. Collaboration with international bodies and organisations can also provide valuable insights and best practices in AI regulation[iii].

One of the limitations of Australia's current regulatory landscape is that remedies are often provided after potential harms have occurred, making it challenging to address systemic or irreversible impacts caused by AI. While existing laws can serve as effective deterrents, there is a need for preventive measures to limit problems before they arise. The Online Safety Act 2021, for instance, introduces Basic Online Safety Expectations and industry codes to enhance transparency and address illegal or harmful content online. These initiatives can be adapted to prevent potential harms arising from AI, such as proactive detection and demotion of harmful content in algorithm recommendations.

Another aspect to consider is the complexity and potential duplication of regulations when AI is combined with other emerging technologies. As AI is often used in conjunction with other components to create innovative products and services, the regulatory landscape becomes more intricate. This complexity can lead to duplication or conflicts between regulatory systems, imposing additional compliance burdens on AI developers and adopters. Context-specific responses may be necessary to address the diverse applications of AI in different sectors while avoiding a one-size-fits-all approach. For example, regulations suitable for medical device regulation may not be applicable or appropriate in the education sector.

To mitigate these challenges and address potential regulatory gaps, it is crucial to identify context-specific governance mechanisms and develop additional AI governance initiatives that support the safe and responsible development and adoption of AI. This requires a collaborative approach involving relevant government portfolios, experts, and stakeholders. The ongoing exploration and consideration of AI developments within specific governance areas, such as education and online safety, demonstrate the recognition of the need for tailored responses[iv].

Australia's existing regulatory approaches acknowledge the risks associated with AI, but additional action is needed to effectively address these challenges. Ensuring fairness, transparency, and accountability through tailored regulations and guidelines is crucial for safeguarding individuals and society in the rapidly evolving AI landscape. By implementing preventive measures, context-specific responses, and collaborative governance initiatives, Australia can promote the safe and responsible use of AI while mitigating potential risks.

3. ARE THERE ANY FURTHER NON-REGULATORY INITIATIVES THE AUSTRALIAN GOVERNMENT COULD IMPLEMENT TO SUPPORT RESPONSIBLE AI PRACTICES IN AUSTRALIA? PLEASE DESCRIBE THESE AND THEIR BENEFITS OR IMPACTS.

In addition to regulatory initiatives, the Australian Government can implement non-regulatory measures to support responsible AI practices in the country. These initiatives can complement existing regulations and foster a culture of ethical and responsible AI development and adoption.

SOME POTENTIAL NON-REGULATORY INITIATIVES INCLUDE:

- *Guidelines and Best Practices:* The government can develop guidelines and best practices for AI development and deployment. These guidelines can outline ethical principles, transparency requirements, and responsible data handling practices. They can serve as a reference for organisations and developers to ensure that their AI systems align with ethical standards and promote trust and accountability.
- *Standards and Certification:* The government can collaborate with industry stakeholders and experts to establish standards and certification frameworks for AI systems. These standards can cover various aspects, such as fairness, accountability, transparency, and robustness. Certification programs can provide assurance to consumers and businesses that AI systems meet specific criteria and adhere to responsible AI practices.
- *Education and Training:* Investing in education and training programs focused on AI ethics and responsible practices can help build capacity and awareness among developers, data scientists, and decision-makers. Workshops, seminars, and online courses can provide guidance on ethical considerations, bias mitigation, privacy protection, and other aspects of responsible AI development and deployment.
- *Collaboration and Knowledge Sharing:* The government can facilitate collaboration between industry, academia, and research institutions to encourage knowledge sharing and exchange of best practices. Creating platforms and networks where experts and stakeholders can come together to discuss AI ethics, share insights, and address emerging challenges can foster a collective approach to responsible AI practices.
- *Public Awareness and Engagement:* Launching public awareness campaigns and initiatives can help educate the public about AI technologies, their potential benefits, and associated risks. By promoting a better understanding of AI and its implications, the government can encourage public engagement, informed decision-making, and dialogue on responsible AI development and usage.

SOME POTENTIAL BENEFITS AND IMPACTS INCLUDE:

- *Enhanced Trust and Consumer Confidence:* Non-regulatory initiatives can contribute to building trust in AI systems by promoting responsible practices. Clear guidelines, standards, and certification programs can provide assurance to consumers and businesses that AI technologies are developed and used in an ethical and accountable manner.
- *Improved Transparency and Accountability:* Guidelines and best practices can promote transparency in AI decision-making processes, ensuring that stakeholders have a clear understanding of how AI systems function and make decisions. This transparency enhances accountability and enables the identification and mitigation of biases or discriminatory outcomes.
- *Stimulated Innovation:* Non-regulatory initiatives that encourage collaboration, knowledge sharing, and capacity building can foster innovation in AI technologies. By providing guidance and resources to developers, the government can facilitate the creation of responsible and socially beneficial AI solutions.
- *International Collaboration and Reputation:* Implementing non-regulatory initiatives aligned with international standards and practices can position Australia as a leader in responsible AI.

Collaboration with other countries and participation in global discussions and forums can help shape international norms and frameworks for AI governance.

- *Sustainable and Ethical AI Ecosystem:* Non-regulatory measures can contribute to the long-term sustainability and ethical development of the AI ecosystem in Australia. By promoting responsible AI practices, the government can mitigate potential risks, ensure compliance with ethical standards, and foster public confidence in AI technologies.

It is important to note that non-regulatory initiatives should work in tandem with regulatory approaches to create a comprehensive framework for responsible AI practices in Australia.

4. DO YOU HAVE SUGGESTIONS ON COORDINATION OF AI GOVERNANCE ACROSS GOVERNMENT? PLEASE OUTLINE THE GOALS THAT ANY COORDINATION MECHANISMS COULD ACHIEVE AND HOW THEY COULD INFLUENCE THE DEVELOPMENT AND UPTAKE OF AI IN AUSTRALIA.

COORDINATION OF AI GOVERNANCE ACROSS GOVERNMENT IN AUSTRALIA:

AI governance is crucial for ensuring the responsible development and uptake of artificial intelligence technologies. In Australia, governance responses to date have primarily been voluntary with the release of the AI Ethics Framework in 2019 as a notable step which I was delighted to have contributed to. The following explores the coordination mechanisms for AI governance across the Australian government, outlines the goals they aim to achieve, and assesses their potential influence on the development and uptake of AI.

GOALS OF COORDINATION MECHANISMS:

- *Establish a Centralised AI Governance Body:* Create a dedicated government agency or organisation responsible for coordinating and overseeing AI governance efforts. This centralised body can provide expertise, guidance, and enforcement of AI-related regulations, ensuring consistency and effectiveness across different sectors and government agencies.

- *Foster Interagency Collaboration:* Facilitate collaboration and information-sharing among various government departments and agencies involved in AI governance. This collaboration can help align efforts, share resources, and promote a unified approach to addressing AI-related challenges.

- *Building Trust and Confidence:* The voluntary AI Ethics Framework, consisting of eight principles, aims to guide businesses and government in responsibly designing and implementing AI systems. By promoting the principles of human well-being, fairness, privacy protection, transparency, and accountability, these mechanisms seek to build trust and confidence in the use of AI.

- *Ensuring Ethical and Inclusive AI:* The coordination mechanisms emphasise human-centered values, respect for human rights and diversity, and the prevention of unfair discrimination. They aim to ensure that AI systems are inclusive, accessible, and do not harm individuals, communities, or groups.

- *Privacy Protection and Data Security:* With a focus on respecting privacy rights and data protection, the mechanisms aim to ensure that AI systems uphold privacy standards and safeguard data from unauthorised access or misuse.

- *Reliability and Safety:* The coordination mechanisms seek to ensure that AI systems operate reliably and safely according to their intended purpose. They promote responsible disclosure and transparency to help individuals understand when they are significantly impacted by AI systems.

- *Contestability and Accountability:* The mechanisms emphasise the importance of providing individuals with the ability to challenge the use or outcomes of AI systems when they significantly impact them. They also stress the need for clear identification and accountability of individuals responsible for different phases of the AI system lifecycle.

- *Enhance Stakeholder Engagement:* Engage industry, academia, civil society organisations, and the public in the development and implementation of AI governance frameworks. Soliciting diverse perspectives and expertise can improve the effectiveness and legitimacy of AI regulations and ensure that they align with the needs and expectations of different stakeholders.
- *Regular Review and Updates:* Establish a mechanism for regular review and updates of AI governance frameworks to keep pace with technological advancements and emerging risks. This process should involve ongoing consultation with experts and stakeholders to ensure that regulations remain relevant, effective, and adaptable to changing circumstances.
- *International Collaboration:* Foster collaboration and information-sharing with international partners on AI governance best practices, standards, and regulatory approaches. This collaboration can help Australia stay informed of global developments and ensure that its governance frameworks align with international norms and standards.

INFLUENCE ON AI DEVELOPMENT AND UPTAKE:

The coordination mechanisms for AI governance in Australia outlined above can have several significant influences on the development and uptake of AI in the country:

- *Consistency and Effectiveness:* The establishment of a centralised AI governance body can ensure consistent and effective implementation of AI-related regulations across different sectors and government agencies. This centralised coordination can provide clear guidelines, expertise, and enforcement mechanisms, promoting responsible AI development and uptake.
- *Collaboration and Resource Sharing:* Facilitating interagency collaboration enables the sharing of resources, knowledge, and best practices among different government departments involved in AI governance. This collaboration fosters a unified approach to addressing AI-related challenges, streamlines processes, and avoids duplication of efforts.
- *Building Trust and Confidence:* The voluntary AI Ethics Framework and the principles it promotes help build trust and confidence in the use of AI technologies. By emphasising human well-being, fairness, transparency, and accountability, these mechanisms assure businesses and the public that AI systems are designed and implemented responsibly, thereby encouraging their adoption and acceptance.
- *Ethical and Inclusive AI:* The coordination mechanisms prioritise human-centered values, diversity, and the prevention of unfair discrimination. By ensuring that AI systems are inclusive, accessible, and respectful of human rights, these mechanisms promote the development and uptake of AI technologies that align with ethical standards and societal expectations.
- *Privacy Protection and Data Security:* The mechanisms' focus on privacy rights and data protection ensures that AI systems uphold privacy standards and safeguard sensitive information. By addressing concerns related to privacy and data security, these mechanisms can enhance public trust and encourage the responsible use of AI technologies.
- *Reliability and Safety:* The coordination mechanisms emphasise the reliable and safe operation of AI systems. By promoting responsible disclosure, transparency, and accountability, these mechanisms help individuals understand the impact of AI systems on their lives, fostering trust and ensuring the responsible development and uptake of AI technologies.
- *Stakeholder Engagement*: Engaging industry, academia, civil society organisations, and the public in the development and implementation of AI governance frameworks ensures diverse perspectives and expertise are considered. This inclusive approach enhances the legitimacy and effectiveness of regulations, and it aligns AI development and uptake with the needs and expectations of various stakeholders.

Overall, the coordination mechanisms outlined above contribute to fostering responsible AI development, building public trust, and ensuring that AI technologies in Australia are developed and adopted in an ethical, inclusive, and secure manner[v].

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?

As the development and adoption of AI continues to accelerate globally, many countries are taking measures to govern AI technology safely and responsibly. The following explores governance measures implemented or considered by countries worldwide, assessing their relevance, adaptability, and desirability for Australia. It examines specific examples, case studies, and relevant insights into potential approaches Australia can consider.

INTERNATIONAL GOVERNANCE MEASURES:

- *Canada:* Canada's Directive on Automated Decision-Making focuses on transparency, accountability, and privacy in AI systems. It requires federal agencies to document the use of AI and conduct algorithmic impact assessments. Australia can adapt these requirements to enhance transparency and accountability in government AI systems. *Case Study: Algorithmic Impact Assessment Framework - Canadian Human Rights Commission.*

- *European Union (EU):* The EU has introduced the General Data Protection Regulation (GDPR), which establishes strict regulations for data protection and privacy. The GDPR has implications for AI systems that process personal data. Australia can draw lessons from the GDPR in strengthening its own privacy and data protection laws to address AI-related challenges. The EU is working on passing its draft Artificial Intelligence Act, a comprehensive piece of legislation intended to govern nearly all uses of AI. The Act groups AI applications into four risk categories, each governed by a predefined set of regulatory tools. *Case Study: EU GDPR and Its Impact on AI Governance.*

- *Singapore:* Singapore has implemented the Model AI Governance Framework, which provides comprehensive guidance on ethical and responsible AI development. The framework emphasises fairness, transparency, accountability, and human-centricity. Australia can consider adopting similar frameworks to ensure responsible AI practices across sectors. *Case Study: Model AI Governance Framework - Infocomm Media Development Authority, Singapore.*

- *United States:* The United States has established the National Artificial Intelligence Research and Development Strategic Plan, which outlines strategic objectives for AI research, development, and governance. The plan promotes public-private partnerships, interdisciplinary collaboration, and investments in AI. Australia can explore similar strategic plans to foster AI innovation and governance. *Case Study: National Artificial Intelligence Research and Development Strategic Plan - National Science and Technology Council, USA.*

- *United Kingdom:* The UK's Centre for Data Ethics and Innovation (CDEI) conducts research and provides recommendations on AI governance. The CDEI explores issues such as algorithmic bias, online targeting, and AI in the criminal justice system. Australia can establish a similar independent body to provide expertise and guidance on AI governance. *Case Study: Centre for Data Ethics and Innovation - UK Government.*

- *China:* Over the past year, China has rolled out some of the world's first nationally binding regulations targeting algorithms and AI. It has taken a fundamentally vertical approach: picking specific algorithm applications and writing regulations that address their deployment in certain contexts. For example, China's internet regulator has created a mandatory registration system for recommendation algorithms. This system is part of China's 2022 regulation on recommendation algorithms, which came into effect in March of this year and was led by the Cyberspace Administration of China (CAC). *Case Studies: The Cyberspace Administration of China (CAC) has released a draft set of thirty rules for regulating internet recommendation algorithms, the software powering everything from TikTok to news apps and search engines;*

The China Academy of Information and Communications Technology has developed tools for testing and certification of "trustworthy AI" systems; The Ministry of Science and Technology has established AI ethics principles and is creating tech ethics review boards within companies and research institutions.

KEY INSIGHTS FOR AUSTRALIA'S AI REGULATION:

- *Blend of Horizontal and Vertical Approaches:* Both the EU and China have adopted a combination of horizontal and vertical approaches to regulate AI. A purely horizontal approach is insufficient to address specific requirements, while a collection of freestanding vertical regulations can result in compliance challenges. A blended approach, incorporating both horizontal and vertical elements, is more effective.
- *EU's Horizontal Approach:* The EU's AI Act takes a broadly horizontal approach by grouping AI applications into risk categories and defining regulatory tools for each category. It provides flexibility for domain-specific bodies to determine compliance strategies and parameters. The EU's focus on essential requirements and technical standards allows for precise regulation across sectors and evolving technologies.
- *Risks of the EU Approach:* The EU's horizontal approach faces risks such as inconsistent interpretations and enforcement by individual regulators, as well as potential industry dominance in the standards-setting process. The effectiveness of the proposed European AI Office in supplementing national and sectoral regulators is also a factor to watch.
- *China's Vertical Approach:* China has implemented vertical regulations targeting specific AI applications, such as recommendation algorithms and deep synthesis technology. These regulations address specific contexts and requirements. China's algorithm registry serves as a horizontal tool to gather information on algorithms across various vertical regulations.
- *Precision and Compliance Challenges in China's Approach:* China's vertical regulations allow for targeted requirements but risk falling behind rapidly evolving technology. Vague requirements may shift power from technology companies to regulators but can also create a patchwork of regulations that are costly to comply with and poorly considered collectively.
- *Considerations from the United States:* The United States has adopted a blend of horizontal and vertical elements in its AI regulation efforts. Political guidance and principles have been provided, but coordination and resources are necessary to ensure effective sector-specific adaptation. Federal rulemaking risks reversal with changing administrations, but funding AI research and developing best-practice guidelines can support future vertical regulations.
- *Lessons for AI Regulation:* The effective regulation of AI requires a combination of horizontal and vertical elements. Horizontal approaches provide predictability and address common themes of transparency, robustness, and accountability. Vertical regulations allow tailored mitigation of specific harms and can inform future horizontal regimes. Legislative and interagency coordination is crucial for vertical regulations to minimise costs and ensure effective enforcement.

These insights highlight the need for a flexible and adaptive approach to AI regulation, combining horizontal and vertical elements while considering legislative structures, coordination, and resource allocation. Adapting lessons from the US, EU and China can help Australia develop effective and responsible AI governance strategies.

ADAPTABILITY AND RELEVANCE FOR AUSTRALIA:

The blend of horizontal and vertical approaches, as observed in the US, EU and China, can serve as valuable models. The EU's horizontal approach, as demonstrated by the AI Act, categorises AI applications into risk categories and defines regulatory tools for each category. This approach provides flexibility for domain-specific bodies to determine compliance strategies and parameters, allowing for precise regulation across sectors and evolving technologies. Australia could consider adopting a similar approach to address specific requirements and ensure effective regulation.

However, it is important to be mindful of the risks associated with the EU's horizontal approach, such as inconsistent interpretations and enforcement by individual regulators, as well as potential industry dominance in standards-setting. Australia should take these risks into account and develop mechanisms to mitigate them, potentially through effective coordination among different regulatory bodies.

China's vertical approach, on the other hand, focuses on implementing targeted regulations for specific AI applications, addressing their unique contexts and requirements. Australia could consider adopting vertical regulations to address specific AI use cases and mitigate associated harms. However, it is crucial to balance precision with compliance challenges, ensuring that the regulations keep pace with rapidly evolving technology and do not create a burdensome patchwork of requirements.

Lessons from the United States, which has adopted a blend of horizontal and vertical elements in its AI regulation efforts, can also inform Australia's approach. Providing political guidance and principles while ensuring effective coordination and resource allocation is necessary to support sector-specific adaptation. Funding AI research and developing best-practice guidelines can also support the development of future vertical regulations.

For Australia, a flexible and adaptive approach to AI regulation, combining both horizontal and vertical elements, is recommended. By considering lessons from the EU, China, and the United States, Australia can develop effective and responsible AI governance strategies that address specific requirements while remaining adaptable to technological advancements.

These international governance measures offer valuable insights and examples for Australia to consider. They address various aspects of AI governance, including transparency, accountability, privacy, ethics, and public engagement. Australia can adapt and customise these measures to align with its unique socio-cultural context and regulatory landscape.

DESIRABILITY AND CHALLENGES:

As recommended previously, the Australian government should consider a blended approach to AI regulation, combining horizontal and vertical elements, as observed in the US, EU and China. This approach offers desirability by allowing for addressing specific requirements while maintaining flexibility and adaptability to evolving technologies. By striking the right balance between horizontal and vertical elements, Australia can ensure consistent interpretation and enforcement, while avoiding compliance challenges through coordination and harmonisation among regulatory bodies.

The EU's horizontal approach, demonstrated by the AI Act, provides desirability for Australia. As mentioned previously, it categorises AI applications into risk categories and defines regulatory tools, enabling domain-specific bodies to determine compliance strategies. However, challenges such as inconsistent interpretations and enforcement, as well as potential industry dominance in standards-setting, need to be addressed through effective coordination and oversight.

Adopting China's vertical approach can also benefit Australia. As mentioned previously, this approach targets specific AI applications, addressing unique contexts and requirements, and promotes responsible AI practices. However, challenges arise in maintaining precision and keeping up with rapidly evolving technology. Regular updates to vertical regulations are necessary to remain relevant and adaptable and to avoid compliance challenges and a fragmented regulatory landscape.

Drawing lessons from the United States' blend of horizontal and vertical elements, Australia can provide political guidance, coordination, and resources for effective sector-specific adaptation. However, the risk of federal rulemaking reversals with changing administrations must be addressed. Establishing stability and continuity in AI regulation through mechanisms such as best-practice guidelines and long-term commitments to AI research is crucial.

While these blended governance measures offer desirable approaches, their implementation may face challenges. Balancing innovation with regulation, ensuring cross-sector collaboration, and addressing the evolving nature of AI pose ongoing challenges for any governance framework. Moreover, the international examples cited may not directly address Australia's specific challenges and priorities, necessitating careful adaptation[vi].

TARGET AREAS

6. SHOULD DIFFERENT APPROACHES APPLY TO PUBLIC AND PRIVATE SECTOR USE OF AI TECHNOLOGIES? IF SO, HOW SHOULD THE APPROACHES DIFFER?

The use of AI technologies in both the public and private sectors has raised important governance considerations. The following examines whether different approaches should apply to the public and private sector use of AI technologies and explores how these approaches might differ.

DIFFERENTIATING APPROACHES:

Public Sector: The public sector operates under distinct responsibilities, such as safeguarding public interest, accountability, and transparency. As a result, AI governance in the public sector must prioritise principles such as fairness, equity, and citizen engagement. Key considerations for public sector AI governance include:

- *Accountability and Transparency:* The public sector should prioritise transparency in AI decision-making processes, ensuring that algorithms are explainable and accountable to the public. Estonia's "AI Black Box" project is an example of a government initiative aiming to increase transparency in public sector AI.
- *Ethical and Social Implications:* Public sector AI governance should incorporate ethics and address potential societal implications. For instance, the use of facial recognition technology by law enforcement agencies requires careful considerations to prevent bias and privacy violations, as demonstrated by controversies surrounding facial recognition systems used by police departments in various countries.
- *Public Participation:* Governments should actively involve citizens in the development and deployment of AI systems, promoting public trust and ensuring inclusive decision-making processes. The citizen engagement initiatives of South Korea's National AI Strategy exemplify efforts to involve the public in AI governance.

Private Sector: The private sector operates within a different context, driven by market dynamics, innovation, and profit motives. While accountability and ethical considerations remain important, the private sector's governance approach may differ due to the following factors:

- *Intellectual Property and Trade Secrets:* Private sector AI governance must balance the need for transparency with the protection of intellectual property and trade secrets. Companies may need to safeguard proprietary algorithms or datasets while still ensuring accountability and transparency in their AI systems.
- *Industry Self-Regulation:* The private sector can play a role in developing industry-specific standards and best practices. Collaborative efforts such as the Partnership on AI, which includes major technology companies, demonstrate self-regulation initiatives aimed at promoting responsible AI development.
- *Risk Management and Innovation:* The private sector emphasises risk management and innovation. Governance approaches should enable responsible innovation while managing risks associated with biased algorithms, discriminatory outcomes, or negative societal impacts. Case studies, such as the adoption of responsible AI practices by major tech companies like Google or Microsoft, highlight efforts to integrate ethical considerations into private sector AI.

Differentiating approaches to AI governance in the public and private sectors acknowledges their distinct contexts, responsibilities, and priorities. While the public sector should prioritise accountability, transparency, and citizen engagement, the private sector should balance market dynamics, intellectual property protection, and responsible innovation. However, it is essential to maintain cross-sector collaboration, information sharing, and regulatory oversight to ensure overall alignment with societal values and goals[vii].

7. HOW CAN THE AUSTRALIAN GOVERNMENT FURTHER SUPPORT RESPONSIBLE AI PRACTICES IN ITS OWN AGENCIES?

As AI technologies continue to advance, it is crucial for the Australian Government to foster responsible AI practices within its own agencies. The following explores how this can be achieved.

STRATEGIES FOR SUPPORTING RESPONSIBLE AI PRACTICES:

- *Establish Clear Guidelines and Standards:* The Australian Government should develop clear guidelines and standards for AI implementation in government agencies. These guidelines should encompass principles of transparency, accountability, fairness, and ethical considerations. The framework should draw from international best practices and engage with experts, academia, and industry stakeholders to ensure comprehensiveness and relevance. For instance, the European Commission's Ethics Guidelines for Trustworthy AI and the OECD's Principles on AI provide valuable references for developing AI guidelines.

- *Enhance AI Education and Training:* To promote responsible AI practices, the Australian Government should invest in education and training programs focused on AI ethics, bias detection, and responsible AI development. These programs should target government employees involved in AI decision-making processes and implementation. Collaborations with academic institutions, industry experts, and AI research centers can facilitate the development of tailored training programs. The Responsible AI Certification Program launched by the State Services Commission of New Zealand can serve as a valuable case study in this regard.

- *Encourage Cross-Agency Collaboration:* The Australian Government should foster collaboration and knowledge sharing among government agencies to develop a holistic approach to responsible AI practices. Establishing cross-agency working groups or task forces can facilitate the exchange of best practices, lessons learned, and challenges faced in AI implementation. The establishment of the United States Federal AI Community of Practice is a notable example of interagency collaboration to promote responsible AI adoption.

- *Establish Ethical Review Mechanisms:* The Australian Government should introduce ethical review mechanisms to assess the potential risks, impacts, and ethical implications of AI projects within government agencies. These review mechanisms can involve multidisciplinary committees or ethics boards that provide independent assessments and recommendations. The establishment of the Data Ethics Advisory Group in the United Kingdom's Cabinet Office can serve as a case study for implementing such mechanisms.

- *Foster Public-Private Partnerships:* Collaborating with the private sector can accelerate responsible AI practices in government agencies. The Australian Government should establish partnerships with AI industry leaders and organisations that promote ethical AI development. These partnerships can facilitate knowledge transfer, access to expertise, and collaborative initiatives for responsible AI implementation. The Australian Government's collaboration with the Data61 unit of CSIRO, demonstrates the potential benefits of public-private partnerships in advancing responsible AI practices.

To support responsible AI practices in Australian Government agencies, it is crucial to establish clear guidelines, enhance education and training programs, encourage cross-agency collaboration, introduce ethical review mechanisms, and foster public-private partnerships. By adopting these strategies, the Australian Government can effectively navigate the complexities of AI implementation while ensuring transparency, accountability, and ethical considerations in its agencies' use of AI technologies[viii].

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.

Addressing the risks associated with AI requires a careful balance between generic solutions that apply broadly to AI technologies and technology-specific solutions that address risks unique to specific AI applications. The following explores the circumstances in which generic solutions are most valuable and situations where technology-specific solutions offer better risk mitigation.

GENERIC SOLUTIONS TO AI RISKS:

- *Ethical Guidelines and Frameworks:* Generic solutions, such as ethical guidelines and frameworks, are valuable for addressing broad ethical concerns across AI applications. These guidelines provide high-level principles and standards that guide AI development and deployment. Notable examples include the European Commission's Ethics Guidelines for Trustworthy AI and the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. These frameworks help establish a foundation of responsible AI practices and promote transparency, fairness, and accountability.

- *Bias and Fairness Mitigation:* Mitigating bias and ensuring fairness are critical concerns across various AI applications. Generic solutions, such as algorithmic fairness techniques and bias detection methods, can be employed to identify and address bias in AI systems. Techniques like adversarial debiasing and pre-processing can help minimise discrimination in decision-making processes. These approaches provide valuable tools for addressing bias and fairness concerns across different AI domains, including recruitment, criminal justice, and credit scoring.

- *Explainability and Interpretability:* Promoting AI systems' explainability and interpretability is another generic solution applicable to a wide range of AI technologies. Ensuring that AI algorithms can provide explanations for their decisions is crucial for building trust, accountability, and regulatory compliance. Techniques like model interpretability algorithms, rule-based explanations, and counterfactual explanations enable users to understand the reasoning behind AI-generated outcomes. These generic approaches help address concerns regarding AI opacity and enable human oversight.

