Australasian Cyber Law Institute

Department of Industry, Science and Resources
Consultation paper "Safe and responsible AI in Australia"

https://www.industry.gov.au/news/responsible-ai-australia-have-your-say

https://consult.industry.gov.au/supporting-responsible-ai/submission

ACLI Submission

We are pleased to submit this paper in response to the Department of Industry, Science and Resources ("DISR") "Safe and Responsible AI in Australia" Discussion Paper.

Definitions

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

We consider that the definitions set out in Section 1.2 of the Discussion Paper would be enhanced by the addition of robotic systems including autonomous vehicles, social robots, delivery vehicles and drones, surgical robots, and similar technologies.

The advantage of a wider definition is that the same considerations for trustworthiness will apply consistently across all these technologies. We consider that suitable definitions are set out in IEEE SA's Ethically Aligned Design Ed2, on which Australia's AI Principles are based.

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?

The Australasian Cyber Law Institute considers that the following risks from AI may not be adequately addressed by Australia's existing regulatory approaches:

1. Privacy Breaches
2. Societal Harms: Community Division
3. Societal Harms: Invasive Surveillance and Manipulation

Please refer to Annexure A for our detailed response in respect of each of these risks.

The Australasian Cyber Law Institute suggests the following regulatory frameworks could provide useful parallels in the mitigation of these risks:

Useful parallels:

1. **Consumer Law**
   Consumer law states that businesses are not permitted to do certain things even if the consumer 'consents' (or contractually agrees) to them. For example, businesses must not engage in misleading and deceptive conduct.
2. **Health and Safety Law**
   Worksafe references WH&S standards to enhance the law. This would be useful to allow the law to respond more nimbly. WTO Article 20 of General Agreement on Tariffs and Trade protection for human, animal or plant life or health is another useful reference.
3. **Modern Slavery Law**
4. **Human research approvals**
   HREC system under which safety of research must be approved[1]
5. **International Trade Laws**
6. **Quality Standards such as ISO 9001, ISO 19011, etc**
7. **Ban on biological warfare**
   Banning of biological warfare – Coordination of global governments to achieve global prohibition
8. **CDR (Consumer data right)**

Support for Ex ante measures

Schmidt (2017) identifies two main approaches to handling AI risk:

- 'Ex-ante measures – to understand, guide and implement an ethics and values-based framework for AI research to mitigate the risk of civilisation-ending AI; and
- Ex-post measures – to understand, design and implement countermeasures that operate should the ex-ante measures fail to prevent the creation of civilisation-ending AI'.[2]

On this approach, the question is should we take action beforehand to prevent problems or should we wait for sufficiently severe problems to arise that extensive redress is being sought by the community or a particular section of the community?

ACLI takes the view that the potential harms from poorly implemented AI are severe enough to warrant the adoption of a safety-based approach that imposes responsibility on those deploying AI to ensure it does not do harm. The ethics and values-based framework represented by the Australian AI Principles represents a suitable benchmark against which a safety-based approach could be structured.

[1] https://www.nhmrc.gov.au/research-policy/ethics/human-research-ethics-committees

[2] Schmidt, P 2017, 'Brave new world or the end of it? Regulating artificial intelligence (AI) begins with understanding the real risks', Mondaq Business Briefing, Sept 6, 2017, <http://www.mondaq.com/australia/x/626412/new+technology/Brave+new+world+or+the+end+of+it+Regulating+artificial+intelligence+AI+begins+with+understanding+the+real+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.

The Australasian Cyber Law Institute considers that industry codes setting out the risks that need to be considered and minimum outcomes would be an important element. This allows nuanced treatment of risks – ie, all credit applications should be transparent and fair, but higher standards of review and accountability should apply for more significant transactions such as home loans. We consider that AI may be a valuable tool for processing micro-lending applications and other low risk transactions.

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.

The Australasian Cyber Law Institute considers that the establishment or nomination of a primary regulator is an important part of managing the issues and challenges raised by this technology. The benefit of this would be having a central point of reference for coordination of the many legal, commercial, and societal sectors and issues raised and impacted by AI.

