



Australian Government

**Department of Industry,
Science and Resources**

Online Safety Act Review

Department of Industry, Science and Resources
Written Submission

June 2024

Contents

Contents.....	2
Introduction	2
Background.....	2
Potential risks of AI	3
Supporting the safe and responsible use of AI	4
Regulatory clarity and certainty	5
Supporting and promoting best practice for safety and adoption	7
Related work across Government	8
Leveraging international approaches	9
Glossary	10

Introduction

The Department of Industry, Science and Resources (DISR) welcomes the opportunity to provide a submission and support the Statutory Review into the Online Safety Act (2021).

DISR notes the Statutory Review of the Online Safety Act (2021) Issues Paper, released 26 April 2024, highlights how emerging technologies such as generative artificial intelligence (AI) and algorithms have the potential to amplify online harms.

DISR is leading coordination of the whole-of-government effort on Safe and Responsible AI. DISR is consulting with industry, academia and civil society to deliver advice to government on how to harness the benefits of this technology while mitigating against its serious risks. An objective of this work is to ensure consistent and coherent approach to supporting the safe and responsible use of AI across the economy, recognising individual portfolios are responsible for their own legislation. DISR is working closely with the Department of Infrastructure, Transport, Regional Development, Communications and the Arts (DITRDCA) and other relevant agencies to ensure work on safe and responsible AI is harmonised with work seeking to address online harms.

Background

The potential for AI systems and applications to improve wellbeing, quality of life and grow our economy is well known. The design, development and deployment of AI will transform industries, create new jobs, drive innovation, improve decision-making, enhance customer experiences, and support small businesses. AI and automation could generate between \$170 to \$600 billion a year towards Australia's GDP by 2030¹.

At the same time, AI presents a potential to create or amplify harms to individuals, organisations, communities, and social cohesion. Harms include discrimination, bias, mis- and dis-information, socio-economic disparities, and health and safety risks. This includes amplification of online harms. These harms may disproportionately affect vulnerable and marginalised groups, particularly people with cognitive disability, displaced workers and older people. It may also pose higher risks of harm to culturally and linguistically diverse communities, regional communities, women and girls, and gender equality more broadly.

Through the Safe and Responsible AI agenda, the Australian Government is acting to ensure the design, development and deployment of AI systems in Australia in legitimate, but high-risk settings, is safe and

¹ C Taylor, J Carrigan, H Noura, S Ungur, J van Halder and G Dandona, Australia's automation opportunity: Reigniting productivity and inclusive income growth, McKinsey & Company, 2019.

OFFICIAL

can be relied upon, while ensuring the use of AI in low-risk settings can continue to flourish largely unimpeded.

In addressing the harms from AI and other emergent technologies, it will be important for the Online Safety Act and other regulatory measures, including potential AI regulations, to be consistent and complimentary. Measures should operate alongside each other to support a cohesive and comprehensive regulatory ecosystem that effectively prevents harms whilst limiting unnecessary regulatory burden on businesses.

Potential risks of AI

While AI has a range of legitimate uses, there are documented instances of harm. Through the government's consultation on the *Safe and Responsible AI in Australia* discussion paper, DISR heard that current regulatory frameworks likely do not sufficiently address known risks presented by AI systems.² It is likely that existing laws do not sufficiently prevent AI-facilitated harms *before* they occur. Putting in place guardrails against these risks will create the right settings for AI innovation and adoption in Australia, allowing benefits to be fully realised across all sectors.

Consultations presented diverse views on the most serious, urgent and probable risks of AI. These are broadly categorised as technical risks, unpredictability and opacity risks, domain-specific risks, systemic risks, and unforeseen risks.

Online harms are a key **domain-specific risk** identified during the consultation process. Domain-specific risks arise where AI interacts with existing harms, systems, or legislative frameworks, and new risks can arise, or existing risks can be exacerbated. Examples include the generation and spread of online harms like deepfake pornography or AI generated cyber-abuse, and the undermining of social cohesion through misinformation or disinformation generated, tailored and spread by AI.

These risks and harms may particularly affect vulnerable or marginalised groups (which are also groups identified as particularly at risk in the Online Safety Act issues paper), including First Nations communities, gender diverse Australians, women, people with disability, LGBTQI+ communities, and culturally and linguistically diverse Australians. Accordingly, government is acting to capture the views and needs of these groups when exploring options to support safe and responsible AI.

