NCC Group's response to the Government's discussion paper: Safe and responsible AI in Australia

July 2023

NCC Group welcomes the opportunity to respond to the Australian Government's discussion paper and offer our expertise as a global cyber security business. We are keen to ensure that security considerations are embedded in AI systems from the outset. This is because AI can only be safe if it is also secure.

We support the Government's endeavours to review its approach to governing AI, building public trust and enabling society to reap the benefits AI has to offer. We agree that the myriad of existing technology-neutral laws governing areas such as data privacy, online safety and consumer rights provide a strong foundation for regulating AI. We also agree that there is a need for context-specific regulation, and support ongoing regulatory efforts in sectors such as healthcare and transport. That said, AI presents new challenges and risks that need to be understood and mitigated so that Australia can benefit from the opportunities AI presents. We believe that the **Government's plans could be strengthened in the following ways** to cement Australia's position as a global leader in this field:

- **The Government must clearly define Australia's risk appetite.** AI will never be zero risk if we wish to pioneer its development and, crucially, its deployment. As such, Australia should define its risk appetite so that red lines with regards to AI and ML systems and their security, safety and resilience are known.
- **End-users and consumers should be empowered** to make decisions about the AI systems they use by improving transparency of where and how AI technologies are being deployed. A Government-led **consumer labelling scheme**, backed up by **independent third-party product validation**, would enable end-users to more confidently use AI technologies, knowing steps have been taken to reduce associated risks. We have seen similar schemes deployed for smart devices, which could be built on and learnt from. For higher-risk products, we believe that continuous third-party product assurance should be mandated. This should take into account the context in which AI is deployed, as well as the underlying risks in the technology itself. There should also be clear routes for redress where things go wrong, building on existing legislation.
- **Flexibility, agility and periodic regulatory and legislative reviews** should be built in from the outset to keep pace with technological and societal developments. This could include requirements for regulators to engage regularly with industry experts, drawing from a wide range of backgrounds and generational perspectives.
- In assuming a greater role in regulating the use of AI, **regulators should be strengthened in their powers, resources and capabilities**.
- There remains a significant shortage of the skills we need to develop AI frameworks, and assure systems' safety, security and privacy. If Australia wants to be a global leader in AI, the Government must focus investment on **developing the skills needed** to make its regime a success.
- If the Government wants to ensure that Australian languages, religious outlooks, values and cultural references are protected, while also minimising the risk of adopting biases seen elsewhere in the world, steps must be taken to **make Australian datasets more readily available for use in AI**.
- **The drafting, approval and implementation of technical standards** that underpin the Australia's regulatory framework will be critical. We ask that a clear route map for the development of technical standards is laid out, detailing where those standards are cross-sectoral and where sector-specific standards are required (e.g. medical devices).

We are pleased that the Government is taking the time to review its approach. We are keen to support its development by sharing our expertise and insights from operating at the 'coalface' of cyber security. Below we explore our recommendations in more detail, responding directly to the Discussion Paper's questions.

About NCC Group

NCC Group's purpose is to create a more secure digital future**.** As experts in cyber security and risk management, our c.2,300 people worldwide are trusted by our customers to help protect their operations from cyber threats. As a global business operating in 12 countries, we were delighted to open our regional headquarters in Sydney last year amid a rapidly growing footprint across Australia, with around 85 colleagues now based here.

Each year we dedicate thousands of days of internal research and development enabling us to stay at the forefront of cyber security and ensuring we secure the rapidly evolving and complex technological environment. We have many years' experience researching artificial intelligence (AI), large language models (LLMs) and machine learning (ML) to understand the risks and opportunities these technologies present. We have published whitepapers, blogs and guidance on everything from security auditing of ML systems and the use of LLMs in security code reviews to security vulnerabilities found in AI-based systems such as facial recognition technology and image classification.

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?