AI governance across Australia draws primarily on two sources: human rights and efficiency. Philosophically, these could be considered under the headings of deontological and utilitarian.

The utilitarian approach can be found in the first principle of the CSIRO's *Artificial Intelligence Australia's Ethics Framework: A Discussion Paper*, which states: 'The AI system must generate benefits for people that are greater than the costs.'[3] The human rights approach can be found in the first principle of the final version of Australia's AI Ethics Principles.[4] Principal one states: 'Human, societal and environmental wellbeing: AI systems should benefit individuals, society and the environment.'

ACLI strongly believes that basing AI adoption on universal human rights principles will help facilitate and strengthen trust in the ongoing development of AI technology.

[3] Dawson D and Schleiger E, Horton J, McLaughlin J, Robinson C, Quezada G, Scowcroft J, and Hajkowicz S (2019) Artificial Intelligence: Australia's Ethics Framework. Data61 CSIRO, Australia. <https://www.csiro.au/en/research/technology-space/ai/ai-ethics-framework>.

[4] Australian Government, Department of Industry, Science and Resources, 'Australia's Artificial Intelligence Ethics Framework, Australia's AI Ethics Principles (Report, 2019).

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?

The Australasian Cyber Law Institute considers that the European approach of mandatory risk assessments when AI is used in certain areas is a useful approach. This approach supports innovation while requiring the work and development to be carried out in a way that is considered and provides for the transparent documentation of the basis for identified risks and the corresponding mitigations.

Target areas

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

We consider that both public and private sector use of AI should be subject to the same laws and standards. We note that due to the mandatory nature of some interactions of the community with government, this should inform the risk rating for such AI systems, with a potentially higher risk rating applying if people cannot opt out of use of the system (or application of the system to them).

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

We consider that the Australian Government can support responsible AI practices in its own agencies by leading by example in the application of risk assessment and risk mitigation frameworks. By demonstrating the application of these frameworks, the Australian Government can provide a useful reference point for model handling of AI tools and risks.

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.

Generic solutions may be appropriate for consideration for an end user of an AI-based product to assist them in selecting and acquiring an AI-based application or tool which meets their needs and addresses their risks. For example, for a human resources organisation wanting to use an AI-based tool to generate policies, or a guideline or assessment tool which assesses the local privacy (and other relevant) obligations and regulations against AI-based options would allow them to select an appropriate AI-based tool that reduces their risks of breaching the privacy obligations.

Technology-specific solutions would be valuable in two scenarios:

1. **Use of API for existing AI tool**
   For organisations that are integrating an application with an existing AI solution, for example an organisation that develops an application that utilises and integrates with an existing third-party AI API.
2. **Development of foundation AI tools**
   For organisations that build AI applications from the ground up, they build the application, source data and train the AI application.

In these scenarios, design guidelines or regulations may be more appropriate. One example of a voluntary guideline in secure web-based design is the OWASP Top Ten. A draft version of the OWASP Top 10 for Large Language Model Applications has been released: https://owasp.org/www-project-top-10-for-large-language-model-applications/

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

a. where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI?

Transparency has two specific meanings in the AI context. First, being transparent about when AI is being *used*,[5] , and second, being transparent about what the AI *does*.[6]

ACLI supports transparency in the former sense. We propose that transparency in the latter sense should include information about when the basis of AI decisions can be explained, and when it can only be inferred.

We do not consider that AI should be assumed to meet the requirements of the *Evidence Act* (Cth) S 146 which applies to 'Evidence produced by processes, machines and other devices'. Section 146 (2) narrows this definition to a device that 'ordinarily produces that outcome'. As AI outputs beyond the simplest have a level of uncertainty, it is not a technology that 'ordinarily produces [a particular] outcome'.

In addition, because generative AI draws on an enormous range of data sources, it is important to be transparent in advising where generative AI outcomes may potentially breach privacy rights or data protection rights.

[5] As seen in the Californian Bolstering Online Transparency (BOT) Act.
https://www.darpa.mil/program/explainable-artificial-intelligence

[6] For example, as outlined by Explainable Artificial Intelligence (XAI): Défense Advanced Research Projects Agency, 'Explainable Artificial Intelligence (XAI)' (Web Page, accessed 21 July 2023) <https://www.darpa.mil/program/explainable-artificial-intelligence>.