Additional risks from AI include:

- **Technical and data risks:** the outputs of AI systems can be compromised by technical limitations, including inaccuracies in system design or biases in training data. This can result in inaccurate or unfair outcomes for people or groups. For example, if AI models in healthcare are trained on non-representative data, they may contribute to disparity in health outcomes for underrepresented groups.
 - Online harms can occur from users interacting with poorly trained chatbots and other interactive interfaces, which can reproduce bias within the datasets used to train the models.
- **Unpredictability and opacity:** opaque AI systems can make it difficult to identify harms, predict sources of error, establish accountability, explain model outcomes and assure quality. For example, if job applications are assessed by AI systems where internal workings are automated and invisible, people affected by discriminatory outcomes may have limited ability to understand or question decisions.
- **Systemic risks:** emerging AI developments can create systemic risks including risks associated with the development of highly capable and potentially dangerous frontier models as well as the greater accessibility and useability of generative AI models. These developments can lead to unpredictable harms on a scale and at a speed not previously possible.
- **Unforeseen risks:** AI is evolving with a speed and complexity that will likely pose unforeseen risks. The rate of technological change makes it difficult to ensure that regulation is future-proofed and

² DISR, [Safe and responsible AI in Australia Consultation Australian Government's interim response](#), Jan 2024.

OFFICIAL

can meet unforeseen challenges without stifling innovation and the chance for AI to solve some of our most pressing problems. Accordingly, regulatory approaches must be flexible and responsive to risks as they emerge.

These harms can occur on three broad levels:

- **Individual:** harms to individuals including discrimination, exclusion, bias, defamation and reputational damage, threats to physical or psychological safety, breaches of privacy.
 - AI used to predict recidivism can disproportionately target marginalised groups.
 - Recruitment algorithms can unfairly penalise diverse groups, including women, people with disabilities and culturally and linguistically diverse people.
 - AI may be misused to generate deepfake images and videos, including non-consensual and sexually explicit deepfakes which almost exclusively target women.
- **Organisational:** harms to organisations including reputational or commercial damage.
 - Increased digitisation exposes organisations to greater chances of cyber-attacks which may expose personal data of clients.
 - Profit loss due to AI errors like incorrect valuations.
 - Legal risks from AI that results in biased or discriminatory decision-making for example in recruitment which may breach Commonwealth laws including the *Racial Discrimination Act 1975* and the *Sex Discrimination Act 1984* and expose an organisation to civil liability.
- **Societal:** harms to society at large including environmental damages, perpetuating inequality or bias, and distributing harmful content.
 - Dissemination of misinformation or disinformation at scale.
 - AI-generated videos undermining electoral and democratic processes.
 - AI driven facial recognition technology being used for combatting theft may risk discrimination based on skin colour (with higher rates of error).
 - Increased uptake of AI may lead to increased energy usage and emissions.
 - AI algorithms which seek to suppress harmful content may demonstrate bias against women and women's bodies, or "shadow ban" content with images of women, with implications for online reach and women's online businesses.³
 - First Nations cultural intellectual property rights may be at risk from AI, for example imitations of arts that resemble an 'Indigenous style'.⁴

Supporting the safe and responsible use of AI

Across all AI policy initiatives, government is focused on operationalising Australia's AI Ethics Principles.⁵ These principles provide important values-based guidance for the intent of regulatory design and are strongly aligned to internationally recognised principles on ethical and responsible AI.⁶

With AI systems becoming increasingly powerful and pervasive across our economy and society, the overwhelming view from stakeholders is that voluntary compliance with Australia's AI Ethics Principles and standards is no longer sufficient for high-risk AI applications – implementation of principles through practical strategies, clear obligations and effective enforcement will be key.

³ G Mauro and H Schellmann, [There is no standard': investigation finds AI algorithms objectify women's bodies](#), The Guardian, 8 Feb 2023

⁴ T Worrell, [Generative AI in the classroom risks further threatening Indigenous inclusion in schools](#), *The Conversation*, 6 Feb 2024.

⁵ Department of Industry, Science and Resources, [Australia's AI Ethics Principles](#), Department of Industry, Science and Resources, 2019.