TECHNOLOGY-SPECIFIC SOLUTIONS TO AI RISKS:

- *Autonomous Vehicles and Safety:* In the domain of autonomous vehicles, technology-specific solutions are essential for addressing risks related to safety. Robust engineering practices, sensor redundancy, and fail-safe mechanisms specific to autonomous vehicles are required to mitigate the risk of accidents and ensure passenger safety. For instance, companies like Waymo employ lidar-based perception systems, sensor fusion algorithms, and extensive testing protocols to minimise the risk of accidents in their self-driving cars.

- *Healthcare and Privacy:* AI applications in healthcare present specific risks related to patient privacy and data security. Technology-specific solutions, such as advanced encryption methods, secure data storage, and access control mechanisms, are crucial for protecting sensitive patient information. Healthcare organisations and technology providers invest in specific privacy-preserving technologies like federated learning and differential privacy to ensure that patient data remains confidential and secure.

- *Cybersecurity and AI Threats:* AI technologies themselves can be used as tools for cyberattacks and malicious activities. Technology-specific solutions are necessary to address the unique risks associated with AI-enabled cyber threats. Developing AI-specific cybersecurity tools, such as anomaly detection algorithms and adversarial machine learning defenses, can enhance the resilience of AI systems against attacks. Organisations like IBM Research have been actively working on technology-specific solutions to detect and defend against adversarial attacks on AI models.

In addressing the risks of AI, a balanced approach that combines generic and technology-specific solutions is essential. Generic solutions, such as ethical guidelines, bias mitigation techniques, and explainability methods, provide a foundation for responsible AI practices applicable across diverse domains. Technology-specific solutions, on the other hand, are critical for addressing risks unique to specific AI applications, such as autonomous vehicles, healthcare, and cybersecurity. By understanding the circumstances where generic solutions are most valuable and the areas where technology-specific solutions offer better risk mitigation, stakeholders can effectively manage AI risks and promote responsible AI adoption[ix].

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.

Transparency plays a crucial role in mitigating potential risks associated with AI and in building public trust and confidence in AI systems. The following examines the areas where transparency is most critical and valuable and explores the implications of mandating transparency requirements across the private and public sectors.

A. CRITICAL AND VALUABLE ASPECTS OF TRANSPARENCY:

- *Algorithmic Decision-Making:* Transparency is particularly critical when it comes to algorithmic decision-making processes that have a significant impact on individuals or society. Transparent AI systems allow stakeholders to understand the inputs, processes, and outputs of algorithms, enabling them to assess the fairness, bias, and potential risks associated with the decisions made by AI. This is especially relevant in domains such as criminal justice, lending, and hiring, where AI algorithms have the potential to perpetuate or exacerbate existing biases and discrimination.

- *Data Collection and Use:* Transparency in data collection and use is essential to address concerns related to privacy, consent, and data governance. Individuals and communities should be informed about the types of data collected, how they are used, and the purposes for which they are employed in AI systems. Transparent data practices can help prevent unauthorised use, mitigate the risk of data breaches, and empower individuals to exercise control over their personal information. For example, the General Data Protection Regulation (GDPR) in the European Union mandates transparency requirements for organisations handling personal data.

- *Model Development and Deployment:* Transparency in the model development and deployment process is crucial to understand the limitations, biases, and potential risks associated with AI systems. Providing documentation and insights into the training data, model architecture, and evaluation metrics allows external stakeholders, including researchers and regulators, to assess the reliability, robustness, and potential biases of AI models. OpenAI's GPT language model documentation is an example of providing transparency by detailing its capabilities and limitations.

B. MANDATING TRANSPARENCY REQUIREMENTS:

Transparency is of paramount importance in mitigating potential risks and fostering public trust in AI systems. It is critical in algorithmic decision-making, data collection and use, as well as model development and deployment. Mandating transparency requirements across the private and public sectors can be achieved through regulations, industry standards, and auditing mechanisms. By implementing transparency measures, governments and organisations can enhance accountability, reduce biases, and address public concerns, thereby promoting responsible and trustworthy AI practices[x].

- *Private Sector:* Mandating transparency requirements in the private sector can be achieved through regulations and industry standards. Governments can introduce laws that enforce disclosure of AI system functionalities, data sources, and decision-making processes. Additionally, industry bodies and organisations can collaborate to develop best practices and transparency guidelines specific to different AI domains. For instance, the Partnership on AI, a multi-stakeholder initiative, focuses on developing and promoting responsible AI practices.

- *Public Sector:* In the public sector, transparency requirements can be implemented through legislation and regulatory frameworks. Governments can mandate that AI systems deployed in public services, such as healthcare, education, and law enforcement, undergo transparency assessments and disclose relevant information. The Algorithmic Accountability Act proposed in the United States exemplifies a regulatory approach to ensure transparency and accountability in public sector AI.

- *Auditing and Certification:* To ensure compliance with transparency requirements, independent auditing and certification mechanisms can be established. External auditors can assess AI systems, reviewing the data sources, algorithms, and decision-making processes, and provide certification or compliance reports. This helps build public trust and confidence in AI technologies by providing assurance that systems are designed and implemented in a transparent and responsible manner. The Trusted AI framework by the Singapore government incorporates auditing and certification as part of its approach to responsible AI adoption.

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 TECHNOLOGIES THAT SHOULD BE BANNED, AND IN WHICH CONTEXTS?

The rapid advancement of AI has raised concerns about the potential risks associated with certain applications and technologies. The following explores the question of whether any high-risk AI applications or technologies should be banned completely. It also examines the criteria or requirements that can be used to identify AI applications or technologies that should be banned and in which contexts.

A. BAN OF HIGH-RISK AI APPLICATIONS:

Banning high-risk AI applications requires careful consideration of contextual factors, application-specific risks, and the availability of alternatives. A nuanced approach based on the precautionary principle, continuous evaluation, and the identification of severe harms, lack of ethical frameworks, and public consensus can help determine when a complete ban is necessary. It is crucial to strike a balance between protecting individuals and society from potential risks while fostering innovation and the responsible use of AI technologies[xi].

- *Contextual Considerations:* The decision to ban high-risk AI applications should be context-specific, considering factors such as the potential harm to individuals or society, the availability of alternative solutions, and the feasibility of implementing regulatory measures. While outright bans may be warranted in certain cases, a blanket ban on an entire technology may hinder innovation and the development of beneficial applications. Therefore, a nuanced approach that carefully assesses the risks and benefits is necessary.

- *Precautionary Principle:* The precautionary principle suggests that if there are reasonable grounds to believe that an AI application poses significant risks, preventive measures should be taken even in the absence of full scientific certainty. In situations where the potential harms outweigh the benefits and there is insufficient evidence to guarantee safety, a complete ban on high-risk AI applications may be justified. For instance, the European Union's ban on AI-enabled systems for indiscriminate surveillance exemplifies the application of the precautionary principle.

- *Continuous Evaluation:* Given the evolving nature of AI technologies and their potential impact, a ban on high-risk AI applications should be subject to continuous evaluation. Regular assessments can help determine the effectiveness of regulatory measures, assess emerging risks, and ensure that the ban remains relevant and proportional to the risks involved. This adaptive approach allows for a more nuanced response to the dynamic AI landscape.

B. CRITERIA FOR IDENTIFYING BANNED AI APPLICATIONS:

- *Potential for Severe Harm:* High-risk AI applications that have the potential to cause severe harm, such as endangering human lives, compromising fundamental rights, or exacerbating social inequalities, should be considered for a ban. Examples include AI systems used in autonomous weapons, AI-enabled social scoring systems that violate privacy and civil liberties, or AI-driven financial systems that facilitate fraud or market manipulation.
- *Lack of Ethical or Legal Frameworks:* AI applications that operate in the absence of clear ethical or legal frameworks pose significant risks and may warrant a ban. This can include AI technologies that violate privacy, engage in unethical data collection or manipulation, or perpetuate discriminatory practices. Clear criteria and guidelines should be established to identify such applications and prohibit their use in specific contexts.
- *Public Consensus and Expert Input:* The process of identifying AI applications for a complete ban should involve extensive public engagement, including consultations, debates, and expert input. Engaging multiple stakeholders helps ensure diverse perspectives are considered, prevents undue concentration of power in decision-making, and fosters transparency and accountability. For instance, the European Commission's AI regulatory proposal involves public consultation and stakeholder engagement to shape the criteria for high-risk AI systems.

11. WHAT INITIATIVES OR GOVERNMENT ACTION CAN INCREASE PUBLIC TRUST IN AI DEPLOYMENT TO ENCOURAGE MORE PEOPLE TO USE AI?

Public trust is essential for the widespread adoption and acceptance of AI technologies. The following explores the initiatives and actions that the Australian government can undertake to increase public trust in AI deployment. The aim is to provide an analysis of the measures that can be taken to foster trust and encourage more people to use AI.

- *Independent Oversight:* Establishing an independent oversight body or regulatory authority dedicated to monitoring and ensuring the responsible use of AI can enhance public trust. This body can provide oversight, investigate complaints, and enforce compliance with AI regulations, thereby instilling confidence in the transparency and accountability of AI systems.
- *Data Governance and Privacy Protection:* Strengthening data governance frameworks and privacy protections is crucial for fostering public trust. The government can enforce stringent data protection regulations, ensure individuals have control over their personal data, and implement measures to prevent unauthorised access or misuse of data in AI systems.
- *Transparency and Explainability:* To enhance public trust, the Australian government should prioritise transparency and explainability in AI deployment. This includes providing clear explanations of how AI systems make decisions, ensuring that individuals understand the basis on which they are being assessed or served by AI algorithms. Initiatives such as algorithmic impact assessments, where AI systems are audited for fairness, bias, and accountability, can contribute to increased transparency. The use of open-source AI algorithms and making public datasets available can also foster transparency and enable independent scrutiny.
- *Robust Security Measures:* Addressing cybersecurity risks associated with AI systems is essential for building public trust. The government can promote the adoption of cybersecurity best practices and establish clear guidelines for AI developers and users to ensure the security and integrity of AI systems and protect against potential threats and vulnerabilities.

- *Ethical Guidelines and Standards:* The development and adherence to ethical guidelines and standards can help build public trust in AI deployment. The Australian government can establish a framework that promotes responsible and ethical AI practices, ensuring that AI technologies are designed, developed, and deployed with considerations for privacy, security, and fairness. By engaging with experts, academia, and industry stakeholders, the government can create comprehensive guidelines that address the specific challenges and risks associated with AI deployment in various domains.
- *Independent Auditing and Certification:* Introducing independent auditing and certification processes for AI systems can provide assurance to the public that AI technologies meet ethical and technical standards. Independent auditors can assess the fairness, transparency, and accountability of AI systems, and certifications can serve as a mark of trustworthiness for AI deployments.
- *Public Engagement and Collaboration:* Engaging the public in the decision-making processes around AI deployment is crucial for building trust. The Australian government can facilitate public consultations, forums, and citizen deliberation exercises to gather diverse perspectives and address concerns related to AI. By involving citizens in the development of policies and regulations, the government can demonstrate transparency, accountability, and responsiveness to public needs. Collaborative initiatives between the government, industry, civil society organisations, and academia can also promote responsible AI practices and foster public trust.
- *Regulatory Framework:* A robust regulatory framework is essential to ensure accountability and mitigate risks associated with AI deployment. The Australian government can establish clear guidelines and regulations that address data protection, privacy, bias, and discrimination concerns. Regulatory measures can include mandatory impact assessments, regular audits, and compliance requirements for AI systems. The government should collaborate with international organisations and align its regulatory efforts with global standards to ensure consistency and promote trust in cross-border AI applications.
- *Responsible Procurement Practices:* The government can lead by example in AI procurement by adopting responsible AI practices and considering ethical considerations in the selection and use of AI technologies. Implementing guidelines that encourage responsible procurement practices can influence private sector organisations to follow suit, thereby enhancing public trust in AI.
- *Encouraging Ethical AI Innovation:* The government can incentivise and support the development of ethical AI technologies by providing funding and resources for research and development initiatives. Supporting projects that prioritise ethical considerations, fairness, and societal impact can demonstrate the government's commitment to responsible AI deployment.
- *Education and Awareness:* Promoting AI literacy and awareness among the public is crucial for building trust and encouraging more people to use AI. The Australian government can invest in educational initiatives that aim to demystify AI, explain its benefits and limitations, and address misconceptions. By providing accessible information and resources, the government can empower individuals to make informed decisions about AI usage and alleviate concerns related to its deployment.

Building public trust in AI deployment requires a multifaceted approach that prioritises transparency, ethical guidelines, public engagement, regulatory frameworks, and education. The Australian government can take proactive measures to foster trust by implementing initiatives such as transparency requirements, ethical guidelines, public consultations, robust regulations, and educational campaigns. By addressing concerns related to AI ethics, accountability, and fairness, the government can create an environment that encourages the responsible and beneficial use of AI technologies[xii].

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?

Banning high-risk AI activities, such as social scoring or facial recognition technology in certain circumstances, can have significant implications for Australia's tech sector and its trade and export relationships with other countries. The following explores the potential impact of such bans to provide an analysis of the consequences for the tech sector and international trade.

IMPACT ON THE TECH SECTOR:

- *Innovation and Competitiveness:* Banning high-risk AI activities may hinder innovation within Australia's tech sector. Restrictive regulations can discourage local and foreign investment and research in these technologies, limiting the development of cutting-edge AI solutions. As a result, Australia's tech sector may face challenges in maintaining competitiveness in the global market.

- *Loss of Market Opportunities:* Banning high-risk AI activities could limit the domestic market opportunities for Australian tech companies. If these technologies are in demand globally, Australian companies may miss out on valuable business prospects and partnerships. This loss of market share may impede the growth and expansion of the tech sector.

- *Talent Drain:* Restrictive policies on high-risk AI activities could lead to a brain drain of talented professionals. Skilled individuals may seek opportunities in countries with more favourable regulatory environments, where they can freely engage in research and development of high-risk AI technologies. This talent exodus may weaken Australia's tech ecosystem.

