Blackwell AI

31st July 2023

Department of Industry, Science and Resources

Safe and responsible AI in Australia

Dear Reader,

We welcome the opportunity to respond to the discussion paper on supporting responsible Artificial Intelligence (AI). As a start-up at the forefront of AI technology, we understand the critical importance of addressing the potential risks associated with AI while fostering a safe and responsible AI environment in Australia.

AI has undoubtedly brought significant advancements and improvements to various aspects of our lives. However, the rapid pace of innovation in this domain has also raised concerns and uncertainties about its potential consequences. We believe that effective governance mechanisms are essential to ensure the development and deployment of AI in a manner that upholds privacy and security.

In light of the Government's focus on regulations, standards, tools, frameworks, principles, and business practices, we commend their commitment to fostering safe and responsible AI practices in the country. As a stakeholder in this AI landscape, we are eager to share our perspectives and contribute to the collective effort to mitigate risks and build a secure and thriving AI ecosystem.

We believe we are in a unique position to respond to this discussion paper as we are an AI start-up with an alternate view on AI systems and technology, especially surrounding data and privacy risks in a range of business environments.

Once again, we welcome the consultation and the opportunity to respond on this crucial matter. We look forward to collaborating with all relevant stakeholders to promote the responsible use of AI in Australia and to collectively address the challenges and opportunities that AI brings to our nation.

Regards,

Blackwell AI Team

Blackwell AI blackwellai@outlook.com **ABN 33 665 427 976**

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

We would appreciate the opportunity to share our insights on refining and clarifying these definitions to ensure that they are precise, comprehensive, and accessible to a wide range of stakeholders. These clarifications are essential because they help set boundaries on the subject, preventing confusion with similar software that employs similar terminology but relies on different methodologies.

The definition of artificial intelligence states that it refers to an "engineered system" that generates predictive outputs without explicit programming. While this is generally accurate, the term "engineered system" might be too broad and could include non-AI systems as well. A more precise and explicit definition of AI, emphasising its ability to learn from data and adapt to new situations, could be beneficial.

Specificity in Machine Learning: The definition of machine learning correctly highlights that it involves patterns derived from training data for prediction or decision-making. However, it would be helpful to specify that machine learning is a subset of AI, focusing on algorithms that improve their performance through experience and data. A distinction should also be made between machine learning and statistical learning as the current definition could be interpreted to include all predictive models.

Clarity on the Range of Automated Systems: The definition of automated systems seems to encompass a broad spectrum of technologies, from traditional non-technological rules-based systems to more advanced AI-based tools. Providing clearer distinctions between these different types of systems and their respective applications could enhance understanding.

We would emphasise the need for comprehensive and nuanced definitions of AI, machine learning, and automated systems, along with a focus on the responsible development and deployment of these advanced AI models. It is vital for the Australian Government's policies and frameworks to account for the evolving nature of AI technologies and their potential impact on various sectors, while maintaining a strong commitment to privacy, security, and ethical AI practices.

Blackwell AI

What potential risks from AI are not covered by Australia's existing regulatory approaches? Do you have suggestions for possible regulatory action to mitigate these risks?

We acknowledge the importance of regulatory approaches to address potential risks associated with AI in Australia. While existing regulations provide a foundation for safeguarding privacy and data protection, it is essential to recognise that some risks related to AI may not be fully covered by the current regulatory landscape.

While the Privacy Act addresses many privacy concerns, AI's increasing influence requires a closer examination of existing protections. We recommend updating the Privacy Act to incorporate specific provisions that address data privacy concerns related to AI technologies, especially those involving large-scale data processing and profiling.

Another potential risk not extensively addressed by existing regulations is the ethical use of AI, particularly in sensitive domains such as healthcare and financial services. As AI technology evolves rapidly, it is crucial to develop regulatory frameworks that guide the responsible and ethical development, deployment, and decision-making processes surrounding AI systems. Most importantly, how data can be collected and used needs to be clearly defined, as most AI tools actively rely on user input data to further train and develop their models.

Recognising the significance of regulatory measures in addressing potential AI risks, we emphasise the value of frameworks in safeguarding privacy and data protection. However, it must also find a harmonious equilibrium that promotes innovation, experimentation, and pilot programs. Allowing AI developers the room to address potential challenges and issues will cultivate a vibrant AI environment while providing protection against severe risks.

Successful regulatory action should be based on a collaborative approach involving industry stakeholders, government agencies, and experts, while simultaneously allowing developers to innovate and experiment. By engaging in constructive dialogue and considering diverse perspectives, we can design effective regulatory frameworks that strike a balance between managing risks and fostering innovation in the AI landscape.

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.