⁶ Organisation for Economic Co-operation and Development, [OECD AI Principles Overview](#), OECD, 2019.

OFFICIAL

In line with the Australian Government's Interim Response on *Safe and Responsible AI in Australia (the Interim Response)*, government is taking action to ensure the Australian economy and community secure the benefits of safe and responsible AI (Figure 1), through:

- Delivering regulatory clarity and certainty.
- Supporting and promoting best practice for safety.
- Ensuring government is an exemplar in the use of AI.
- Engaging internationally on how to govern AI.

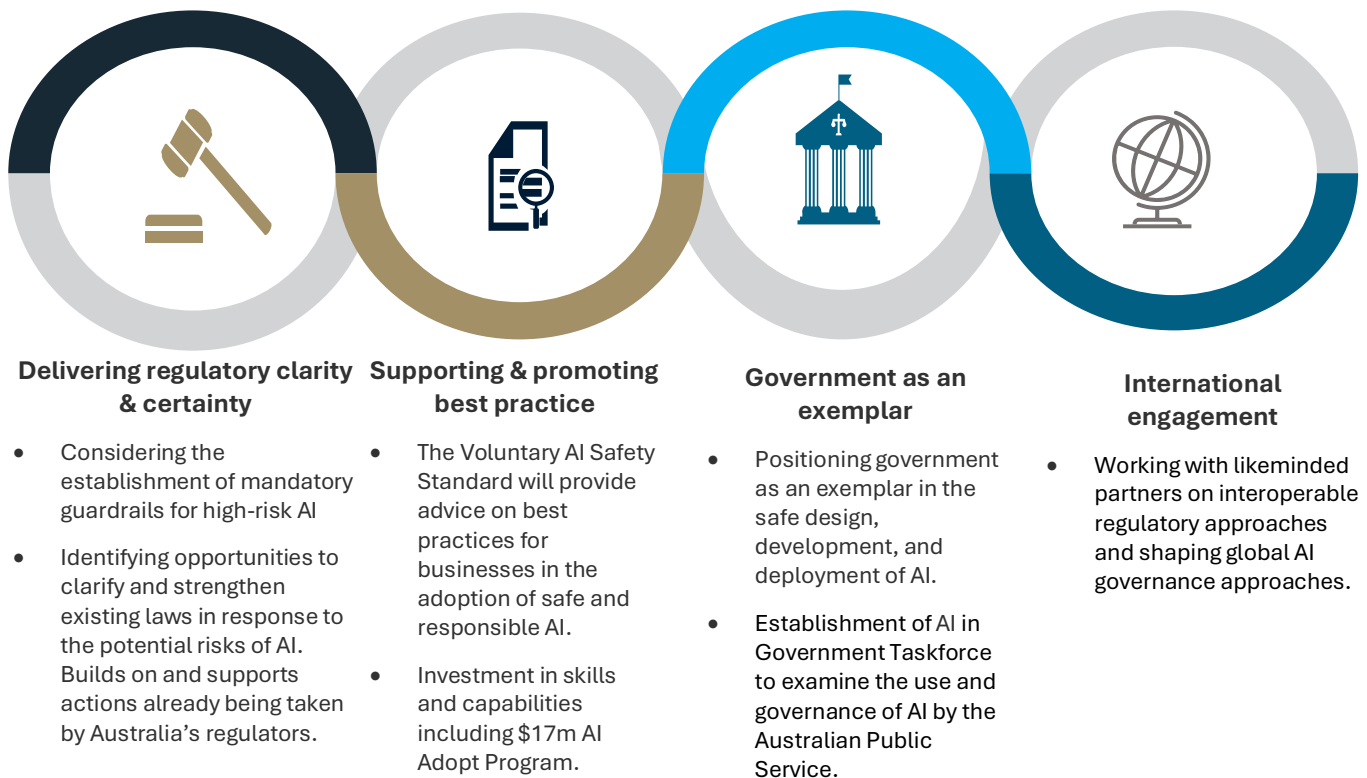


Figure 1: Safe and Responsible AI in Australia: government action.

DISR is focussed on ensuring a consistent and comprehensive approach to AI regulation, ensuring work on new laws, or strengthening and updating existing legislation and regulation is aligned to the government's agenda for safe and responsible use of AI across the economy.