TRADE AND EXPORT RELATIONS:

- *Trade Barriers:* Bans on high-risk AI activities can create trade barriers and strain diplomatic relationships. Countries that are proponents of these technologies may view such bans as protectionist measures or unjustified restrictions. This can lead to retaliatory actions, trade disputes, and hindered collaborations in the tech sector.

- *Export Limitations:* Banning high-risk AI activities may limit Australia's ability to export AI technologies to countries that have not implemented similar bans. This could curtail potential revenue streams and negatively impact Australia's export market. The loss of export opportunities may weaken Australia's position in the global tech trade landscape.

- *International Harmonisation:* Australia's decisions to ban high-risk AI activities may influence international discussions and negotiations regarding the regulation of these technologies. By taking a proactive stance on risk mitigation, Australia can contribute to shaping global standards and norms, potentially leading to a more harmonised and responsible AI ecosystem.

- *Case Study: Facial Recognition Technology.* The ban on facial recognition technology in certain circumstances, such as in public spaces or for mass surveillance purposes, has gained traction in several countries. For example, the city of San Francisco in the United States banned the use of facial recognition technology by government agencies. This decision sparked debates on privacy, civil liberties, and the potential biases and risks associated with facial recognition.

While such bans aim to protect individual rights and mitigate risks, they can also impact the tech sector. Facial recognition technology has applications in various industries, including security, marketing, and healthcare. Banning the technology entirely may hinder innovation and limit the capabilities of Australian companies operating in these sectors.

Banning high-risk AI activities, like social scoring or facial recognition technology in certain circumstances, can have significant implications for Australia's tech sector and trade relations.

While such bans may address concerns related to privacy and ethics, they can hinder innovation, limit market opportunities, and strain international trade relations. Balancing risk mitigation with the promotion of responsible AI practices and international cooperation is essential for Australia to navigate the complex landscape of AI regulation while fostering a thriving tech sector and maintaining strong trade ties[xiii].

13. WHAT CHANGES (IF ANY) TO AUSTRALIAN CONFORMITY INFRASTRUCTURE MIGHT BE REQUIRED TO SUPPORT ASSURANCE PROCESSES TO MITIGATE AGAINST POTENTIAL AI RISKS?

As the deployment of AI technologies becomes increasingly prevalent, ensuring robust assurance processes is crucial to mitigate potential risks. The following section examines the potential changes required in Australian conformity infrastructure to support effective assurance processes in AI deployment.

STRENGTHENING REGULATORY FRAMEWORKS:

Strengthening regulatory frameworks by updating conformity assessment procedures to address specific AI risk factors such as algorithmic bias, privacy concerns, and transparency requirements maybe required in Australia's conformity assessment procedures. This tailored approach ensures effective risk mitigation and compliance. Another important step is introducing pre-market assessments for AI technologies, which evaluate ethical and technical standards before market entry, enhancing conformity infrastructure.

- *Updating Conformity Assessment Procedures:* Australian conformity infrastructure needs to adapt to the unique challenges posed by AI technologies. This includes revising conformity assessment procedures to encompass specific AI risk factors such as algorithmic bias, privacy concerns, and transparency requirements. Tailoring conformity assessment processes to AI applications will ensure effective risk mitigation and compliance.
- *Regulating Pre-market Assessments:* Introducing pre-market assessments for AI technologies can enhance conformity infrastructure. This would require manufacturers or developers to undergo rigorous evaluation of their AI systems before market entry, ensuring compliance with ethical and technical standards. Similar approaches have been implemented in sectors such as medical devices to assess safety and effectiveness prior to commercialisation.
- *Addressing explainability and interpretability of AI systems:* Developing standards and guidelines to ensure the transparency and interpretability of AI algorithms.
- *Evaluating the scalability of conformity assessments:* Incorporating mechanisms to assess the scalability of AI systems and adapt assurance processes accordingly.

ESTABLISHING INDEPENDENT THIRD-PARTY AUDITS:

Establishing independent third-party audits is another crucial element. By incorporating provisions for these audits, Australian conformity infrastructure can verify the compliance of AI systems in an unbiased manner. These audits evaluate algorithmic fairness, data privacy, and security measures, fostering public trust and confidence. Encouraging collaboration and knowledge-sharing among experts from different domains further enhances the effectiveness of audits and assessments.

- *Incorporating External Audits:* Australian conformity infrastructure should include provisions for independent third-party audits to verify the compliance of AI systems. These audits could evaluate algorithmic fairness, data privacy, and security measures. Independent auditors would provide an unbiased assessment of AI technologies, enhancing public trust and confidence.
- *Promoting interdisciplinary expertise:* Encouraging collaboration and knowledge-sharing among experts from different domains to enhance the effectiveness of audits and assessments.

- *Learning from Global Practices:* International examples, such as the EU's General Data Protection Regulation (GDPR) and its certification framework, can serve as valuable references for Australian conformity infrastructure. Embracing global best practices in AI conformity assessment can streamline processes, promote harmonisation, and facilitate international trade.

ENCOURAGING COLLABORATION AND STANDARDS DEVELOPMENT:

Encouraging collaboration and standards development is essential. Building a collaborative ecosystem involving government agencies, industry stakeholders, academia, and civil society organisations enables comprehensive approaches to address AI risks and ensure accountability. Actively supporting the development and adoption of industry-specific standards for AI technologies provides clear benchmarks for risk assessment, data handling, and system performance.

- *Collaboration between Stakeholders:* Building a collaborative ecosystem involving government agencies, industry stakeholders, academia, and civil society organisations is essential for effective conformity infrastructure. Joint efforts in developing standards, guidelines, and benchmarks will enable a comprehensive approach to address AI risks and ensure accountability.
- *Promoting Industry Standards:* Australian conformity infrastructure should actively support the development and adoption of industry-specific standards for AI technologies. Standards can provide clear benchmarks for risk assessment, data handling, and system performance, facilitating transparency, interoperability, and responsible AI practices.
- *Monitoring and evaluating conformity infrastructure:* Establishing mechanisms for ongoing monitoring and evaluation of the conformity infrastructure's effectiveness in mitigating AI risks.
- *Case Study: European Union's AI Regulation.* The European Union's proposed AI regulation exemplifies efforts to adapt conformity infrastructure for AI risk mitigation. The regulation sets out requirements for high-risk AI systems, including conformity assessments, transparency obligations, and human oversight. By establishing clear rules and conformity procedures, the EU aims to foster public trust and ensure responsible AI deployment.

To ensure continuous improvement, mechanisms for monitoring and evaluating conformity infrastructure's effectiveness in mitigating AI risks should be established. Ongoing assessment enables the infrastructure to adapt to evolving AI technologies and risks, ensuring its efficacy in supporting assurance processes.

Drawing inspiration from the European Union's proposed AI regulation, which sets out requirements for high-risk AI systems, including conformity assessments, transparency obligations, and human oversight, Australia gains valuable insights for shaping its own policies and frameworks. By leveraging this case study, Australia can strengthen its conformity infrastructure and support effective assurance processes and risk mitigation in AI deployment, ultimately fostering public trust and responsible AI practices.

To support effective assurance processes and mitigate potential AI risks, Australian conformity infrastructure requires adaptation and enhancement. Strengthening regulatory frameworks, establishing independent audits, promoting collaboration, and encouraging industry standards are vital components of an effective conformity infrastructure. By implementing these changes, Australia can foster responsible AI practices, enhance risk mitigation, and build public trust in the deployment of AI technologies[xiv].

RISK-BASED APPROACHES

14. DO YOU SUPPORT A RISK-BASED APPROACH FOR ADDRESSING POTENTIAL AI RISKS? IF NOT, IS THERE A BETTER APPROACH?

The management of potential risks associated with AI technologies is a critical concern for policymakers. The following examines the suitability and effectiveness of a risk-based approach in addressing AI risks in Australia and aims to provide a comprehensive evaluation of the advantages and limitations of a risk-based approach and explore alternative strategies, if applicable.

ADVANTAGES OF A RISK-BASED APPROACH:

A risk-based approach offers several advantages in the regulation of AI applications, enabling targeted mitigation, flexibility, and adaptability, and promoting industry innovation and growth. By focusing resources on high-risk AI applications, regulatory bodies can effectively manage potential harm while avoiding excessive regulations on low-risk applications. This approach allows regulators to adjust mitigation strategies and regulatory requirements as technologies and contexts evolve rapidly. Rather than stifling innovation, a risk-based approach fosters a conducive environment for the development and adoption of AI technologies by providing clear guidelines and expectations. The United States Food and Drug Administration's (FDA) risk-based approach in regulating medical devices serves as a relevant case study, demonstrating how a balanced approach between safety and innovation can be achieved through appropriate oversight and timely access to new technologies.

- *Targeted Mitigation:* A risk-based approach enables the allocation of resources and efforts towards high-risk AI applications, focusing on areas that pose significant potential harm. By prioritising high-risk activities, regulatory bodies can concentrate their attention on critical areas, ensuring effective risk management while avoiding overly burdensome regulations on low-risk applications.

- *Flexibility and Adaptability:* A risk-based approach allows for flexibility and adaptability to evolving technologies and contexts. As AI technologies rapidly advance, a rigid regulatory framework may struggle to keep pace. A risk-based approach empowers regulators to adjust mitigation strategies and regulatory requirements based on emerging risks and changing circumstances.

- *Industry Innovation and Growth:* By focusing on risks rather than stifling innovation, a risk-based approach can foster a conducive environment for the development and adoption of AI technologies. It encourages responsible innovation by providing clear guidelines and expectations, enabling businesses to navigate the regulatory landscape and develop safe and ethical AI solutions.

- *Case Study: United States Food and Drug Administration (FDA).* As mentioned previously, the FDA's risk-based approach in regulating medical devices provides a relevant case study. The FDA assesses medical devices based on their potential risks to patient safety and public health. High-risk devices, such as implantable pacemakers, undergo a stringent regulatory process, including pre-market approval, while low-risk devices follow a less burdensome regulatory pathway. This risk-based approach strikes a balance between safety and innovation, ensuring appropriate oversight while facilitating timely access to new technologies.

LIMITATIONS AND CONSIDERATIONS:

The implementation of a risk-based approach in AI regulation comes with limitations and considerations that need to be addressed. Assessing and quantifying AI risks can be challenging due to the complexity and subjectivity involved, requiring robust methodologies and data-driven frameworks. Furthermore, a risk-based approach may overlook systemic risks associated with AI technologies, necessitating the inclusion of broader ethical considerations and societal impact assessments to ensure comprehensive risk management. These limitations and considerations highlight the need for a multidimensional approach in AI regulation that combines risk assessment with broader societal implications.

- *Risk Assessment Challenges:* Assessing and quantifying AI risks can be complex and subjective, especially when dealing with emerging technologies. Determining the severity and likelihood of potential harms may involve uncertainty and require expert judgment. Robust methodologies and data-driven risk assessment frameworks are essential for effective implementation of a risk-based approach.
- *Addressing Systemic Risks*: A risk-based approach may overlook systemic risks inherent in AI technologies, such as societal implications and unintended consequences. It is crucial to complement the risk-based approach with broader ethical considerations and societal impact assessments to capture these systemic risks and ensure comprehensive risk management.

COMPLEMENTARY APPROACHES:

In addition to a risk-based approach, complementary approaches are essential for effective AI governance. Incorporating ethical and human-centric frameworks alongside risk assessment provides a more holistic approach, addressing broader societal concerns beyond individual risks. Principles such as fairness, transparency, and accountability guide the development and deployment of AI technologies. Furthermore, collaborative governance that involves multiple stakeholders, including industry, academia, civil society, and the public, enhances decision-making processes. By leveraging diverse perspectives and expertise, comprehensive frameworks can be developed to address risks, ethical considerations, and public concerns, ensuring responsible and inclusive AI governance.

- *Ethical and Human-Centric Frameworks:* Incorporating ethical guidelines and human-centric considerations alongside a risk-based approach can provide a more holistic approach to AI governance. This includes principles such as fairness, transparency, and accountability, which address broader societal concerns beyond individual risks.
- *Collaborative Governance:* Emphasising multi-stakeholder collaboration and engagement can enhance AI governance. By involving industry, academia, civil society, and the public in the decision-making process, diverse perspectives and expertise can be leveraged to develop comprehensive frameworks that address risks, ethical considerations, and public concerns.

A risk-based approach offers several advantages for addressing AI risks in Australia, including targeted mitigation, flexibility, and fostering innovation. However, it is crucial to acknowledge the limitations and challenges associated with risk assessment and systemic risks. Complementary approaches, such as ethical frameworks and collaborative governance, can enhance the effectiveness of AI governance. Striking a balance between risk-based approaches and broader ethical considerations will help ensure responsible and beneficial AI deployment in Australia[xv].

15. WHAT DO YOU SEE AS THE MAIN BENEFITS OR LIMITATIONS OF A RISK-BASED APPROACH? HOW CAN ANY LIMITATIONS BE OVERCOME?

The following aims to assess the benefits and limitations of a risk-based approach in AI governance. Additionally, potential strategies to overcome these limitations are explored to ensure effective risk management in the context of AI technologies.

BENEFITS OF A RISK-BASED APPROACH:

A risk-based approach in addressing potential AI risks offers several benefits that contribute to effective governance. As mentioned previously, by efficiently allocating limited resources, regulatory bodies can focus their efforts on areas of high potential risk, ensuring that mitigation measures are implemented where they are most needed. The flexibility and adaptability of a risk-based approach allow regulatory frameworks to stay current with evolving technologies and emerging risks, enabling them to remain relevant and effective over time. Additionally, a risk-based approach promotes innovation by providing tailored regulations that address specific risks, creating a conducive environment for responsible AI development and exploration of the technology's potential while mitigating associated risks.

- *Targeted Resource Allocation:* A risk-based approach allows for the efficient allocation of limited resources, focusing efforts on areas of high potential risk. By prioritising high-risk AI applications, regulatory bodies can ensure that mitigation measures are implemented where they are most needed, maximising the effectiveness of oversight and control.
- *Flexibility and Adaptability:* A risk-based approach provides flexibility to adapt regulatory measures in response to evolving technologies and emerging risks. This adaptability enables regulatory frameworks to stay current with the rapid advancements in the AI landscape, ensuring that governance practices remain relevant and effective over time.
- *Promotion of Innovation:* By taking a risk-based approach, innovation is not stifled but encouraged. Regulations that are tailored to specific risks foster a conducive environment for responsible AI development, allowing businesses to innovate and explore the potential of AI technologies while simultaneously addressing associated risks.
- *Case Study: European Union's Risk-Based Approach in Data Protection*. The General Data Protection Regulation (GDPR) in the European Union (EU) provides a relevant case study of a risk-based approach in AI governance. The GDPR requires organisations to implement appropriate data protection measures based on the level of risk associated with the processing of personal data. This risk-based approach ensures that resources are focused on high-risk activities, such as sensitive data processing, while reducing the burden on low-risk activities.

LIMITATIONS AND OVERCOMING CHALLENGES:

While a risk-based approach offers advantages in addressing AI risks, it is important to recognise and overcome its limitations and challenges. As mentioned previously, assessing and quantifying risks in the AI domain can be subjective and uncertain, necessitating the development of robust methodologies and transparent frameworks to enhance consistency and reduce subjectivity. Additionally, a risk-based approach may overlook systemic risks, highlighting the need for broader ethical considerations, societal impact assessments, and interdisciplinary collaboration to ensure comprehensive AI governance. Striking a balance between innovation and risk is crucial, and agile regulatory approaches, along with collaboration between regulators, industry, and academia, can foster responsible innovation while effectively managing risks.

- *Subjectivity and Uncertainty:* Assessing and quantifying risks in the AI domain can be subjective and uncertain. Overcoming this challenge requires the development of robust risk assessment methodologies that incorporate expert input, empirical evidence, and ongoing evaluation of AI technologies. Transparent guidelines and frameworks can enhance consistency and reduce subjectivity in risk assessments.
- *Addressing Systemic Risks:* A risk-based approach may focus primarily on individual risks associated with AI applications and overlook systemic risks. To address this limitation, it is crucial to complement the risk-based approach with broader ethical considerations, societal impact assessments, and interdisciplinary collaboration. Ethical frameworks and human-centric principles can help capture systemic risks and ensure a comprehensive approach to AI governance.
- *Balancing Innovation and Risk:* Striking a balance between fostering innovation and managing risks is essential. To address this, regulatory bodies can adopt agile and adaptive regulatory approaches that promote responsible innovation. Encouraging collaboration between regulators, industry, and academia can facilitate the development of guidelines and best practices that support innovation while mitigating risks effectively.

A risk-based approach in AI governance offers numerous benefits, including targeted resource allocation, flexibility, and the promotion of innovation. However, limitations such as subjectivity, systemic risks, and balancing innovation and risk must be addressed. By developing robust risk assessment methodologies, incorporating ethical considerations, and promoting collaborative governance, these limitations can be overcome. It is crucial to strike a balance between risk management and innovation to ensure the responsible and beneficial deployment of AI technologies[xvi].

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?

The following aims to assess the suitability of a risk-based approach in AI governance based on organisational factors such as size, AI maturity, and resources and provides insights into the effectiveness and applicability of a risk-based approach in different sectors, AI applications, and organisations.

SIZE OF ORGANISATION:

- *Large Organisations:* Larger organisations often possess greater resources and capabilities to implement comprehensive risk management frameworks. A risk-based approach allows them to allocate resources efficiently by prioritising high-risk AI applications and deploying advanced risk mitigation strategies e.g. multinational companies can leverage their extensive resources to conduct thorough risk assessments and invest in robust control mechanisms.
- *Small and Medium-sized Enterprises (SMEs):* SMEs may have limited resources and expertise to implement complex risk management frameworks. In such cases, a risk-based approach can be tailored to their specific context, focusing on the most critical AI risks. Regulatory bodies can provide guidance and support to SMEs, facilitating compliance with risk management requirements while considering their unique constraints.

AI MATURITY:

- *Emerging Technologies:* In sectors where AI technologies are still in their early stages of development, a risk-based approach can facilitate a proactive approach to identify and address potential risks. Regulatory frameworks can encourage organisations to implement risk mitigation measures in parallel with technology advancement. For instance, in the autonomous vehicles sector, early identification of safety risks can guide regulatory actions to ensure responsible deployment.
- *Mature Technologies:* In sectors where AI technologies have reached a higher level of maturity, a risk-based approach can help refine governance practices by targeting specific risks associated with the technology's application. For example, in healthcare, mature AI applications like medical image analysis systems may require risk assessment and mitigation measures to ensure patient safety and data privacy.

RESOURCE AVAILABILITY:

- *High-Resource Organisations:* Organisations with ample resources can invest in advanced risk assessment methodologies, comprehensive monitoring systems, and dedicated risk management teams. A risk-based approach allows them to leverage their resources effectively to mitigate potential AI risks. Examples include large financial institutions implementing robust anti-money laundering systems or government agencies utilising sophisticated surveillance technologies.
- *Limited-Resource Organisations:* Organisations with limited resources may struggle to implement complex risk management frameworks. In such cases, regulatory bodies can provide guidance, best practices, and standardised risk assessment tools to assist organisations in managing AI risks within their resource constraints. Collaboration with industry associations and knowledge-sharing platforms can help smaller organisations overcome resource limitations.

A risk-based approach in AI governance can be adapted to suit different sectors, AI applications, and organisations based on their size, AI maturity, and available resources. While larger organisations and those with more advanced AI maturity and resources may be better equipped to implement comprehensive risk management frameworks, regulatory bodies should provide support and guidance to smaller organisations and emerging sectors. Tailoring the risk-based approach to the specific context and resource availability of organisations can ensure effective risk management and responsible deployment of AI technologies[xvii].

17. WHAT ELEMENTS SHOULD BE IN A RISK-BASED APPROACH FOR ADDRESSING POTENTIAL AI RISKS? DO YOU SUPPORT THE ELEMENTS PRESENTED IN ATTACHMENT C?

This following examines the elements proposed in Attachment C as part of a draft risk-based approach for addressing potential AI risks in Australia. It aims to assess the suitability and effectiveness of these elements in promoting safe and robust AI practices, as well as increasing public trust and confidence.

IMPACT ASSESSMENTS:

The inclusion of impact assessments is crucial to ensure organisations properly consider and mitigate potential risks associated with AI. By publishing the results of these assessments, transparency is enhanced, enabling stakeholders to understand how risks are being managed. Additionally, external peer review of impact assessments for high-risk AI applications can provide an additional layer of scrutiny and expertise, further improving the quality and reliability of risk assessment processes.

- *Example:* The European Union's General Data Protection Regulation (GDPR) mandates impact assessments for certain high-risk processing activities, including those involving AI technologies. This requirement ensures organisations thoroughly evaluate the potential risks and implement appropriate safeguards.

NOTICES:

Notices play a vital role in informing individuals when AI systems are used in ways that significantly impact them. By providing this notification, individuals can exercise their rights, seek reviews of decisions, and ultimately develop trust in the system. Transparent communication about the use of AI helps bridge the information gap and avoids surprises that may undermine public trust.

- *Example:* In the financial sector, where AI algorithms are increasingly used for credit scoring, providing notices to individuals about the factors influencing their credit decisions fosters transparency and enables customers to understand and challenge those decisions if necessary.

HUMAN IN THE LOOP/OVERSIGHT ASSESSMENTS:

Determining when human oversight is appropriate is essential for minimising risks and ensuring public trust. Assessments should consider the complexity of decisions, level of discretion, potential damage of incorrect decisions, and the expertise required. However, it is important to strike a balance between meaningful human involvement and efficiency, especially in cases where automation and scale are necessary.

- *Example:* The use of automated facial recognition technology by law enforcement agencies highlights the need for human oversight assessments. While automation can enhance efficiency, introducing human involvement in the form of review or monitoring can mitigate potential biases and ensure ethical use.

EXPLANATIONS:

Explanations contribute to building public trust by providing clarity on AI decisions or outcomes. Individuals affected by AI decisions should be able to understand the factors influencing those decisions, fostering transparency and accountability. Clear explanations can help address concerns related to bias, fairness, and ethics, enabling individuals to assess the legitimacy of AI-driven outcomes.

- *Example:* Explainable AI techniques, such as interpretable machine learning models, can provide explanations for decisions made by AI systems, enabling affected individuals to understand the rationale behind those decisions.

TRAINING:

Proper training is essential for employees involved in AI design, implementation, and oversight. Adequate training equips employees with the necessary knowledge to understand potential risks, mitigate them effectively, and provide oversight. The level of training should align with the level of potential risk, ensuring competent and qualified individuals are responsible for AI systems.

- *Example:* In the healthcare sector, where AI algorithms are used to assist in medical diagnoses, healthcare professionals require specific training to interpret and validate AI recommendations, ensuring patient safety and avoiding over-reliance on automated systems.

MONITORING AND DOCUMENTATION:

Ongoing monitoring is critical to ensure AI systems operate as intended and detect any adverse or unintended impacts, such as bias or unintended consequences. The intensity of monitoring should correspond to the level of risk, with more frequent tests for high-risk AI applications. Documentation of the AI's design, function, and implementation facilitates understanding of potential risks, accountability, and decision-making.

- *Example:* The financial industry implements comprehensive monitoring systems to detect anomalies and potential risks in automated trading algorithms, ensuring market stability and minimising the impact of erroneous AI-driven trades.

The elements proposed in Attachment C of a risk-based approach for addressing AI risks in Australia provide a comprehensive framework for promoting safe and robust AI practices. Impact assessments, notices, human in the loop/oversight assessments, explanations, training, and monitoring and documentation are all crucial components that contribute to risk mitigation and public trust. However, their successful implementation will depend on clear guidelines, collaboration between stakeholders, and appropriate regulatory oversight. Striking the right balance between risk management and innovation will be essential to ensure the benefits of AI are realised while minimising potential harms[xviii].

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?

The following examines the integration of an AI risk-based approach into existing assessment frameworks, such as privacy, and risk management processes and aims to explore how the incorporation of AI risk considerations can streamline and reduce potential duplication in assessment and management practices.

ALIGNMENT WITH PRIVACY ASSESSMENT FRAMEWORKS:

Integrating AI risk considerations into existing privacy assessment frameworks can provide a comprehensive approach to address the potential privacy risks associated with AI technologies. By leveraging the existing structures and processes of privacy assessments, organisations can streamline their efforts and avoid duplicative assessments. This integration ensures that privacy risks related to AI are properly evaluated, mitigated, and monitored.

- *Example:* The Privacy Impact Assessment (PIA) framework, widely used in various jurisdictions, can be expanded to incorporate specific AI risk factors. This integration enables organisations to assess the privacy implications of AI technologies while leveraging the established processes and methodologies of PIAs.

HARMONISATION WITH RISK MANAGEMENT PROCESSES:

Integrating an AI risk-based approach into existing risk management processes, such as enterprise risk management (ERM), can enhance the overall effectiveness and efficiency of risk mitigation efforts. By aligning AI-specific risks with broader organisational risk frameworks, duplication of efforts and resources can be minimised. This integration allows organisations to holistically manage and prioritise risks across different domains, including AI.

- *Example:* The ISO 31000 standard for risk management provides a flexible framework that can accommodate the incorporation of AI-specific risks. By extending the risk assessment and treatment processes to include AI-related risks, organisations can effectively integrate AI risk management into their existing risk management practices.

SYNERGY WITH COMPLIANCE FRAMEWORKS:

Integrating AI risk considerations into existing compliance frameworks, such as data protection regulations or industry-specific standards, can facilitate a cohesive and unified approach to risk management. By aligning AI risks with existing compliance requirements, organisations can ensure compliance while effectively addressing the unique risks posed by AI technologies.

- *Example:* The integration of AI risk considerations into the requirements of the General Data Protection Regulation (GDPR) enables organisations to assess and manage AI-related privacy risks within the broader compliance framework. This approach ensures that organisations meet their legal obligations while addressing the specific risks associated with AI.

CHALLENGES AND CONSIDERATIONS:

- *Standardisation:* Ensuring consistency and standardisation across different frameworks and processes can be challenging, as AI risks may require specific assessment methodologies and mitigation strategies.
- *Expertise and Awareness:* Organisations need to develop AI-specific expertise and awareness among risk management and assessment professionals to effectively identify and address AI-related risks.
- *Continuous Monitoring:* Given the rapidly evolving nature of AI technologies, continuous monitoring is crucial to detect and mitigate emerging risks, necessitating regular updates and revisions to existing frameworks.

Integrating an AI risk-based approach into existing assessment frameworks and risk management processes holds significant potential for streamlining efforts and reducing duplication. By aligning with privacy assessment frameworks, risk management processes, and compliance frameworks, organisations can effectively address AI-related risks while leveraging established methodologies. However, standardisation, expertise, and continuous monitoring remain critical challenges that need to be addressed for successful integration[xix].

19. HOW MIGHT A RISK-BASED APPROACH APPLY TO GENERAL PURPOSE AI SYSTEMS, SUCH AS LARGE LANGUAGE MODELS (LLMS) OR MULTIMODAL FOUNDATION MODELS (MFMS)?

The following explores how a risk-based approach can be applied to general-purpose AI systems, specifically focusing on large language models (LLMs) and multimodal foundation models (MFMs). It aims to provide insights into effectively managing the risks associated with these powerful AI technologies.

IDENTIFYING AND ASSESSING RISKS:

A risk-based approach involves identifying and assessing potential risks associated with LLMs and MFMs, considering their broad range of applications and impact on various domains. This includes evaluating risks such as biased outputs, misinformation propagation, and ethical concerns arising from the deployment and use of these models.

- *Example:* OpenAI's GPT model, a widely known LLM, demonstrated risks of generating biased or misleading content, potentially influencing public opinion, or reinforcing stereotypes. Risk assessment methodologies should involve conducting comprehensive audits and evaluations of these models to understand their limitations, vulnerabilities, and potential consequences.

RISK MITIGATION STRATEGIES:

Once risks are identified, a risk-based approach focuses on developing appropriate mitigation strategies to minimise or manage those risks. This includes implementing safeguards, guidelines, and interventions to address specific risks associated with LLMs and MFMs.

- *Example:* Implementing bias detection and mitigation techniques in LLMs can help reduce the generation of biased or discriminatory outputs. OpenAI's work on fine-tuning GPT to reduce gender and racial bias is an example of a risk mitigation strategy that aims to enhance the fairness and inclusivity of the model's outputs.

CONTINUOUS MONITORING AND EVALUATION:

A risk-based approach acknowledges the dynamic nature of AI systems and emphasises the importance of continuous monitoring and evaluation. This involves regularly assessing the performance and impact of LLMs and MFMs, identifying emerging risks, and implementing necessary updates and improvements.

- *Example:* Ongoing audits and third-party evaluations of LLMs, such as the Common Voice initiative by Mozilla, contribute to monitoring their performance and identifying potential risks related to privacy, security, and biased outputs. This continuous evaluation process helps ensure the responsible and accountable use of these AI systems.

CHALLENGES AND CONSIDERATIONS:

- *Complexity of Risk Assessment:* Evaluating the risks associated with LLMs and MFMs requires a multidimensional approach due to their diverse applications and potential societal impact. Developing comprehensive risk assessment frameworks specific to these models is crucial.

- *Ethical Considerations:* LLMs and MFMs can raise ethical concerns, including potential misuse, information manipulation, and infringement of privacy rights. Incorporating ethical considerations into risk assessments is necessary to ensure responsible deployment and usage.

- *Collaboration and Transparency:* Engaging stakeholders, including AI researchers, policymakers, and the public, in the risk assessment and decision-making processes is essential for transparency and fostering trust in the governance of LLMs and MFMs.

A risk-based approach provides a framework for identifying, assessing, and mitigating risks associated with general-purpose AI systems like LLMs and MFMs. By conducting comprehensive risk assessments, implementing appropriate mitigation strategies, and continuously monitoring their performance, these AI technologies can be utilised responsibly and ethically. However, addressing the complexity of risk assessment, incorporating ethical considerations, and promoting collaboration and transparency remain critical challenges to ensure effective risk management[xx].

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?

The following examines the question of whether a risk-based approach for responsible AI a voluntary self-regulation tool should be or mandated through regulation. Additionally, it explores the scope of application, considering whether it should apply to public or private organisations, developers or deployers, or both. It aims to provide insights into the merits and drawbacks of voluntary and mandated approaches.

VOLUNTARY SELF-REGULATION:

Advantages:

- *Flexibility and Adaptability:* Voluntary self-regulation allows organisations to tailor their risk management strategies to their specific contexts, fostering innovation and accommodating diverse needs.

- *Early Adoption:* Organisations can voluntarily adopt responsible AI practices before formal regulations are established, demonstrating their commitment to ethical AI deployment.

- *Collaboration and Best Practices:* Voluntary frameworks encourage knowledge sharing, collaboration, and the development of industry-wide best practices.

Limitations:

- *Lack of Uniformity:* Voluntary self-regulation can result in inconsistent practices and standards across organisations, potentially leading to gaps in addressing AI risks.
- *Limited Compliance:* Without regulatory mandates, some organisations may choose not to prioritise responsible AI practices, undermining the overall effectiveness of self-regulation.
- *Public Trust Concerns:* Voluntary measures may not sufficiently address public concerns regarding potential risks and may be viewed as inadequate in ensuring ethical AI deployment.
- *Example:* The Partnership on AI is a voluntary alliance of organisations committed to advancing responsible AI. It fosters collaboration, knowledge sharing, and the development of guidelines for AI ethics, but its effectiveness depends on the willingness of organisations to actively participate and adhere to the principles.

MANDATED REGULATION:

Advantages:

- *Consistency and Accountability:* Mandated regulations ensure a level playing field, establishing consistent standards and accountability across all organisations.
- *Public Trust and Protection:* Regulatory mandates provide assurance to the public that AI systems are subject to scrutiny, mitigating risks and safeguarding against potential harms.
- *Enforcement and Penalties:* Regulatory frameworks empower authorities to enforce compliance and impose penalties for non-compliance, incentivising organisations to prioritise responsible AI practices.

Limitations:

- *Adaptability Challenges:* Regulatory mandates may struggle to keep pace with the rapidly evolving AI landscape, potentially hindering innovation and creating compliance burdens.
- *Resource Requirements:* Compliance with regulatory requirements can impose financial and administrative burdens, particularly for smaller organisations with limited resources.
- *Potential Regulatory Capture:* The risk of regulatory capture exists, where powerful organisations influence the regulatory process to serve their interests, potentially undermining the effectiveness of regulation.
- *Example:* The General Data Protection Regulation (GDPR) in the European Union mandates certain requirements for AI systems that process personal data. It aims to protect individuals' privacy rights and enhance transparency and accountability in AI deployments.

APPLICABILITY TO PUBLIC AND PRIVATE SECTORS:

Responsible AI practices should be applicable to both public and private organisations. As mentioned previously, the public sector has a responsibility to lead by example in deploying AI systems that uphold ethical standards, protect citizen rights, and ensure fairness and transparency. By implementing robust regulatory measures, the government can set a benchmark for responsible AI practices, while also addressing potential biases and discriminatory impacts that could disproportionately affect marginalised communities.

Similarly, the private sector plays a significant role in AI development and deployment, and its responsible practices are critical to safeguarding individuals and maintaining public trust. While voluntary initiatives within the industry can encourage responsible behaviour, a regulatory framework is necessary to ensure consistent adherence, accountability, and protection of consumer rights.

Appropriate regulations can mitigate potential risks, including biased algorithms, unauthorised data usage, and lack of transparency.

DEVELOPERS AND DEPLOYERS: SHARED RESPONSIBILITIES:

Developers and deployers share responsibilities in ensuring the responsible deployment of AI systems. Developers play a crucial role in designing AI algorithms and models, and they have the responsibility to ensure fairness, transparency, and compliance with ethical standards. They should address algorithmic biases, conduct impact assessments, and promote transparency throughout the AI development lifecycle.

To address algorithmic biases, developers need to be aware of potential biases present in AI algorithms and take steps to mitigate them. For example, in the recruitment industry, AI-powered hiring systems have been found to perpetuate biases based on gender, race, or other protected characteristics. Responsible developers actively work to identify and rectify these biases to ensure fair and equitable outcomes.

Conducting impact assessments is another responsibility of developers. Thorough impact assessments evaluate the potential risks and consequences of AI system deployment, considering factors such as privacy, security, fairness, and social impacts. For instance, in the healthcare sector, developers need to assess the potential risks of AI systems used for diagnosing diseases, ensuring that the technology doesn't compromise patient privacy or lead to incorrect diagnoses.

Promoting transparency is essential throughout the AI development process. Developers should provide clear documentation of the system's functionalities, limitations, and potential risks to users and stakeholders. Transparency allows users to understand how AI models operate and identify potential biases or limitations in the technology. Companies like OpenAI have developed guidelines for transparency, contributing to responsible AI development.

Deployers, on the other hand, are responsible for the ethical and responsible use of AI systems in real-world applications. They must ensure proper implementation, monitoring, and accountability for the outcomes of AI systems. Deployers should comply with regulations, undertake risk assessments, and establish mechanisms for ongoing monitoring and auditing to detect and rectify unintended consequences.

Proper implementation is crucial for deployers. They need to ensure that AI systems are implemented correctly, considering factors such as data quality, system configuration, and user interface design. For example, autonomous vehicle deployers must ensure that self-driving cars are properly calibrated, have robust safety features, and comply with traffic regulations to minimise the risk of accidents.

Ongoing monitoring and auditing are necessary for deployers to ensure the performance of AI systems in real-world contexts. Regular audits and feedback collection help detect unintended consequences or biases. Social media platforms, for instance, have a responsibility to monitor their AI algorithms for content moderation to prevent the spread of misinformation or harmful content.

Deployers should also comply with relevant regulations and guidelines governing the use of AI systems. Compliance with regulations, such as the General Data Protection Regulation (GDPR) in the European Union, ensures the protection of individuals' privacy and rights when personal data is processed by AI systems.

By combining voluntary self-regulation with mandated regulation and involving both developers and deployers, a comprehensive approach to responsible AI deployment can be achieved. Voluntary initiatives and guidelines encourage responsible behaviour and innovation within the industry, while regulatory mandates provide accountability and assurance to the public. This balanced approach ensures that both the design and use of AI systems prioritise fairness, transparency, and ethical considerations, fostering trust and responsible adoption of AI technologies[xxi].

RECOMMENDATIONS

CREATING REGULATION AND GOVERNANCE FOR SAFE AND RESPONSIBLE AI IN AUSTRALIA

Ensuring safe and responsible AI deployment in Australia requires the implementation of effective regulation, engagement, participation and governance measures. These recommendations highlight key actions and considerations for the Australian government to create the necessary framework:

1. COMPREHENSIVE REGULATORY FRAMEWORK:

1.1. Develop a comprehensive regulatory framework that addresses different stages of the AI lifecycle, including development, deployment, and use. This framework should cover areas such as algorithmic fairness, transparency, accountability, and privacy.

1.2. Incorporate risk-based approaches that consider the potential risks associated with AI applications, considering factors such as the level of harm, complexity, and societal impact. This will help prioritise resources and efforts where they are most needed.

1.3. Draw insights from international best practices and learnings, such as the European Union's General Data Protection Regulation (GDPR), to inform the development of regulations that protect personal data in AI systems. International collaboration and alignment can foster harmonisation and interoperability in the global AI landscape.

2. RISK ASSESSMENT AND MITIGATION:

2.1. Establish a standardised framework for conducting comprehensive risk assessments of AI systems before their deployment. This includes assessing potential biases, privacy concerns, security vulnerabilities, and societal impacts.

2.2. Develop guidelines and best practices for mitigating identified risks and ensuring that AI systems are designed and deployed in a manner that minimises harm and negative consequences.

3. TRANSPARENCY AND ACCOUNTABILITY:

3.1. Mandate transparency requirements to increase public trust and confidence in AI systems. This includes publishing impact assessments, providing explanations for AI decisions, and ensuring meaningful human involvement in critical decision-making processes. Transparency fosters accountability and allows users and stakeholders to understand the functioning of AI systems.

3.2. Implement mechanisms for independent auditing and external peer review of high-risk AI applications to enhance accountability and mitigate potential risks. Independent audits can provide objective assessments of AI systems' performance, identify biases or unintended consequences, and promote responsible behavior among developers and deployers.

3.3. Encourage organisations to adopt ethical AI principles and adhere to responsible AI practices through awareness campaigns and incentives. Promoting responsible behavior can be achieved by providing guidelines, training programs, and incentives that encourage the integration of ethical considerations throughout the AI lifecycle.

4. AUDITABILITY AND EXPLAINABILITY:

4.1. Encourage the development and adoption of AI systems that are auditable and explainable. This involves ensuring that AI algorithms and models can be understood, verified, and audited by independent entities to assess their fairness, transparency, and compliance with ethical standards.

4.2. Support research and development efforts in explainable AI techniques and technologies that provide insights into the decision-making processes of AI systems, particularly in high-risk applications.

5. REGULATION AND ACCOUNTABILITY FOR AUTONOMOUS SYSTEMS:

5.1. Develop specific guidelines and regulations for AI systems with high autonomy levels, such as autonomous vehicles or advanced robotics. These guidelines should address issues related to liability, accountability, and decision-making transparency.

5.2. Ensure that deployers of autonomous systems are accountable for any accidents, errors, or unintended consequences caused by the AI system's actions. This may involve defining clear lines of responsibility and establishing mechanisms for compensation and dispute resolution.

5.3. Develop specific regulations and guidelines for autonomous AI systems that have the capability to make decisions without human intervention. These regulations should ensure transparency, accountability, and safety in the deployment of such systems.

5.4. Define clear boundaries and responsibilities for the decision-making authority of autonomous AI systems to avoid potential ethical and legal challenges.

6. ETHICAL AI REVIEW BOARDS AND ADVISORY GROUPS:

6.1. Establish independent ethical AI review boards composed of multidisciplinary experts from academia, industry, and civil society. These boards can also provide guidance and oversight on the ethical implications of AI systems and ensure that they align with societal values and principles including guidance, recommendations, and oversight on AI-related policies, regulations, and decision-making processes.

6.2. The ethical AI review boards should review and evaluate high-risk AI applications, assess their compliance with ethical standards, and make recommendations for improvements or mitigations where necessary.

6.3. Ensure transparency and accountability in the functioning of ethics boards, with regular reporting and public disclosure of their activities and recommendations.

7. DATA PRIVACY, PROTECTION, GOVERNANCE AND SHARING:

7.1. Strengthen privacy laws and regulations to address the specific challenges posed by AI systems, particularly in handling personal data. This includes ensuring that individuals' rights to privacy are protected, and their personal data is handled in a responsible and secure manner.

7.2. Establish clear guidelines on data collection, storage, and usage in AI applications, ensuring that individuals have control over their data and consent to its use. Striking a balance between data access for innovation and safeguarding privacy rights is essential in the AI ecosystem.

7.3. Foster the development of privacy-preserving AI techniques and technologies to minimise the risks associated with data breaches and unauthorised access. Encouraging research and innovation in privacy-preserving AI can enable the responsible use of data while protecting individuals' privacy.

7.4. Establish robust data governance frameworks to ensure the responsible and ethical use of data in AI applications. Promote data sharing practices that prioritise privacy, security, and informed consent while enabling innovation and societal benefits.

7.5. Encourage collaborations between data providers, AI developers, and deployers to ensure responsible data usage. Foster partnerships that adhere to data protection principles, promote data diversity, and mitigate biases in AI training data.

7.6. Encourage open data practices to facilitate transparency and accountability in AI development and deployment. This includes promoting the sharing of datasets, algorithms, and methodologies (where appropriate) to foster collaboration, peer review, and independent scrutiny.

8. INCENTIVES FOR SAFE AND RESPONSIBLE INNOVATION:

8.1. Provide incentives, such as tax breaks or grants, for organisations that demonstrate responsible AI practices and actively contribute to the development of safe and ethical AI systems.

8.2. Recognise and reward organisations that prioritise transparency, fairness, and accountability in their AI development and deployment efforts, fostering a culture of responsible AI adoption.

8.3. Allocate funding for research initiatives focused on ethical AI, fairness, accountability, and transparency. Encourage collaboration between researchers, industry partners, and government agencies to develop innovative approaches and tools for responsible AI.

8.4. Support research projects that aim to address AI biases, enhance interpretability and explainability of AI systems, and develop methods for detecting and mitigating unintended consequences. Invest in long-term research efforts to ensure the continuous improvement of AI technologies.

8.5. Provide incentives, such as grants, funding, and tax breaks, for organisations and researchers who prioritise responsible AI innovation. Encourage the development of AI systems that align with ethical principles, demonstrate fairness, transparency, and accountability, and address societal challenges.

8.6. Recognise and promote AI initiatives and projects that have a positive societal impact, such as those focused on healthcare, environmental sustainability, or social equality. Encourage collaboration between public and private sectors to foster responsible AI innovation.

9. REGULATORY AGILITY AND ADAPTABILITY:

9.1. Foster a regulatory environment that is agile, adaptable, and responsive to technological advancements and evolving AI applications. Develop processes and frameworks that can accommodate new AI developments while ensuring that they adhere to ethical, legal, and societal considerations.

9.2. Establish mechanisms for ongoing dialogue and engagement with industry stakeholders, AI experts, and academia to stay informed about emerging technologies, challenges, and opportunities. Encourage open collaboration to address regulatory gaps and enable innovation within the boundaries of responsible AI.

9.3. Develop mechanisms to regularly review and update AI regulations to adapt to emerging challenges, address new risks, and incorporate evolving international best practices.

9.4. Foster collaboration between regulatory bodies and technology experts to maintain an up-to-date understanding of AI developments and trends. This enables regulators to make informed decisions and respond effectively to the evolving AI landscape.

10. REGULATORY SANDBOX AND PILOT PROJECTS:

10.1. Establish regulatory sandboxes or pilot projects that allow for controlled experimentation and testing of AI technologies. This provides a platform for developers and deployers to explore innovative AI solutions while ensuring compliance with regulations and minimising potential risks.

10.2. Regularly evaluate the outcomes and lessons learned from these sandboxes and pilot projects to inform the refinement of regulatory frameworks and promote responsible AI practices.

10.3. Regulatory sandboxes provide a space for collaboration between regulators, industry stakeholders, and researchers to understand the implications of AI technologies and iterate on regulatory approaches in a controlled and iterative manner.

11. CYBERSECURITY AND RESILIENCE:

11.1. Integrate cybersecurity measures into the design and implementation of AI systems to protect against potential threats, data breaches, and malicious attacks.

11.2. Foster collaborations between AI developers, deployers, and cybersecurity experts to enhance the security and resilience of AI systems, including regular vulnerability assessments and proactive risk mitigation strategies.

11.3. Strengthen cybersecurity measures to protect AI systems from potential cyber threats and attacks. This includes robust encryption, secure data storage, and regular vulnerability assessments.

11.4. Establish incident response plans and protocols to ensure a swift and effective response in the event of a cybersecurity breach or compromise of AI systems.

12. INTERNATIONAL COLLABORATION AND HARMONISATION:

12.1. Actively participate in international forums and collaborations to harmonise AI regulations, standards, and best practices. This includes engaging with organisations like the United Nations, OECD, and other international bodies to shape global norms and promote responsible AI governance.

12.2. Seek opportunities for bilateral or multilateral agreements with other countries to facilitate information exchange, regulatory cooperation, and mutual recognition of AI systems' safety and ethical standards.

13. CROSS-SECTOR COLLABORATION AND STAKEHOLDER ENGAGEMENT:

13.1. Foster collaboration among government agencies, industry experts, academia, and civil society organisations to ensure a holistic approach to AI regulation. This collaboration can lead to the development of informed policies that balance innovation and ethical considerations.

13.2. Establish mechanisms for ongoing dialogue and consultation with stakeholders to gather diverse perspectives and address emerging challenges and concerns. Regular engagement with stakeholders, including public consultations, can help identify potential risks, improve understanding, and build public trust.

13.3. Engage in international partnerships and collaborations to align with global standards and promote knowledge sharing. By participating in international discussions and collaborations, Australia can contribute to the development of responsible AI practices while staying informed about global developments and potential risks.

14. SUPPORT FOR SAFE AND RESPONSIBLE AI PRACTICES:

14.1. Provide resources and support for organisations to implement responsible AI practices, including employee training, capacity building, and access to expert guidance. Investing in education and skill development can help organisations adopt ethical AI principles and ensure responsible deployment.

14.2. Encourage the adoption of ethical frameworks and guidelines, such as those developed by the Partnership on AI, to promote responsible AI development and deployment. These frameworks provide practical guidance on various aspects of responsible AI, including fairness, transparency, and accountability.

14.3. Foster a culture of continuous learning and improvement, encouraging organisations to regularly monitor and document AI systems' performance and address any unintended biases or negative impacts. Learning from real-world deployments can lead to iterative improvements and enhance the overall responsibility of AI systems.

15. ETHICAL AI DESIGN AND LABELLING:

15.1. Develop and promote clear ethical AI design principles that guide the development and deployment of AI systems. These principles should encompass values such as fairness, transparency, accountability, privacy, and human-centeredness.

15.2. Encourage developers and deployers to integrate ethical considerations into the design and development process of AI systems. This can be achieved through awareness campaigns, training programs, and providing resources that emphasise the importance of ethical AI practices.

15.3. Introduce a system for labelling AI systems to provide transparency and inform users about the underlying algorithms, data sources, and potential limitations or biases of the AI system.

15.4. The labelling system can follow standardised formats and include relevant information such as the purpose of the AI system, data collection practices, and details on the human oversight or intervention involved in decision-making.

16. PUBLIC AWARENESS AND EDUCATION:

16.1. Launch public education campaigns to increase awareness and understanding of AI technologies, their potential benefits, and the associated risks and challenges. This can help empower individuals to make informed decisions and engage in responsible AI use.

16.2. Conduct public awareness campaigns to educate citizens about AI technologies, their capabilities, and potential risks. This includes fostering a better understanding of AI's limitations, biases, and ethical implications.

16.3. Promote AI literacy and education across different sectors, including schools, universities, and workplaces. This can involve integrating AI education into curriculum, offering training programs, and encouraging lifelong learning to equip individuals with the knowledge and skills to engage with AI technologies responsibly.

16.4. Prioritise AI education and literacy programs to enhance public understanding of AI technologies, their capabilities, and potential implications. This includes raising awareness about AI ethics, biases, and the responsible use of AI systems.

16.5. Develop educational initiatives that target different age groups and sectors, including schools, universities, businesses, and government agencies. Promote AI literacy to empower individuals to make informed decisions about AI adoption and usage.

16.6. Promote AI education in schools and universities, ensuring that students are equipped with the necessary knowledge and skills to understand and critically evaluate AI technologies.

17. INCLUSIVE AND DIVERSE AI DEVELOPMENT AND PARTICIPATION:

17.1. Foster diversity and inclusivity in AI development teams to ensure a wider range of perspectives and mitigate biases. Encouraging diverse teams can lead to more comprehensive AI systems that consider a broader range of user needs and societal impacts.

17.2. Invest in initiatives that support underrepresented groups in AI, such as women and minority communities, by providing scholarships, mentorship programs, and research grants. This can help address the diversity gap and promote fairness and equality in AI development.

17.3. Foster public participation and engagement in AI governance processes. Establish mechanisms for soliciting public input on AI policy development, regulation, and deployment, such as public consultations, citizen assemblies, or advisory boards.

17.4. Promote inclusivity and diversity in AI decision-making processes by involving representatives from various communities, including marginalised groups, in discussions and decision-making related to AI development and deployment.

18. EQUITY AND EQUITY ACCESSIBILITY:

18.1. Ensure that AI systems do not perpetuate or exacerbate existing social inequalities. Developers and deployers should proactively address issues of bias and discrimination, considering the potential impact of AI systems on marginalised groups and vulnerable populations.

18.2. Promote accessibility by designing AI systems that are inclusive and usable for individuals with disabilities. This includes considering diverse user needs and providing appropriate accommodations and interfaces.

18.3. Address the potential biases and disparities in access to AI technologies. Ensure that AI systems are developed and deployed in a manner that promotes equitable access and benefits for all segments of society.

18.4. Monitor and mitigate any biases or discriminatory impacts that may arise from AI algorithms and models, particularly in areas such as healthcare, employment, and financial services.

19. PUBLIC PARTICIPATION, ACCOUNTABILITY, AND GOVERNANCE:

19.1. Promote public participation and involvement in the decision-making processes related to AI development and deployment. This can be achieved through public consultations, citizen panels, or participatory design approaches.

19.2. Establish mechanisms for individuals to report concerns or issues related to AI systems, ensuring that their feedback is taken into account and appropriate actions are taken to address them.

19.3. Promote participatory governance models for AI decision-making processes. Involve citizens, end-users, and stakeholders in the development of AI policies, regulations, and guidelines to ensure diverse perspectives are considered and democratic values are upheld.

20. CONTINUOUS MONITORING AND ADAPTION:

20.1. Implement mechanisms for ongoing monitoring and evaluation of AI systems in real-world contexts. This includes collecting feedback from users, monitoring performance, and assessing the system's impact on individuals and society.

20.2. Regularly update and adapt AI regulations and governance frameworks based on emerging risks, technological advancements, and societal changes. This ensures that the regulatory framework remains responsive and effective in addressing evolving challenges.

20.3. Encourage research and development in AI auditing and monitoring tools that can detect and mitigate biases, ensure fairness, and assess the overall performance and impact of AI systems.

21. ETHICAL USE OF AI IN GOVERNMENT:

21.1. Introduce mandatory impact assessments for AI procurement processes in government agencies and organisations to evaluate the potential risks, biases, and social implications of AI systems before their deployment. These impact assessments should consider factors such as privacy, fairness, security, and the potential impact on vulnerable populations, ensuring responsible and accountable AI adoption.

21.2. Establish guidelines and frameworks for the ethical use of AI in government decision-making processes. This includes ensuring transparency, fairness, and accountability when AI systems are used to support policy development, resource allocation, or public service delivery.

21.3. Implement safeguards to prevent the misuse of AI technologies for surveillance, invasion of privacy, or discriminatory practices within government agencies.

21.4. Develop guidelines and frameworks for the ethical use of AI within government departments and agencies. This includes ensuring transparency, fairness, and accountability in the

21.5. Establish mechanisms for independent oversight and audit of AI systems used within the government to ensure compliance with ethical standards and protect against potential biases or discriminatory practices.

21.6. Provide training and capacity building for government officials to understand the ethical considerations of AI and make informed decisions regarding its implementation. Foster a culture of responsible AI use within the government.

22. RESPONSIBLE AI PROCUREMENT:

22.1. Develop guidelines for responsible AI procurement by government agencies. This includes considering ethical considerations, transparency requirements, and accountability mechanisms when acquiring AI systems or services.

22.2. Encourage government agencies to prioritise the procurement of AI systems from vendors who adhere to responsible AI practices and demonstrate a commitment to fairness, transparency, and ethical considerations.

22.3. Encourage private organisations and businesses to adopt responsible AI procurement practices by incentivising the use of AI systems that meet certain ethical standards and compliance requirements.

22.4. Integrate responsible AI considerations into government procurement processes. Prioritise the selection of AI systems that align with ethical standards, transparency, and accountability requirements. Encourage private organisations to adopt similar responsible AI procurement practices.

22.5. Include specific clauses in procurement contracts that require vendors to adhere to responsible AI practices, provide transparency, and undergo independent audits to ensure compliance. Regularly evaluate vendors' performance and ethical practices to maintain high standards.

22.6. Incorporate responsible AI criteria into government procurement processes. Require vendors to adhere to ethical AI principles and demonstrate transparency, fairness, and accountability in their AI systems.

22.7. Develop guidelines and evaluation frameworks for assessing the ethical and responsible use of AI in government contracts. Consider factors such as privacy protection, algorithmic bias mitigation, and accountability mechanisms when selecting AI vendors.

23. ETHICAL AI RESEARCH AND DEVELOPMENT:

23.1. Promote and support research and development initiatives that prioritise ethical considerations in AI. Allocate funding and resources to encourage the development of AI technologies that are aligned with ethical principles, fairness, transparency, and social good.

23.2. Foster interdisciplinary collaborations between AI researchers, ethicists, social scientists, and other relevant disciplines to ensure that AI technologies are developed with a comprehensive understanding of their potential impact on individuals, communities, and society as a whole.

23.3. Allocate funding and resources for research and development in ethical AI, including areas such as bias mitigation, fairness, explainability, and accountability.

23.4. Support collaborations between academia, industry, and government to advance ethical AI research and promote the adoption of responsible AI practices.

23.5. Introduce incentives and recognition programs to encourage organisations to prioritise responsible AI practices. Offer tax incentives, grants, or other forms of support to businesses that demonstrate commitment to ethical AI development, deployment, and transparency.

