ASPI
AUSTRALIAN STRATEGIC POLICY INSTITUTE

LEVEL 2, 40 MACQUARIE STREET, BARTON ACT 2600

Submission to DISR's public consultation on supporting responsible AI

4 August 2023

Australian Strategic Policy Institute
Cyber, Technology & Security (CTS)

Jacinta Keast
Analyst, Cyber, Technology & Security

This submission does not reflect a single Australian Strategic Policy Institute (ASPI) perspective. It is the opinion of the individual author.

Introduction

The Australian Strategic Policy Institute (ASPI) is an independent, non-partisan think tank that produces expert and timely advice for Australian and global leaders. ASPI generates new ideas for policy makers, allowing them to make better-informed decisions.

Analysts at ASPI CTS aim to inform and influence policy debates in the Indo-Pacific through original, rigorous, and data-driven research. CTS is a leading voice in global debates on cyber, emerging, and critical technologies, foreign interference and issues related to information operations and disinformation. CTS also has a capacity-building team that conducts workshops, training programs and large-scale exercises for the public, private and civil society sectors.

ASPI CTS welcomes the opportunity to provide its thoughts on how the Government should formulate its regulatory and policy responses to responsible and safe AI.

Some of ASPI's published reports on AI include:

- (2023) De-risking authoritarian AI: a balanced approach to protecting our digital ecosystems
- (2023) The Critical Technology Tracker
- (2023) Strengthening Indo-Pacific techdiplomacy in critical technologies
- (2022) Artificial intelligence and policing in Australia
- (2022) Artificial intelligence: your questions answered
- (2020) Weaponised deep fakes
- (2018) Social credit in China
- (2018) Technological entanglement: Cooperation, competition and the dual-use dilemma in artificial intelligence

DISR has asked for responses to a list of 20 consultation questions. Each one of these would require extended study, analysis, and assessment. Our submission provides insight on a number of these questions. As many of the answers to these questions cross over more than one category, our responses are presented in an integrated fashion, with the relevant question numbers they respond to indicated in a box below each recommendation.

ASPI CTS encourages DISR to continue to liaise with stakeholders in formulating its regulatory and policy responses to responsible AI. We are open to further consultation and briefing for the department on any of our areas of expertise and are reachable through the contact details provided in our submission.

Key Recommendations

1	The Australian Government should recognise AI as a dual-use technology. DISR should coordinate with relevant Australian government departments on the implications of AI as a dual-use technology for policy and regulatory responses.
2	DISR should regularly update its definition of AI to encompass new and major technologies and applications because AI is a rapidly evolving technical field. This should be done in consultation with a 'Responsible AI council' through DISR.
3	DISR should not follow the European Union's AI act to approaching risk. It should develop its own risk framework, based on likelihood and consequence, and this should be used by other Australian government agencies to inform policy and legislation.
4	The use of AI in certain harmful activities, such as proliferation of child sex abuse material (CSAM) or electoral interference, should be specifically banned through new or amended legislation.
5	The use of non-regulatory initiatives like sandboxing, funding for AI safety research, and a 'Responsible AI council' through DISR would support responsible AI practices in Australia, while maintaining incentives for innovation.
6	To increase transparency and public trust in AI, DISR's National Measurement Institute, like the National Institute of Standards and Technology (NIST) in the US, should conduct technical tests and publish the results of bias in high-risk algorithms used in the public sector. This testing function could extend to governments in Southeast Asia and the Pacific.
7	Relevant Australian government agencies, such as the Department of Foreign Affairs and Trade (DFAT) and DISR, should coordinate with like-minded partners in the Indo-Pacific on AI governance initiatives and engage in 'techdiplomacy'.

Recommendation 1: The Australian government should recognise AI as a dual-use technology. DISR should coordinate with relevant Australian government departments on the implications of AI as a dual-use technology for policy and regulatory responses.

Corresponds to q1, q2, q4

DISR's discussion paper should foreground AI's fundamental nature as a dual-use technology. 'Dual-use' refers to an item, technology, or service's ability to be used for both civil and military purposes. Dual-use technology erases a fundamental regulatory distinction that treats military weapons and consumer products as two distinct categories.

Several examples of AI with both civil and military use include:

AI-application or technology	**Civil use**	**Military use**
Chatbots based on large language models	Use in writing editors or language learning	State-backed hacking campaigns or information operations
Unmanned aerial vehicles ('drones') with AI	Improve crop yields in agriculture	Combat or reconnaissance scenarios
Personalised medicine and biotechnology	Drug discovery and personalised medicine	Genetic engineering or development of bioweapons

Failing to recognise AI as dual-use could mean policy and legislation would fail to consider serious national security risks. These could include risks to:

- Cybersecurity
- Critical infrastructure
- Weapons proliferation and terrorism
- Social cohesion

DISR should coordinate with relevant Australian government departments on the implications of AI as a dual-use technology for policy and regulatory responses. A shared understanding would assist in coordination of regulation across government, particularly in harmonising definitions and understanding of AI between national security and non-national security policymakers. In the long run, this would decrease coordination costs for the government, and compliance costs for businesses subject to regulation.