Regulatory clarity and certainty

Industry and civil society have been clear that clarity and certainty on how AI will be regulated is required to fully realise the opportunities AI presents. Government is progressing two key streams of work:

- Considering options for mandatory guardrails for organisations designing, developing and deploying AI systems in high-risk settings.
- Strengthening and clarifying existing laws to address risks and harms of AI.

Considering options for mandatory guardrails

Business and individuals who design, develop and deploy AI are already subject to various Australian laws. These include economy-wide laws such as those relating to privacy, online safety, corporations,

OFFICIAL

intellectual property and anti-discrimination laws and sector-specific laws such as those applying to medical devices, motor vehicles, airline safety and financial services.

These laws will respond, and in some cases already are, to some of the risks of AI. This includes reform processes such as the review of Australia's privacy regime, or the Online Safety Act. However, many of our laws, including our consumer laws, are primarily focussed on providing redress after a harm has taken place. This may not be appropriate when harm is generated at speed and scale before any enforcement activity can take place. The fact that individuals may not know they are engaging with AI systems means they may not know they are being harmed and will be unable to seek redress.

Governments around the world are considering regulatory responses that focus on preventing harms before they arise. This is a seismic shift in the way governments approach regulation of technology. There has been increasing recognition globally of the potential value of preventative interventions, for example, in the context of competition regulation of digital markets.⁷⁸

As outlined in the Interim Response, the government is working to ensure the design, development and deployment of AI systems in Australia in legitimate, but high-risk settings, is safe and trustworthy, while allowing the use of AI in low-risk settings to continue to thrive within the bounds of existing laws.

The government has committed to a risk-based approach in the consideration of new mandatory guardrails for AI. In designing a risk-based regulatory regime for AI, consideration needs to be given to:

- The levels of risk and key characteristics of known risks.
- The balance of preventative and remedial regulatory measures to effectively target and mitigate known risks.

Several features of AI make this technology well-suited to a risk-based and preventative approach to regulation:

- Potential for harms to spread across the economy and community at speed.
- Potential for catastrophic harm, such as AI being used to manipulate electoral processes on a mass scale or causing wide scale disruptions to the energy grid.
- Highly context-specific harms - An AI system deployed within one sector for a particular purpose may present very low risk of harm, yet once applied within a different sector present a high-risk of harm due to differences in domain-specific risks or impacted parties. For example, an AI driven chatbot system could be deployed with low risk in a customer service setting but its deployment in a health care context may lead to serious consequences if a recommendation is incorrect.
- Potential harms arising not only to individuals, organisations, communities, social cohesion, and society at large.
- Uncertainty about how and what types of AI harms might arise as technology evolves. This uncertainty will require regulatory measures and enforcement tools which can successfully adapt to new forms of high-risk AI.

The government has established a temporary **AI Expert Group**⁹ (Expert Group) with 12 appointees spanning industry, academia, and legal expertise to advise on options ahead of the government's consideration and public consultation. The government has asked the Expert Group to provide advice on options for mandatory guardrails for high-risk AI with a focus on testing, transparency and accountability measures. If adopted, mandatory guardrails could place obligations on those in the AI supply chain (that is designers, developers and deployers) that are best placed to take early action to prevent harms.

The Expert Group is considering the following issues:

⁷ OECD, *Ex Ante Regulation and Competition in Digital Markets*, Accessed 10 May 2024

⁸ The Treasury, *Government Response to ACCC Digital Platform Services Enquiry*, The Australian Government

⁹ Department of Industry, Science and Resources, *New expert group will help guide the future of safe and responsible AI in Australia*, Australian Government, 14 February 2024

OFFICIAL

- **A definition of high risk.** Internationally, several jurisdictions have, or are, considering how to define high-risk. This consideration has included:
 - A principles-based approach grounded in individual's safety, health and rights.
 - Flexibility to ensure low risk applications are not inadvertently captured.
 - Interoperability with international definitions and markets.
 - Applicability to General Purpose AI, the risks of which are not defined by specific applications.
- **Options for mandatory guardrail measures** for high-risk systems with a focus on testing, transparency and accountability. Examples of guardrails, based on approaches overseas, include:
 - Conducting risk or impact assessments.
 - Requiring conformity assessments.
 - Ensuring organisational accountability frameworks are in place including clear roles, responsibilities and reporting structures.
 - