Supporting responsible AI discussion paper

SUBMISSION

AISA

AUSTRALIAN INFORMATION SECURITY ASSOCIATION

EXECUTIVE SUMMARY

AISA

The Australian Information Security Association (AISA) champions the development of a robust information security sector by building the capacity of professionals and advancing the cyber security and safety of the Australian public as well as businesses and governments in Australia.

Established in 1999 as a nationally recognised and independent not-for-profit organisation and charity, AISA has become the recognised authority and industry body for information security, cyber security, and security-related privacy matters in Australia. AISA caters to all domains of the information security industry with a particular focus on sharing expertise from the field at meetings, focus groups and networking opportunities around Australia.

AISA's vision is a world where all people, businesses and governments are educated about the risks and dangers of invasion of privacy, cyber-attack, and data theft and to enable them to take all reasonable precautions to protect themselves. AISA was created to provide leadership for the development, promotion, and improvement of our profession, and AISA's strategic plan calls for continued work in the areas of advocacy, diversity, education, and organisational excellence.

AISA submissions represent our 10,000+ strong member association, most are professionals in cyber security, information technology, and privacy, and allied professionals in legal, regulatory, financial, and prudential sector, as well as cyber and IT enthusiasts and students around Australia. AISA members are tasked with protecting and securing public and private sector organisations including national, state and local governments, ASX listed companies, large enterprises, NGO's as well as SME/SMBs across all industries, verticals and sectors.

AISA proactively works to achieve its mission along with its strategic partners. These include: the Australian Women in Security Network (AWSN); Australian Institute of Company Directors (AICD); Australian Security Industry Association Limited (ASIAL); grok academy; the Oceania Cyber Security Centre (OCSC); Risk Management Institute of Australia (RMIA); untapped; as well as international partner associations such as $(ISC)^2$, ISACA and the Association of Information Security Professionals (AISP). AISA also works closely with both federal and state / territory governments to ensure a robust and safe sector.

Definitions

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

AISA holds the view that the definition of Artificial Intelligence (AI) presented in this discussion paper is limiting in its scope. Certain instances have already demonstrated AI's ability to exhibit knowledge beyond its programmed parameters, leading to our suggestion for the exclusion of "generate predictive outputs" from the definition. AI possesses emergent properties, where it can spontaneously acquire new skills and deliver unforeseeable results. In one such example, Google engineers were surprised when an experimental AI learned a language it was never trained for. AI hallucination exemplifies one such outcome of this characteristic, where AI could make up false information or facts which aren't based on real data or events.

Furthermore, it is crucial to incorporate Deep Learning in the definition, AI extensively employs this technique. Deep Learning involves training artificial neural networks with vast datasets to discern patterns and make informed decisions.

Potential gaps in approaches

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?

Use of existing regulations in certain sectors to protect consumers can be a pragmatic approach to address AI-related concerns. Many sectors already have established regulatory bodies with expertise in specific domains. Using these existing regulations could help avoid duplication of efforts and minimise additional regulatory burden.

However, it is also important to consider that AI is a rapidly evolving technology, and some of its applications might not fit into existing frameworks. In such cases, some guidelines might be necessary to address the challenges posed by AI. The key is to strike a balance between leveraging existing regulations and creating targeted guidelines.

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.

To safeguard the interests of Australian citizens between the opportunities and the risks presented by AI, it is essential to promote general awareness. Therefore, it is advisable for the Government to collaborate with industry bodies, partners, and the general public to enhance understanding of how major tech companies are employing AI, along with its potential benefits and harms.

AISA also recommends that the Government should debate establishing an AI standards body responsible for evaluating AI-enabled technologies for their potential impact on public use. This body could potentially conduct scrutiny of these platforms and services, with a particular focus on:

- Providing users with an option to opt-out of AI usage.
- Reviewing the appropriateness of AI implementation and setting suitable guardrails within the Australian context.
- Reuse and development of foundation models and the ability to manage potential risks and biases.

Target areas

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

AISA believes that alignment between the public and private sectors regarding the use of AI technologies is of important. This alignment is critical to safeguard the interests of Australian citizens. When both sectors work together collaboratively, they can establish clear guidelines, standards, and best practices that promote responsible and ethical AI use. Collaborative co-regulation with industry as partners, led by the government is essential. The establishment of regulatory sandboxes facilitates swift testing and iteration. It is crucial to ensure that small businesses are not left out through this process either. A common criteria or baseline (e.g., Falcon versus ChatGPT and a way to compare for usage).