Recommendation 2: DISR should regularly update its definition of AI to encompass new and major technologies and applications because AI is a rapidly evolving technical field. This should be done in consultation with a 'Responsible AI' council through DISR.

Corresponds to q1, q2

Policymakers' ideas about how to regulate AI stem from an understanding of what the market looks like today, as well as forecasts of what the market will look like in the near future. There are assumptions about gains in technical capability of existing and emergent AI technologies, as well as the market share of certain technologies and how important foundational AI (such as large language models) will be to the market relative to applied AI. Policymakers also forecast which countries will have the most powerful AI companies, and how humans will use technology. For example, in August 2023, while generative AI companies and large-language models are major players in the AI market, should we assume these companies will be the dominant players in industry in five years' time? It could be the case that the dominant AI companies in five years' time will focus on automated and mechanised processes, such as manufacturing and agriculture.

Similarly, predictions of what AI will look like in five years' time make assumptions as to what inputs will be available and affordable. This might include inputs such as the raw computing power available to companies (and governments), high-performance computing chips, as well as a labour force with the technical skills to continue making scientific advances.

To ensure policy on responsible and safe AI keeps pace with innovation, CTS recommends DISR regularly update its definition of AI. This would ensure that regulation or policy made for the field is adaptive to technical advances and challenges. As per recommendations 2 and 5, DISR could update its definition by consulting with a Responsible AI Council. To build transparency and inform industry and the public, the Council would publish a short annual public report on the State of AI, describing the Council's forecasts for the AI industry.

Recommendation 3: DISR should not follow the European Union's AI act to approaching risk. It should develop its own risk framework, based on likelihood and consequence, and this should be used by other Australian government agencies to inform policy and legislation.

Corresponds to q4, q7, q8, q14, q15, q16, q17

CTS recommends that an Australian approach to responsible and safe AI avoid emulating after the European Union's AI act and a risk-based approach. Australia has a strong environment for innovation and business. The European regulatory approach would hamper Australia's ability to innovate in Ai, including reducing our attractiveness for foreign investment and collaboration.

As AI is a foundational technology, it would make Australian businesses less competitive in global markets. Many of these businesses would be in Australia's small and medium business sector. The Australian military may become less competitive, procurement of products and services may become more expensive and Australians' access to quality public services may diminish.

Australia is a much smaller economy than the European Union, has comparatively few major international technology firms and is a net importer of technologies that use AI. While Australia's university research sector is strong relative to its commercial players, it is still dwarfed by larger economies. According to ASPI's Critical Technology Tracker, Australia produces 1.8% of high-impact research in artificial intelligence algorithms and

hardware accelerators, whereas the US and China produce almost half of the world's most high-impact research.[1]

This means Australia has less of an ability to affect firm behaviour in AI than economies such as the EU, the United States or China. The 'Brussels Effect' sees firms comply with EU laws outside of the EU, as it is not economically practical to have multiple compliance systems. Strong legislative action in Australia could have a 'Reverse Brussels Effect' whereby companies and individuals decide they cannot comply and move away.

As noted in Recommendation #2, Australia's ability to have regulation keep pace with technical breakthroughs in AI and technologies that use AI will be crucial in minimising harm to Australian citizens. However, heavy-handed regulation that categorises tomorrow's AI products based on today's technical possibilities, and adopts a narrow view of risk, would not allow enough flexibility for new products to market.

What should be done? As explained in Recommendation #4, CTS recommends that the riskiest activity be subject to specific legislation; less risky activity should be regulated through existing legislation in Australia and sector-specific non-regulatory instruments.

A framework that measures risk should consider both likelihood and consequence as appropriate inputs. Risk as currently presented in Box 4 is presented as a function of three inputs: level of impact, duration of impact and reversibility of impact. While acknowledging consequence, this ignores the *likelihood* that a risk or harm will occur, an input generally recognised as crucial to risk management frameworks.

If likelihood is not considered, regulatory or policy responses be too onerous that they stifle innovation or too light that they fail to mitigate major harms. It may also not provide businesses with appropriate guidance as to where they can invest in mitigations to negate appropriate risk.