Most AI-based products in use today are 'black box' appliances that are placed onto networks and configured to consume data, process it and output decisions without humans having much knowledge of what's happening. This means understanding and explaining why an AI system reached a certain decision can be very difficult and could be used maliciously by some organisations to provide cover for why a decision has been reached. To enable effective governance, contestability and redress, we believe that any new regulatory framework should ensure that organisations are transparent about:

- **The algorithms they are using and the sources of their training data:** There are two core components to AI-based applications: (1) the algorithms themselves and (2) the data they use. In the end, any application is only as good and fair as the algorithms used (i.e. the right questions must be asked) and the quality of data used to train it. Organisations should outline in general terms how the AI makes decisions, what factors it takes into account and any biases it may have. This could be standardised through a labelling scheme.
- **The AI's decision-making process:** Explainability[1] will be difficult to achieve for many AI-based systems. Indeed, there is an inevitable trade-off between the explainability of a system and its effectiveness as an AI-based product. Nevertheless, organisations should be able to provide insights into how the AI system makes decisions. This might not mean exposing the intricacies of the algorithm, but rather explaining the type of data inputs the AI uses and the broad logic it follows. For example, a bank using AI for loan approvals could explain that the AI uses information like credit score, income and employment history to make its decision.
- **The processes in place to minimise risks, including human oversight:** Even the highest quality AI system, with well-designed algorithms and good training data, will, on occasion, deliver unpredictable outcomes. We therefore believe that AI technologies should be deployed to supplement, rather than replace human decision-making. We also believe that, particularly where explainability is difficult to (proportionately) achieve and/or the output of an AI is likely to impact humans in a notable way, organisations must be clear about the processes in place to oversee the system's outputs and minimise risk. Organisations should also explain how they continue to

[1] The extent to which it is possible for relevant parties to access, interpret and understand the decision-making processes of an AI system.

monitor the AI system's performance over time and how they make updates to the AI system to ensure its continued accuracy and fairness.
- **Performance metrics:** Sharing metrics related to the performance and accuracy of the AI can be beneficial, particularly where end-users are not the developers of the system. For instance, an AI health diagnostic tool could share its accuracy rate or the percentage of diagnoses that are later confirmed by human doctors.
- **Data handling:** In alignment with existing data protection legislation, organisations should disclose what data the AI is using, how this data is being collected, processed and stored. It's also important to communicate whether data is shared with third parties and how the user's privacy is being protected.

There is a danger that Australia adopts AI-based systems, particularly large language models (LLMs), that are linguistically, legally and culturally based on non-Australian datasets. This might mean that **biases and views seen in other countries inform increasingly important AI systems that are operational in Australia**. If the Government wants to ensure that Australian languages, religious outlooks, values and cultural references are protected, while also minimising the risk of adopting biases seen elsewhere in the world, steps must be taken to make Australian datasets more readily available for use in AI. There is a wealth of data already gathered by the Australian public sector; however, much of it is either not in a usable format or is protected by copyright and cannot be used commercially without payment of individual license fees (which might be prohibitively expensive to obtain for innovators). To overcome this issue, the Government should consider the benefits of:

- An initiative to publish more copyright information under a Creative Commons ‘BY’ licence in the interests of innovation and protection of Australian culture and values; and/or,
- Legislation to make all information published under the Commonwealth’s copyright available under a Creative Commons ‘BY’ licence (with some exceptions).

As the Government’s discussion paper rightly highlights, the **potential for bias in AI is a key risk**. Removing or reducing inherent existing biases, while balancing data privacy needs and taking steps to ensure that social issues are not exacerbated, will be crucial. In our view, this is not sufficiently addressed through Australia’s current legal framework. Future regulatory intervention should encourage organisations to take the following steps:

- Establishing clear processes and mechanisms through which applications can be carefully vetted and their respective data supply chains sanitised, particularly where data originates from untrusted sources, such as the Internet and end-users.
- Where proportionate to do so, updating and retraining applications with the latest available data.
- Analysing datasets to ensure they are representative and appropriate for the jurisdiction in which they are used. This should take into account the diversity of the development team responsible for sourcing the datasets, as this may result in unconscious biases. Creating synthetic data (i.e. information that is artificially manufactured rather than generated by real-world events) that is representative could be a future solution to this, and should be a focus for R&D investment.
- A multidisciplinary approach to reviewing the decision criteria used for automated decisions - which offers a legal, policy and operational perspective in addition to a technical review – should be taken to reduce bias wherever possible.
- Responsibility for issues of bias shouldn’t end when products or systems have been released. There should be a clear reporting process that allows organisations to receive and act on information about potential biases in a system. Lessons can be learnt from the security industry where there are established protocols for disclosing vulnerabilities in a system.