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

ACLI notes that the Food Standard Code labelling standards provides a useful regulatory approach to transparency requirements and may provide a sensible parallel for AI transparency.[7]

10. Do you have suggestions for:

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

When reviewing the potential impact to society and the environment, we consider potential harms are sufficiently severe that prohibition or stricter regulation should be considered for:

- facial recognition and other real-time biometric tracking;
- instances where the application of this technology is used to adversely affect individual, societal or community access to certain essential services and products; and
- any risk of violation of human rights.

Examples of such use include social scoring in connection with social entitlements, judgements or adverse outcomes about an individual based on their human characteristics (i.e. profiling by facial features, including age, background, ethnic group, disability).

Community safety needs to be very carefully balanced with the potential for over-reach that will impact the lives and livelihoods of citizens.

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

Criteria should be based upon both evaluation for risk to human values and/or human rights, where these are evaluated according to an independent and accredited standards methodology such as IEEE 7000, and against holistic wellness criteria, using a methodology such as IEEE7010-2020, or in the case of human rights, against the United Nations' Universal Declaration of Human Rights.

Banning AI applications or technologies should be considered where the adverse effects are evaluated as providing a severe risk to human values, ethics and wellbeing from an individual, community and/or a societal view.

[7] Food Standards Australia and New Zealand, Labelling (Web Page, 7 September 2020).

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

The Australasian Cyber Law Institute considers that Australia's AI Principles are a critical centrepiece in increasing public trust in AI deployment and encouraging more people to use AI. The AI Principles need to be operationalised within AI services developed, produced or procured by an organisation, and may require adjustments to the corporate values espoused by public and private organisations in order to demonstrate their commitment to the AI Principles. ACLI's view is that the government should encourage alignment between Australia's AI trustworthy principles and the values and principles of organisations that acquire, procure, sell, produce or develop AI products. This type of alignment does not generally happen organically, but rather needs to be actively encouraged and incentivised, or alternatively regulated.

Introducing an equivalent of the Consumer Data Right for any products utilising data profiles would improve the ability of people to vote with their feet when they consider that a company's approach isn't acceptable. This would contribute to consumers being engaged and informed of alternative choices, allowing them to select the options that best reflect their own personal values.

Other initiatives that would assist in increasing public trust in AI deployment and encouraging more people to use AI include:

- Engagement and encouragement of universal AI literacy within all communities and Australian society, through an introductory campaign to introduce AI/Autonomous systems, Trustworthy AI, and expected normative behaviour of Trustworthy AI systems.
- Incentivise and encouragement of responsible innovation practices ie. 'ethics & human values by design', 'safety by design', 'wellness by design', 'privacy by design' and 'transparency by design' principles associated with the design, development, procurement and governance of AI and autonomous systems.
- Embed these practices into educational courses from the beginning, ie. Secondary and tertiary education, particularly for any technology-oriented courses.
- Tax breaks and government funding incentives, procurement and the development of trustworthy and ethical AI.
- Government procurement could prioritise systems that demonstrate trustworthiness.
- Certification schemes such as the EU-AI act certification scheme or that of IEEE SA's independent CertifAIEd program, where an independent body is accredited with certifying a system as 'trustworthy' according to certain trustworthy characteristics.

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?

The Australasian Cyber Law Institute considers that there is a possibility that banning high-risk activities such as facial recognition may affect aspirations to export local AI products to China, given China's utilisation of social scoring. However, it would send an important message to the local and

international communities that Australia places a high priority on preserving and promoting ethics, human values and human rights (including privacy and the right to freedom).

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

The consideration and handling of issues relating to responsible AI requires a reasonably high level of technical understanding of the technology and its operation. Regulators will need to be educated in appropriate technological principles and application. There is no one regulator at present for whom responsible AI would already be wholly within scope – regardless of which regulator is appointed or allocated, the office of the regulator will need to have appropriate technical understanding.