CTS notes that determining likelihood of harm or risk in AI technologies will be a necessarily more difficult undertaking than evaluating risks from traditional technologies. AI-based systems are prone to complex technical failures, such as distributional bias, generalisation failures and adversarial data. The presence of these harms, the likelihood of harm and the degree to which they may impact users can often only be evaluated by technical experts. Evaluating harm could also be hard due to unforeseen combinations of AI technology with technical advances in other areas, for example a breakthrough in robotics could augment the capacity of AI software used in robots to generate harm or benefit to users.

As risk and harm in AI technologies is dynamic, CTS would advocate a dynamic framework that is updated at least yearly, rather than making static risk assessments. DISR should lead on developing this harm-based framework and it should be used by other Australian government agencies to inform policy and legislation. This would assist in coordinating regulation across government, particularly in an understanding of harm that is shared between national security and non-national security policymakers. For example, as per Recommendation #1, DISR might consider particular harms that arise from a technology being dual-use.

[1] Measured by Australia's performance in the category 'Artificial intelligence algorithms and hardware accelerators. 'High-impact research' is defined as the proportion of publications in the top 10% of the most highly cited papers See: (2023) ASPI Critical Technology Tracker, 'Artificial intelligence algorithms and hardware accelerators'.

Recommendation 4: The use of AI in certain harmful activities, such as proliferation of child sex abuse material (CSAM) or electoral interference, should be specifically banned through new or amended legislation.

Corresponds to q10(a), q10(b)

There are certain harmful activities that will have their impacts and reach turbocharged by AI. Australia already criminalises or penalises these activities through federal legislation, but it should introduce legislative amendments or new legislation to criminalise or penalise cases of when AI is used in these activities.

A non-exhaustive list of activities that should be considered unlawful when they use AI technology would be:

- Proliferation of AI-generated child sex abuse material (CSAM)
- Production of deep fake pornography or impersonation of political figures
- Online-based election interference
- Cybercrime
- Online coordinated incitement to racial violence or terrorism
- Production of bioweapons or nanoweapons

While existing legislation may cover some of the most egregious uses of AI, it should be evaluated whether they provide significant enough protection against the speed and scale of harm that a technology such as AI can provide. For example, while existing legislation targets online-based election interference, such remedies under the legislation may not allow lawmakers to respond rapidly enough to a coordinated, mass inauthentic online campaign of AI-generated content. While an actor may face prosecution, the damage may already be done before legal and policing authorities can act.

Under a risk-based framework, there may also be activity that has a high potential to harm individuals, organisations, or systems but that may be still in the public interest, if appropriate mitigations are put in place. CTS suggests legislation that adopts a 'why not' approach rather than a 'why yes' approach, where firms or universities are encouraged to innovate provided that they meet mitigation requirements. Getting this right would ensure that scientific breakthroughs, which are low probability and high impact, are encouraged in Australia. This balance is particularly important because AI research in academia is one of Australia's most promising areas for growth in the AI field.

Examples of activity that uses AI that has a high potential to harm individuals but could still be in the public interest include:

- Personalised medicine and drug discovery
- Autonomous vehicles and self-driving cars
- Video surveillance or biometric technology for public safety

Only the most harmful activity should be regulated through specific legislation. Other kinds of activity should be regulated through existing legislation in Australia -- e.g., *Privacy Act 1988*, *Competition and Consumer Act 2010*, *Online Safety Act 2021* -- and the use of sector-specific non-regulatory instruments. As discussed in recommendations #5 and #6, this might include initiatives such as regulatory sandboxes that would maintain a balance between innovation and safety.

Recommendation 5: The use of non-regulatory initiatives like sandboxing, funding for AI safety research, and a 'Responsible AI council' through DISR would support responsible AI practices in Australia, while maintaining incentives for innovation.

Corresponds to q3

As noted in recommendation #2, the pace of technical innovation in AI poses challenges for regulators and policymakers, many of whom will not have the same technical depth and first-hand understanding of the technology and its risks, as those working in industry. Regulatory sandboxing would be a useful non-regulatory initiative to encourage innovation and help policymakers understand the balance of benefits and harms of new technology.

Regulatory sandboxing is a particularly useful tool to spur Australian innovation, considering many early-stage startups rely on international science and technology collaboration. This policy tool would help ensure Australian startups remain internationally competitive and allow them to grow.

CTS also recommends that DISR set up a 'Responsible AI council' to help the department and the government understand how to make policy for responsible AI. DISR should learn from other department's experience in establishing advisory councils for issues that cut across several sectors. Members of the council would be appointed by the Minister for Industry and Science, with diverse representation encompassing industry leaders, AI safety academics, military technologists, ethicists, and leaders from human rights groups.