The Australian Government can further support responsible AI practices by making additional investment in AI research. By allocating resources and funding to establish specialised AI programs, research centres, and scholarships, the government can cultivate a new generation of AI professionals with the necessary skills and knowledge to develop and deploy AI systems responsibly. This initiative will not only address the growing demand for AI expertise in various industries but also foster a culture of ethical AI development, privacy protection, and security consciousness. By encouraging research and collaboration within academia, the government can promote best practices and ensure that Australia remains at the forefront of AI innovation while adhering to principles of privacy and security.

For enhancing responsible AI practices, the Australian Government should actively participate in global coordination efforts concerning data validity and training of Language Models (LLMs) and Multimodal Models (MfMs) used in chatbots and other public tools. Collaboration with international partners, industry stakeholders, and experts can establish common standards for data collection, validation, and model training, reducing potential biases and ensuring the development of fair and trustworthy AI systems. Engaging in global initiatives will also enable Australia to learn from other countries' experiences, share best practices, and collectively address challenges related to AI's ethical use. This approach fosters transparency, accountability, and fairness in AI applications, while simultaneously positioning Australia at the forefront of leading a global example in responsible and safe use AI.

Are there any governance measures being taken or considered by other countries (including any not discussed in this paper) that are relevant, adaptable and desirable for Australia?

The governance measures taken by Italy regarding the banning and subsequent unbanning of ChatGPT hold significant relevance for Australia in the context of data collection and age verification, especially when it comes to publicly accessible AI platforms and the potential risks of harm and damage. Italy's approach in addressing concerns about private data used to train ChatGPT demonstrates the importance of robust data privacy conditions and transparency in AI systems. By requiring increased transparency on data processing, opt-out rights, age verification, and notices about potential inaccuracies, Italy's regulatory actions exemplify measures that prioritise user safety and privacy. Such measures could serve as valuable lessons for Australia in formulating its own governance mechanisms to ensure responsible AI practices and foster public trust in AI technologies.

Blackwell AI

Given the importance of transparency across the AI lifecycle, please share your thoughts on where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI and mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.

Transparency is most critical and valuable in the AI lifecycle during the model pre-training and model fine-tuning stages. In the pre-training phase, transparency becomes essential to ensure data quality and prevent potential biases from being ingrained into the AI model. Understanding the sources and types of data used for pre-training can help identify and address any ethical or representational concerns. Similarly, during the fine-tuning stage, transparency is crucial to shed light on the intentions and objectives of the model's optimisation. This transparency enables an assessment of potential risks and biases introduced through the fine-tuning process, helping to improve the model's fairness and accountability. By focusing on transparency at these early stages, AI developers can instil public trust and confidence in the technology by showcasing responsible and ethical practices.

Mandating transparency requirements across the private and public sectors poses challenges due to the inherent "blackbox" nature of AI technology and the sensitivity of sharing data sets and AI models. Complete transparency may not always be feasible or appropriate, as it could inadvertently reveal sensitive business practices or information. Instead, a balanced approach could be implemented through a combination of regulatory guidelines and ethical frameworks. During the pre-training and fine-tuning stages, a light mandate for transparency can be encouraged, focusing on providing explanations of the general data sources and objectives without divulging proprietary details. Additionally, creating industry standards and best practices for transparency reporting can enable organisations to voluntarily disclose relevant information without compromising competitive advantages.

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

A risk-based approach for addressing potential AI risks is generally favourable as it allows for a nuanced and adaptable strategy that considers the context-specific risks associated with different AI deployments. By categorising AI uses based on their risk levels, organisations can implement less onerous regulation for lower risk AI applications, enabling innovation and development in these areas without unnecessary burdens.

However, it is essential to focus on balance and avoid an overly heavy enforcement of risk-based requirements. Excessive regulatory burdens on medium and high-risk AI applications may hinder innovation and hamper the development of AI technologies in critical areas, particular for small and medium sized businesses. Over-regulation of AI applications could discourage investment and research in these fields due to increased costs and complexity. Therefore, it is vital to foster a collaborative environment between regulators, industry stakeholders, and researchers to continually evaluate and refine risk-based frameworks. This dynamic approach can ensure that AI remains safe and accountable while fostering a conducive environment for innovation and technological advancement. Moreover, promoting transparency and accountability in AI development can complement the risk-based approach, further enhancing public trust and confidence in the technology.

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

A risk-based approach to general purpose AI systems, such as LLMs or MfMs, would initially categorise them as low-risk due to their general applicability. However, their risk level can escalate quickly based on specific use cases and deployment contexts. When applied in settings involving sensitive data, decision-making, or critical infrastructure, these AI systems could transition into medium to high-risk categories. Therefore, it is essential to assess the potential impacts and context of their applications carefully to determine appropriate risk management requirements.