In our experience from operating within the world of cyber security, industry does not always learn from the lessons of others. Indeed, despite daily publicised data breaches, many **organisations continue to make the same mistakes** that eventually result in their own data breach or cyber incident. Incidents that could have been avoided by following industry best practice and learning from the mistakes of others. To avoid this happening with AI technologies, assurance and testing, backed up by investment in R&D, will play an important role in ensuring organisations involved in the development and deployment of AI are taking the right steps and learning from the mistakes of the past. Such activities need to be undertaken on a continuous basis to ensure vulnerabilities are addressed and the latest threat landscape is understood and acted upon. In addition, where the risk profile necessitates, we believe **independent continuous product validation should be mandated**. In our experience, many claims made by AI product vendors, predominantly about products' effectiveness in detecting threats, can be unproven or lack independent verification. Vendors should also be required to promote continuous mitigation throughout the product design, development, and post-market lifecycle, avoiding a 'tick-box' approach to compliance.

The Government should consider how it might address the issue of AI alignment. AI alignment refers to the **challenge of ensuring AI systems act in accordance with human values throughout their operational life**. This is especially important as AI technologies become more autonomous and sophisticated, increasing the potential for their objectives to diverge from those of human operators or society at large. In order to facilitate this, we recommend that any regulatory intervention incorporates an explicit requirement for the disclosure of details about how an AI system has been trained for alignment with human values, as well as the measures in place to maintain this alignment as the system learns and evolves. Such a requirement could help build trust and facilitate external oversight of AI systems, while emphasising the responsibility of developers and operators to continually monitor their systems and correct any misalignment that arises. However, it should be noted that AI alignment is a complex problem, and our understanding of it continues to evolve. The practicalities of regulatory implementation should be carefully considered, and the requirements should be flexible enough to adapt as our understanding of AI alignment improves. Regular consultation with the AI research community, as well as industries using AI, would be crucial to ensuring the continued relevance and efficacy of these regulations.

If new regulation is introduced, regulators will need to consider how they will **retrospectively address the existing use of AI** technologies in their sectors. Most organisations, often unknowingly, already use some form of AI (e.g. the technology used to blur or change backgrounds on video calls). How will these uses be regulated? Is investment in education required to drive up understanding of what constitutes AI, and why steps need to be taken to mitigate the associated 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.

While we note that data protection legislation does already exist, **further legal clarity may be needed as to who owns, or has right of ownership over, the datasets that are used in AI systems**. Many systems use web crawls to collate datasets available on the web, but not all publicly-available data is meant for public consumption. For example, confidential information may be intentionally or accidentally leaked on the internet. The Government should consider how scenarios where such data is (unintentionally) fed into AI systems are to be dealt with from a data privacy and legal standpoint.

A **consumer labelling scheme, backed up by independent third-party product validation**, would enable end-users to more confidently use AI technologies, knowing steps have been taken to reduce associated risks. Indeed, in our experience, many claims made by AI product vendors, predominantly about products' effectiveness in detecting threats, can be unproven or lack independent verification. This scheme could be voluntary, replicating the cybersecurity labelling scheme for smart devices. For higher-risk products, we believe that third-party product validation should be mandated. This is in line with best practice we see in the world of cybersecurity, and will help to ensure a level playing field between those AI technology developers who are taking steps to mitigate risks and those who may not be.

The Government must **equip sectoral regulators with the resources and skills** to effectively oversee the development and use of AI in their sectors. While some regulators are more advanced in their understanding of and ability to regulate AI, there remains a significant skills gap across authorities which will only widen as technologies and applications evolve. In addition to skills investment, a requirement should be established within all regulatory frameworks to regularly and systematically engage with the AI ecosystem to understand technological developments. This includes academics, incubators and accelerators, disruptors, and other innovation centres. This engagement could take a number of forms, including: secondment models; formal and informal government consultations and discussion papers; and regular sounding board mechanisms such as advisory groups and councils.

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.

While we support a context-driven, flexible approach to regulating AI, there is a risk of **reinventing the wheel when regulating cross-sectoral AI systems**. In the same way that a significant proportion of organisations use Office 365, it's likely that some AI systems will be widely adopted across the economy. In such cases, there should be a concerted effort to pool resources and coordinate regulatory responses – otherwise there is a danger of regulatory overlap and/or conflict.

There is a danger that, under a sector-based approach, **some AI systems will not fit neatly within one regulator's remit, and as such their regulation and oversight may fall through the net**. Lessons could be learnt from the world of cyber security, where the sector may dictate what types of activities need to be undertake to improve resilience, but coordination is required across sectors to avoid fragmentation. The Government could explore establishing a central function to oversee the implementation of AI regulations. Part of the function's remit could include regular assessments of the AI landscape to ensure that there is sufficient regulatory oversight across AI applications. This responsibility may need to be established on a statutory footing to ensure effective implementation.