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.

AISA believes that for improved adoption transparency throughout the AI lifecycle is critical. This is crucial for building trust and providing users with confidence. Ethical aspects are of paramount importance for building trust. Few suggestions that could assist in transparency are:

- Country of Origin must be disclosed. It is important to know whether the AI model was developed in a Western country or elsewhere in the world. This information can influence users' perceptions and understanding of potential cultural biases or differences in AI systems.
- Transparency about use cases and their associated risks is important. Different use cases may carry varying levels of risk, depending on whether they are employed by governments, businesses, or in social/racial demographics.
- While the technology itself may originate from any location, the data used to train AI models should either originate from Australia or be appropriately contextualised and relevant to the Australian context. And aligned to Australian Community expectations.
- Providing insights into AI systems' error rates and any potential hallucination phenomena will help users assess the reliability and limitations of AI technologies.

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?

Just as the government will take measures to ban applications that threaten national security or disseminate misinformation, a similar approach should be adopted for AI-based solutions. This means that AI applications that pose significant risks to individuals, society, or national interests should be carefully evaluated and, if necessary, restricted, or banned.

Implications and infrastructure

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?

There could be various potential factors effecting the tech sector and trade and export with other countries. However, it is crucial and fundamental to safeguard the interests of Australian citizens by considering the following two very important aspects:

• Understanding the purpose behind use of high-risk activities. Clarifying the purpose helps prevent potential misuse and ensures alignment with ethical and legal principles.

• It is essential to verify whether the provider, government, or industry has obtained proper authorisation to engage in these high-risk activities. Ensuring appropriate awareness and permissions are in place.

Risk-based approaches

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

Yes, however, it is important to gain a comprehensive understanding and clarity regarding the challenges and the problem we aim to resolve. A standardised risk framework may not be sufficient to bridge the gaps and tackle the issues. Contextual risk analysis is imperative to address the complexities involved.

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

Adopting a holistic perspective that considers intervention points across the entire AI supply chain is crucial. Instead of focusing solely on a single point, understanding the end-to-end process of AI development, deployment, and usage helps identify potential risks and points of intervention more effectively.

While risk frameworks for specific applications have their place, they are just one component of a comprehensive strategy. AI's complexity and impact require multifaceted approaches that encompass various aspects, including technical standards, ethical considerations, legal frameworks, and social implications.
Taking an integrated and comprehensive approach, government and regulators could create a robust AI governance and risk framework that addresses the challenges holistically and ensures responsible and ethical AI development and deployment.

Is a risk-based approach better suited to some sectors, AI applications or organisations than others based on organisation size, AI maturity and resources?

Risk-based approach is contextual and needs to be considered to address the issue and concerns. Each AI application, adoption and use case will have its unique set of risks and challenges, and a one-size-fits-all approach may not be suitable.

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

Yes, there are potentially more elements that should be included, some examples are:

- Country of Origin must be disclosed.
- Transparency about use cases and their associated risks is important.
- The data used to train AI models.
- Error rates and any potential hallucination phenomena.
- Privacy concerns
- Copyright infringement
- Purpose and permission aspects of risks associated with the use of AI.

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?

Establishing a coordinator role to enforce linking various existing frameworks, laws and regulations with a risk-based approach for use of AI. The coordinator could also be tasked with identifying gaps and focus on those areas to address them quickly.

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

Establishing rights to declaration and identification of AI use would be the first step. And Tthen, applying requirements for all AI-based technologies to undergo risk assessments to ensure responsible and safe deployment. Making the risk assessment reports available on a Government portal providing visibility into the safety measures and potential risks associated with the use of the AI technologies. A Ppublic framework for evaluation and risk management would also be required. While a risk-based approach is a sensible approach, a lot depends on the execution. For example, would consumers only want to know that a system or service they use has AI enabled features and will make decisions without a human in the loop or do consumers want to know if the system has passed ethical testing, was built in Australia and that there are detailed processes where issues or grievances can be escalated to a human for review, action and accountability. In the food packaging example in Australia, the five-star food health rating is useless compared to the nutritional information, ingredients list and country of origin information. Hence the relevance and usefulness of a risk-based approach will depend largely on the design, education, and information it conveys to consumers.

Contributors

Akash Mittal
Board Director, AISA

Damien Manuel
Board Director, AISA

Joshua Craig
Board Director, AISA