23.6. Recognise and showcase organisations that excel in responsible AI practices, encouraging others to follow suit. Establish awards and certifications that validate and promote ethical AI approaches and highlight organisations that prioritise societal well-being.

24. EVALUATION AND CERTIFICATION:

24.1. Develop an evaluation and certification process for AI systems to assess their compliance with ethical standards, fairness, and transparency. This can help users and organisations make informed choices when selecting AI technologies.

24.2. Establish independent certification bodies or regulatory agencies responsible for evaluating and certifying AI systems based on predefined criteria and standards.

24.3. Develop mechanisms for the evaluation and certification of AI systems to ensure compliance with ethical and safety standards. Establish independent certification bodies or accreditation programs that assess AI systems based on predefined criteria.

24.4. Certification processes should consider factors such as fairness, transparency, privacy, security, and robustness of AI systems. This can provide assurance to users and stakeholders that AI technologies have undergone rigorous evaluation and meet recognised standards.

25. THIRD-PARTY CERTIFICATION:

25.1. Encourage the development of third-party certification programs for AI systems to verify their compliance with ethical and responsible standards. Independent organisations can assess AI systems based on criteria such as fairness, transparency, accountability, and privacy protection, providing a trusted certification mark.

25.2. Promote the adoption of certified AI systems by offering incentives or preferences in government procurement processes. This encourages organisations to prioritise the use of AI systems that have been rigorously assessed and certified for their ethical and responsible practices.

26. ETHICAL AI IMPACT ASSESSMENTS:

26.1. Mandate the integration of ethical impact assessments as part of the AI development and deployment process. Require organisations to evaluate the potential social, economic, and environmental impacts of AI systems and address any ethical concerns.

26.2. Establish clear guidelines and frameworks for conducting ethical impact assessments, ensuring consistency and thorough evaluation of AI systems' implications on various stakeholders and society as a whole. Consider the long-term consequences and unintended effects of AI technologies.

27. SOCIAL IMPACT ASSESSMENTS:

27.1. Incorporate social impact assessments as a mandatory requirement for high-risk AI systems and applications. Assess the potential societal implications of AI deployment, including economic, environmental, and social consequences.

27.2. Engage with diverse stakeholders, including community representatives, advocacy groups, and marginalised communities, in the social impact assessment process. Ensure that the perspectives and concerns of all stakeholders are considered in decision-making processes.

28. PUBLIC PARTICIPATION AND ETHICAL IMPACT ASSESSMENTS:

28.1. Implement mechanisms for public participation in the development, deployment, and regulation of AI systems. This can involve soliciting public input, engaging citizens in decision-making processes, and conducting public consultations on AI-related policies and regulations.

28.2. Conduct comprehensive ethical impact assessments for high-risk AI applications, involving multidisciplinary experts and public representatives. These assessments can evaluate the potential social, economic, and ethical implications of AI systems and inform regulatory decisions.

29. AI LABORATORY NETWORK:

29.1. Establish a network of AI laboratories across universities and research institutions to promote research, development, and testing of AI technologies. Encourage collaboration between academia, industry, and government to drive innovation while ensuring responsible practices.

29.2. Allocate resources for these AI laboratories to conduct independent evaluations of AI systems, assess algorithmic biases, and propose improvements to enhance fairness, transparency, and accountability.

29.3. Establish partnerships with educational institutions, industry organisations, and research centers to offer training programs, certifications, and workshops on AI ethics, responsible AI development, and AI governance.

30. ETHICAL AI CHARTERS AND CODES OF CONDUCT:

30.1. Encourage the development and adoption of industry-specific ethical AI charters and codes of conduct. These guidelines can outline best practices and principles for the responsible use of AI within specific sectors, such as healthcare, finance, and transportation.

30.2. Collaborate with industry stakeholders, professional associations, and experts to establish ethical AI standards and encourage organisations to voluntarily adopt and adhere to these principles.

The recommendations put forth in this submission aim to establish a comprehensive regulatory framework and governance system for safe and responsible AI development and deployment in Australia. These recommendations emphasise the need for risk assessment, transparency, accountability, auditability, and ethical considerations in AI development and implementation. By fostering collaboration, supporting innovation, and ensuring privacy and cybersecurity measures, Australia can navigate the evolving AI landscape while prioritising the protection of human rights and societal well-being. Implementation of these recommendations will contribute to the responsible and beneficial use of AI technology in Australia and serve as a model for other nations facing similar challenges.

ABOUT THE AUTHOR

Rocky Scopelliti is a world-renowned futurologist whose purpose is to *'create confidence in our future'*. His pioneering behavioural economics research on the confluence of demographic change and digital technology, have influenced the way we think about our social, cultural, economic, and technological future.

His new book *'Disruptive Decarbonisation - Australia's technological race to net zero carbon emissions'* is the first major study of its type into how Australia's industries, organisations and leaders can increase their capacity to adapt to a world in accelerated change. His other books include *'Australia 2030! - Where the bloody hell are we?'* a major study of Australian professional's attitudes toward the decade and '*Youthquake 4.0 - A Whole Generation and the New Industrial Revolution'* which has been published in Chinese, Vietnamese, Indonesian and Korean languages.

As a media commentator, his unique insights have featured on SKY Business News, The Australian Financial Review, ABC Radio National, The Economist, Forbes, and Bloomberg. As an international keynote speaker, his presentations have captivated audiences across Asia Pacific, the USA and Europe including Mobile World Congress. As a thought-leader, over 150 boards and leadership teams, including Fortune 100 corporations, each year seek his advice on strategy, emerging technologies, and their future impact.

A distinguished author, his 18-research thought leadership research publications have become internationally recognised for their influence including the World Economic Forum.

In an executive capacity, he is a member of the Optus Enterprise & Business Leadership team as a Chief Scientist for Government where he leads the creation of world class thought leadership and innovation to guide emerging technologies of the Fourth Industrial Revolution.

In a non-executive capacity, he is an Adjunct Professor (Industry) at the University of Technology Business School, a director on the board of Community First Bank, on the technology advisory board for REST Super and Wake by Reach and a former board member of the Australian Payments Council.

Educated in Australia and trained in the USA at Sydney and Stanford Universities respectively, he has a Graduate Diploma in Corporate Management and an MBA. He is also a graduate and member of the Australian Institute of Company Directors.

REFERENCES

[i] International Organization for Standardization (ISO). (n.d.). Artificial intelligence -- Vocabulary. ISO/IEC 2382-37:2020.
Australian Government Department of Industry, Science, Energy, and Resources. (n.d.). Safe and responsible AI in Australia. Retrieved from consult.industry.gov.au/supporting-responsible-ai
International Organization for Standardization (ISO)Website: https://www.iso.org/home.html
[ii] Online Safety Act 2021, Australian Government, eSafety Commissioner.
Therapeutic Goods Administration (TGA), Australian Government.
Australian Privacy Act Review, Office of the Australian Information Commissioner.
European Union General Data Protection Regulation (GDPR).
Directive on Automated Decision-Making, Government of Canada.
[iii]Australian Consumer Law, Australian Competition and Consumer Commission.
AI and Discrimination in Hiring: A Case Study, Australian Human Rights Commission.
Algorithmic Impact Assessments: A Study of Discrimination in AI Hiring, Data61.
Clearview AI and Privacy Law Breach, Office of the Australian Information Commissioner.
Product Safety Australia, Australian Competition and Consumer Commission.
[iv] Trivago found guilty of misleading conduct over hotel pricing claims, Australian Competition and Consumer Commission.
Australian Consumer Law, Australian Competition and Consumer Commission.
Online Safety Act 2021, Australian Communications and Media Authority.
[v] AI Ethics Framework, Australian Government Department of Industry, Science, Energy and Resources.
National AI Centre and Responsible AI Network (RAIN),
[vi] Directive on Automated Decision-Making, Treasury Board of Canada Secretariat.
Model AI Governance Framework, Infocomm Media Development Authority, Singapore.
General Data Protection Regulation (GDPR), European Union.
National Artificial Intelligence Research and Development Strategic Plan, National Science and Technology Council, USA.
Centre for Data Ethics and Innovation, UK Government.
How will China's Generative AI Regulations Shape the Future? A .... https://digichina.stanford.edu/work/how-will-chinas-generative-ai-regulations-shape-the-future-a-digichina-forum/.
O'Shaughnessy, M. (2022). The Global AI Regulator's Dilemma. Carnegie Endowment for International Peace.
Sheehan, M. (2021). What China's Algorithm Registry Reveals about AI Governance. Carnegie Endowment for International Peace.
[vii] AI Black Box, Republic of Estonia.
"Facial Recognition: Controversies, Technical Challenges, and Ethical Dilemmas" - Boulgouris et al. (2020)
National AI Strategy, Ministry of Science and ICT, South Korea; Partnership on AI.
"Responsible AI Practices" – Google.
"The Need for Ethics in AI Research and Innovation: Challenges and Strategies" - Calvo et al. (2021).
"Towards Principles of Ethical AI: A Japanese Perspective" - Hirano (2020).
[viii] Ethics Guidelines for Trustworthy AI, European Commission.
OECD Principles on Artificial Intelligence.
Responsible AI Certification Program, State Services Commission, New Zealand.
United States Federal AI Community of Practice.
Data Ethics Advisory Group, Cabinet Office, United Kingdom.
"Building the Right Foundations for AI: A Roadmap for Governments" - World Economic Forum.
Data61, CSIRO.
[ix] Ethics Guidelines for Trustworthy AI, European Commission.
IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.
Barocas, S., Hardt, M., & Narayanan, A. (2019); Fairness and machine learning. Big data, 7(11), 1-3.
Doshi-Velez, F., & Kim, B. (2017); Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
Waymo. (2022); Waymo Safety Report.
Privacy-Preserving Machine Learning in Healthcare: Case Studies and New Directions – Springer.
IBM Research. (2020). Adversarial Machine Learning Defense: A Survey.
[x] General Data Protection Regulation (GDPR).
OpenAI. (n.d.); GPT-3 Language Model Documentation
Partnership on AI. (n.d.). Home.
U.S. Congress. (2021). Algorithmic Accountability Act of 2021.
Government Technology Agency of Singapore. (2020). Trusted AI: Enabling Ethical and Responsible AI in Singapore.
[xi] European Commission. (2021). Proposal for a Regulation Laying Down Harmonized Rules on Artificial Intelligence.
European Parliament. (2021). Regulation on Artificial Intelligence in the European Union.
UNESCO. (2020). Recommendation on the Ethics of Artificial Intelligence.
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.
O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group.
[xii] Australian Government. (2020). Artificial Intelligence: Australia's Ethics Framework.
Productivity Commission. (2019). Inquiry Report on Data Availability and Use.
Smith, M., & Anderson, E. (2019). Public Views on Trust in Automated Decision-Making. Report for the Ada Lovelace Institute.
Taylor, L. (2017). Data Justice and COVID-19: Global Perspectives. Data & Society Research Institute.
[xiii] Lomas, N. (2019). San Francisco passes city government ban on facial recognition tech. TechCrunch.
O'Grady, A. (2021). Australia's tech sector at risk from world-first AI laws. ZDNet.
Richards, N. M., & King, J. H. (2020). Three Paradoxes of Predictive Policing. Duke Law Journal, 70(6), 1227-1286.
World Trade Organization. (2020). Trade and Artificial Intelligence.
[xiv] European Commission. (2021). Proposal for a Regulation on a European approach for Artificial Intelligence.
International Electrotechnical Commission. (2021). IEC standards for Artificial Intelligence.
National Institute of Standards and Technology. (2018). Recommendation for Developing and Deploying AI Technologies. NIST Special Publication 1640-2.
Organisation for Economic Co-operation and Development. (2019). OECD Principles on Artificial Intelligence.

[xv] Food and Drug Administration (FDA). (2021). Regulatory Framework for Digital Health Technologies and AI.
Information Commissioner's Office (ICO). (2020). Explaining Decisions Made with AI.
National Health Service (NHS). (2021). AI in Health and Care Award - High Level Guidance.
Royal Society and British Academy. (2019). Data management and use: Governance in the 21st century.

[xvi] European Commission. (2018). General Data Protection Regulation (GDPR).
Information Commissioner's Office (ICO). (2020). Explaining Decisions Made with AI.
National Institute of Standards and Technology (NIST). (2021). Framework for Managing Privacy Risks in AI.

[xvii] Bostrom, N., & Yudkowsky, E. (2014). The Ethics of Artificial Intelligence. Cambridge University Press.
European Union Agency for Fundamental Rights (FRA). (2020). Fundamental Rights Implications of Artificial Intelligence.
OECD. (2019). AI in Society.
World Economic Forum (WEF). (2020). Towards Trustworthy AI: Rethinking AI Governance to Reduce Risks and Increase Benefits.

[xviii] European Union General Data Protection Regulation (GDPR). Information Commissioner's Office. (2020). Explaining decisions made with artificial intelligence.
Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.
Wallach, W., & Allen, C. (2010). Moral machines: teaching robots right from wrong. Oxford University Press.

[xix] International Organization for Standardization. (2018). ISO 31000:2018 Risk management—Guidelines.
European Union General Data Protection Regulation (GDPR). Information Commissioner's Office. (2018). Artificial Intelligence, automation, and data protection.
Data Protection Commission. (2020). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679.

[xx] OpenAI. (2021). "GPT-3: Language Models are Few-Shot Learners."
Mozilla. (n.d.). "Common Voice: A Publicly Available Voice Dataset."
Jobin, A., et al. (2019). "AI in the Wild: The Risks of Bias and Errors in Natural Language Processing." arXiv preprint arXiv:1908.10838.
Gebru, T., et al. (2020). "Datasheets for Datasets." Conference on Fairness, Accountability, and Transparency.

[xxi] Partnership on AI. (n.d.). "About the Partnership on AI." European Union. (2016).
"Regulation (EU) 2016/679 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)."