Responses suitable for Australia

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

We note that the UK Government is looking to establish a central authority responsible for the general oversight of AI regulation. If the Australian Government is to pursue a sector/context-based approach, we would advise that it consider creating a similar function, giving it statutory powers to ensure the effective implementation of the Government's regime. This should include ensuring regulators are working together and not duplicating efforts, and that there are no areas of AI that are falling through the gaps.

We are pleased to see the Government emphasise the importance of **international regulatory cooperation**. In aligning Australia's domestic and international approach, we recommend that the Government:

- Utilises existing successful partnerships, including the 'Five Eyes' alliance and with the European Union;
- Invests time in developing practical outcomes with other governments, that go deeper than high-level principles; and,
- Ensures that civil society and industry – who will play a central role in delivering governments' objectives - are involved in discussions from the outset.

Target areas

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

While we do not have a strong view, we would emphasise that any approach, particularly one that is risk-based, should be consistently applied across the public sector, private sector and academia.

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

One area that will require attention is skills. As outlined elsewhere in this response, there is a lack of the technical skills required to develop and use safe and secure AI. In addition to skills investment, the Government might also consider where external expertise might be needed.

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?

As highlighted above, we believe that any new regulatory framework should ensure that organisations are transparent about:

- **The algorithms they are using and the sources of their training data:** There are two core components to AI-based applications: (1) the algorithms themselves and (2) the data they use. In the end, any application is only as good and fair as the algorithms used (i.e. the right questions must be asked) and the quality of data used to train it. Organisations should outline in general terms how the AI makes decisions, what factors it takes into account and any biases it may have. This could be standardised through a labelling scheme.
- **The AI's decision-making process:** Explainability[2] will be difficult to achieve for many AI-based systems. Indeed, there is an inevitable trade-off between the explainability of a system and its effectiveness as an AI-based product. Nevertheless, organisations should be able to provide insights into how the AI system makes decisions. This might not mean exposing the intricacies of the algorithm, but rather explaining the type of data inputs the AI uses and the broad logic it follows. For example, a bank using AI for loan approvals could explain that the AI uses information like credit score, income and employment history to make its decision.
- **The processes in place to minimise risks, including human oversight:** Even the highest quality AI system, with well-designed algorithms and good training data, will, on occasion, deliver unpredictable outcomes. We therefore believe that AI technologies should be deployed to supplement, rather than replace human decision-making. We also believe that, particularly where explainability is difficult to (proportionately) achieve and/or the output of an AI is likely to impact humans in a notable way, organisations must be clear about the processes in place to oversee the system's outputs and minimise risk. Organisations should also explain how they continue to monitor the AI system's performance over time and how they make updates to the AI system to ensure its continued accuracy and fairness.
- **Performance metrics:** Sharing metrics related to the performance and accuracy of the AI can be beneficial, particularly where end-users are not the developers of the system. For instance, an AI health diagnostic tool could share its accuracy rate or the percentage of diagnoses that are later confirmed by human doctors.

[2] The extent to which it is possible for relevant parties to access, interpret and understand the decision-making processes of an AI system.

- **Data handling:** In alignment with existing data protection legislation, organisations should disclose what data the AI is using, how this data is being collected, processed and stored. It's also important to communicate whether data is shared with third parties and how the user's privacy is being protected.

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

We support mandating transparency requirements. End-users and consumers should be empowered to make decisions about the AI systems they use, and this can only be achieved by improving transparency of where and how AI technologies are being deployed.

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?

We would caution against the outright banning of AI applications or technologies, as this could risk stifling innovation in Australia, leaving it behind its global counterparts. Instead, we would propose that very high-risk applications or technologies could be tested and explored within limited testbeds, and under the close supervision of an independent watchdog that ensures the Government's AI Ethics Principles are not breached.

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

A **consumer labelling scheme, backed up by independent third-party product validation**, would enable end-users to more confidently use AI technologies, knowing steps have been taken to reduce associated risks. Indeed, in our experience, many claims made by AI product vendors, predominantly about products' effectiveness in detecting threats, can be unproven or lack independent verification. This scheme could be voluntary, replicating the cybersecurity labelling scheme for smart devices. For higher-risk products, we believe that third-party product validation should be mandated. This is in line with best practice we see in the world of cybersecurity, and will help to ensure a level playing field between those AI technology developers who are taking steps to mitigate risks and those who may not be.