One major strength of Australia's AI research environment is its academic sector and CSIRO, particularly in applied AI capability.[2] DISR should collaborate more with academic research institutes and research groups around the country on AI safety. As well as academic appointments to an advisory council, DISR could look at administering a grants scheme for public research on topics of AI safety and AI alignment. This could be modelled after the Clay Institute's Millenium Prize Problems, and DISR could award prize money through the National Artificial Intelligence Centre as incentives for researchers to solve alignment and safety problems.

Recommendation 6: To increase transparency and public trust in AI, DISR's National Measurement Institute, like the National Institute of Standards and Technology (NIST) in the US, should conduct technical tests and publish the results of bias in high-risk algorithms used in the public sector. This testing function could extend to governments in Southeast Asia and the Pacific.

Corresponds to q3, q5, q7, q11

The National Institute of Standards and Technology (NIST) is a physical science laboratory in the United States, part of the U.S. Department of Commerce. Its mission is to advance US innovation and competitiveness through measurement science and technical standards. It is known for its Facial Recognition Vendor Test (FRVT) which tests algorithms for technical performance, as well as their performance on bias. Both corporations and research universities can submit technology to NIST, and testing has served as a challenge that spurs innovation as research groups try to improve their rankings and performance.

[2] CSIRO, 'Artificial Intelligence'

Crucially, NIST publishes their research methodology, performance, and rankings publicly. Journalists, civil society, and AI safety researchers can use this data to inform research and investigations.[3] The data has also been used in the US Congress to debate the merits of facial recognition.[4] This has built a public conversation around the risks and benefits of AI and helped improve public trust and transparency around AI.

CTS recommends that DISR establish a similar function within its National Measurement Institute, inviting research universities and corporations to submit algorithms that are used or have the potential to be used, in the public sector. This might include, as an example, algorithms used in facial recognition or video surveillance systems. CTS recommends that the Institute publish the results of testing of these algorithms and distribute this information to Australian and regional journalists, civil society, and AI safety researchers to improve Australia's public conversation and discussion around AI safety in the public sector.

As part of Australia's techdiplomacy (see Recommendation #7), the Institute could also serve as a regional centre of excellence, assisting governments in Southeast Asia and the Pacific to test algorithms used by their governments. Added transparency through rigorous, reliable testing in this way could contribute to a nuanced public discussion on the merits of using technology from providers, such as China, with known racial bias issues in facial recognition or video surveillance technologies.[5] As well as improving AI safety and governance in the region, it could benefit Australia's human rights diplomacy in the region.

Recommendation 7: Relevant Australian government agencies, such as the Department of Foreign Affairs and Trade (DFAT) and DISR, should coordinate with like-minded partners in the Indo-Pacific on AI governance initiatives and engage in 'techdiplomacy'.

Corresponds to q3, q4, q5

Because Australia is a net importer of AI technologies, Australian diplomatic efforts to improve governance mechanisms in the Indo-Pacific on AI will have a positive effect on the development of safe and responsible AI in Australia. This will mean the need for increased Australian diplomatic and industry presence in regional and international technical forums, increased training of Australian representatives to these forums, and increased coordination between DFAT and DISR on strategies to harmonise domestic and international governance for responsible and safe AI.

There are numerous international forums for the global governance of AI where Australia can have influence, including multilateral groups, regional groups, mini-laterals and industry groupings. Several examples of where Australian diplomats and representatives from industry could engage in 'techdiplomacy' include:

[3] For example (2021) Federation of American Scientists, 'A More Responsible Digital Surveillance Future'; (2022) ACLU of Massachusetts, 'Request for Information on Public and Private Sector Uses of Biometric Technologies: Responses'; (2020) The Verge, 'Facial recognition systems are getting better at recognizing masked faces'.

[4] For example (2023) 'Testimony of Alexandra Reeve Givens for the U.S. Senate Committee on the Judiciary Subcommittee on Human Rights and the Law Hearing Entitled Artificial Intelligence and Human Rights '; (2019) 'About Face: Examining the Department of Homeland Security's Use of Facial Recognition and Other Biometric Technologies', Committee on Homeland Security

[5] (2019) 'One Month, 500,000 Face Scans: How China is Using A.I. to Profile a Minority', *The New York Times*

- International Organisation for Standardisation (ISO)
- Institute of Electrical and Electronics Engineers (IEEE)
- International Telecommunications Union (ITU).

ASPI CTS has partnered with the Centre for Internet & Society in India to produce a 'techdiplomacy guide' to help those negotiating technical standards in AI.[6] More work between think tanks or at the track II level in the Indo-Pacific is needed.

[6] (2023) ASPI, Strengthening Indo-Pacific techdiplomacy in critical technologies