Risk-based approaches

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

A risk-based approach would require companies / industry to evaluate their profit motive and product development against fundamental human rights. Firstly, human rights and the potential adverse effects are fundamental and cannot be balanced against companies' interest. Secondly, it leaves companies with motivation / unconscious bias to preference their benefit against human rights.

We would recommend an approach more in line with rights-based approach. As mentioned in the referenced AccessNow paper, a similar approach has been used in the GDPR laws. A risk-based approach may be dumbed down to a simple question similar to 'are we (as a company, government department) willing to gamble that everything will be fine?'

A comprehensive risk framework will begin from the concept of 'risk appetite'. For example, when seeking a new technology benefit, how much risk is a company, government department, or society willing to take in the face of uncertainty introduced by AI technologies.

The Australasian Cyber Law Institute proposes that the risk appetite perspective should be that of society, rather than that of an individual enterprise.

Annexure A
Risks

Privacy Breaches

The use of data by firms creating and training new AI products raises serious data privacy issues. It is common practice for SaaS contracts to include a clause such as the following:

> Notwithstanding the confidentiality or any other provisions of this Agreement, the Supplier may use Customer Data for product development.

For example, Microsoft's privacy statement allows data collected 'through our interactions with you and through our products' to be used to 'improve and develop their products'.[8] This type of clause is common among technology companies and has historically been considered benign. However, the clause takes on an extraordinary breath when interpreted as permitting the use of the data in the development and training of AI products.

The clause is normally accompanied by another clause putting responsibility on the customer organisation to obtain the consent of anyone whose data the customer is uploading (the "data subjects"). This type of clause typically requires the customer to have ensured that the data subjects have consented to the use of the data as declared by the vendor in its terms (which could now include using the data to develop and train AI). A common format for this type of clause is:

> The Customer warrants that it has obtained the consent of any person whose Personal Information is uploaded to the Service by the Customer to the use of such Personal Information on the terms of this Agreement.

In some ways, this type of clause is necessary because the vendor doesn't have a direct relationship with the customer organisation's users. However, many organisations are not aware of the legal effect of such clauses. When taken together, these clauses potentially provide a legal basis for SaaS vendors to utilise any data supplied to them to train AI products that they are developing. This is a substantial undermining of both privacy laws and commercial confidentiality.

Societal Harms: Invasive Surveillance and Bias

Real-time biometric tracking (e.g. facial recognition)

Broadly, facial recognition is when a system or device uses biometric data compiled from people's faces to verify and / or identify a particular person or their characteristics.[9]

This technology can potentially infringe on several human rights.[10] For example, AI-assisted facial recognition can lead to widespread monitoring of citizens by governments. Another example, when

---

[8] Microsoft, *Microsoft Privacy Statement* (Web Page, June 2023) < https://privacy.microsoft.com/en-gb/privacystatement>.

[9]University of Tasmania, Human Technology institute, 'Facial Recognition Technology: Towards a Model Law (Report, September 2022) 2.3.

[10] For a comprehensive chart on AI and potential human right contraventions, see John-Stewart Gordon, *The Impact of Artificial Intelligence on Human Rights Legislation: A Plea for an AI Convention* (Palgrave Macmillan, 2023) 4.4.

used in policing, is that machine bias can lead to people to be incorrectly targeted. As well as contravening someone's fundamental human rights, this also leads to decreasing trust in both AI systems and the police more generally.

Due to the many potentials of risk, specifically human rights contraventions, as mentioned above, ACLI believes that real-time biometric tracking should be either prohibited or strictly regulated.

Social Scoring

Combining mass state surveillance with social scoring raises serious human rights issues. Social scoring is where a government tracks and allocates social points depending on the activities of the person. Depending on what is defined as 'anti-social' a person could lose their job, denied a loan, or even imprisonment or executed. ACLI believes such a system would be diametrically opposed to the values of our liberal democracy.

Machine Bias

Machine bias has been described as the most fundamental problems in AI ethics. ACLI considers that it is vital that all people involved in creating and building AI systems, particularly programmers, are made aware of these deep ethical problems that bias can have on autonomous decisions.