Implications and infrastructure

12. How would banning high-risk activities (like social scoring or facial recognition technology in certain circumstances) impact Australia's tech sector and our trade and exports with other countries?

We would caution against the outright banning of AI applications or technologies, as this could risk stifling innovation in Australia, leaving it behind its global counterparts. Instead, we would propose that very high-risk applications or technologies could be tested and explored within limited testbeds, and under the close supervision of an independent watchdog that ensures the Government's AI Ethics Principles are not breached.

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

We agree that **assurance has a core role to play** in driving up safety, privacy and security standards. However, there remains a distinct lack of people in the AI assurance sector with the experience and/or qualifications to undertake assessments, particularly assessments of cyber security risks that are unique to AI systems. We believe that Government should consider how post-16 education can be more appropriately geared toward developing educational programmes that bridge AI and related disciplines such as cyber security.

Risk-based approaches

14. Do you support a risk-based approach for addressing potential AI risks? If not, is there a better approach? 15. What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome?

We broadly support a risk-based approach to addressing potential AI risks. However, in implementing such an approach, the Government should ensure innovation is not stifled and maintain a level of flexibility and agility to future-proof against the fast-changing technological landscape. This can be done by:

- **Clearly defining Australia's risk appetite**. AI will never be zero risk if we wish to pioneer its development and, crucially, its deployment. As such, Australia should define its risk appetite so that red lines with regards to AI systems and their security, safety and resilience are known.
- **Establishing periodic regulatory and legislative reviews from the outset** to keep pace with technological and societal developments. This could include requirements for regulators and the proposed new central Government function to engage regularly with innovation centres and industry experts, drawing from a wide range of backgrounds and generational perspectives.
- **Investing in horizon-scanning activities.** There is already a myriad of horizon scanning activity and initiatives across government, the private sector and academia. This can lead to overlap and duplication of effort, with no central coordination and collation of data. We therefore propose that the Government invest in coordinating and improving existing horizon-scanning activities, rather than reinventing the wheel.

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

Broadly speaking, we support the elements outlined in Attachment C. We would add that **independent third-party product validation should be mandated for high-risk systems**, with other systems encouraged to adopt a consumer labelling scheme.

As outlined in detail under question 2 above, **there are a number of steps that organisations can take to improve the transparency and explainability of their AI systems**. These should be embedded in a risk-based approach, with levels of transparency and explainability increasing for higher risk systems.

We **agree that training should increase proportionate to the level of risk**. This, however, should be backed up with **investment in the key skills** needed to meet the Government's aims. In particular, focused investment is required to ensure genuine Australian leadership in AI which we would define as producing core AI frameworks, as opposed to using AI frameworks developed by others. While there are many AI frameworks available for use that abstract away from the low-level minutiae/mathematics of AI, there is likely a major skills shortage of people with deep technical understanding of AI and its algorithms. There is therefore a danger that, as a nation, Australia will be using AI frameworks developed by other nations, reliant on the assurances that they provide regarding the security, safety and biases of those frameworks. We strongly believe that this is a much less desirable outcome to being in a position where Australia is the producer of the core AI frameworks (that others might then use).

20. Should a risk-based approach for responsible AI be a voluntary or self-regulation tool or be mandated through regulation?

Where the risk profile dictates, we do believe that mandatory safety, security and privacy requirements will be needed. While advice, guidance and voluntary measures have a key role to play, well-crafted regulation provides industry with clarity, establishes a level playing field amongst developers and usually comes with the resources and powers regulators need to govern effectively. In particular, high-risk systems should be subject to mandatory independent third-party product validation.

The drafting, approval and implementation of technical standards that underpin any regulatory framework will be critical. We know that Australia is already taking steps to be at the forefront of developing world-leading AI standards, and this is something we firmly support. We ask that a clear route map for the development of technical standards is laid out, detailing where those standards are cross-sectoral and where sector-specific standards are required (e.g. medical devices).

And should it apply to: a. public or private organisations or both? b. developers or deployers or both?

It should be applied consistently across the public sector, private sector and academia, with both developers and deployers required to take steps to mitigate against the potential risks. We, however, believe that further work is required to work through edge cases and determine how legal responsibility will be defined. Other jurisdictions, such as the EU, have set out firmer views on the liability of AI systems. While we do not offer a view on whether this is the right approach, we would highlight that in the absence of a defined Australian system of legal liability, other jurisdictions' frameworks will likely be adopted.