The Toronto Declaration outlines several solutions to this problem. First, the human in the loop approach - that humans should make the final decision, next, a requirement of including team members of a diverse nature and inclusive training for teams at all stages of development. Then, critically, avoiding the use of historical data where possible, and finally, proper testing phase relating to bias and following up with a comprehensive monitoring and recoding the results of the system.[11]

Machine data bias can impact automated government decisions, for example, immigration decisions, bail risk profiling, or social security, as well as automated commercial decisions such as loan approvals or insurance applications. The impacts these decisions can have on people can be catastrophic. For example, bias in AI risk assessments when considering bail have been criticised as inaccurate, racially biased, and that they lead to higher incarceration rates for certain groups. While human judges also make erratic and biased decisions (sometimes more than an AI),[12] ACLI considers that the use of AI has the potential to result in widespread and embedded bias in a way that does not occur with individual human judges.

Societal Harms: Community Division and Manipulation.

As discussed in ACLI's recent submission to Select Committee on Foreign Interference through Social Media,[13] social media algorithms, including Facebook, Instagram, Twitter, Youtube and TikTok, disproportionally amplify divisive and incendiary posts and comments. Facebook, for example, knew and rejected efforts to curb these negative behaviours as it would 'impede the platforms usage and growth'.[14]

It is this commercial imperative for growth at the expense of public interest that is a key issue regarding AI use, where potential community division takes a back seat to shareholder reports and

---

[11] https://www.torontodeclaration.org/declaration-text/english/

[12] See generally David Arnold, Will Dobbie and Crystal S Yang, Racial Bias in Bail Decisions

[13] Australian Cyber Law Institute, Submission No 41 to Select Committee on Social Media on Foreign Interference Through Social Media, Parliament of Australia, Foreign Interference through Social Media (31 October 2021)

[14] Jeff Horwitz and Justin Scheck, 'The Facebook Files: Facebook Increasingly Suppresses Political Movements It Deems Dangerous', The Wall Street Journal (online, 22 October 2021) <https://www.wsj.com/articles/facebook-suppresses-political-movements-patriot-party-11634937358>.

growth forecasts. Shoshana Zuboff, a professor emerita at Harvard Business School, states that the goal is not creating norms of behaviour, such as conformity and obedience, but 'rather to produce behaviour that reliably, definitively, and certainly leads to desired commercial result'.[15]

In this 'attention economy'[16] there is only so much attention to go around. It is then a 'race to the bottom of the brain stem'[17] to keep users scrolling and engaging. Negative and divisive content readily engages a user's emotion, which ultimately amplifies such content. As the Cambridge Analytica scandal demonstrated, mis- and disinformation is particularly suspectable to this type of attention. Falsehoods travel 'significantly farther, faster, deeper, and more broadly than the truth'.[18]

The economics of extraction in data has led companies seeking growth to penetrate deep into life and society. Human life, via data, is 'converted into the raw material for capital'.[19] This cultivates many problematic human states for the purpose of generating engagement and sales. Ultimately, the only outcomes that are being measured and monitored are the economic ones.

A holistic wellbeing approach needs to be employed, along with appropriate measurement instruments. This can monitor and manage the health of individuals, communities and society to effectively manage the impact of introducing AI. Monitoring and management must be introduced at the earliest stages during development prior to deployment of the AI systems, and then after introduction of the AI, examining the impacts for all aspects of wellness criteria that are relevant. See IEEE 7010-2020 for guidance on process and criteria.

[15] Shoshana Zuboff, The Age of Surveillance Capitalism (Profile Books, 2019) 201.
[16] Testimony of Tristan Harris before the U.S. Subcommittee on Communications, Technology, Innovation, and the Internet as part of the hearing titled 'Optimizing for Engagement: Understanding the Use of Persuasive Technology on Internet Platforms' (25 June 2019) <https://www.commerce.senate.gov/services/files/96E3A739-DC8D-45F1-87D7-EC70A368371D>.
[17] Ibid
[18] Peter Dizikes, 'Study: On Twitter, false news travels faster than true stories', *MIT News* (online, 8 March 2018) <https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308>.
[19] Shoshana Zuboff, The Age of Surveillance Capitalism (Profile Books, 2019) 201.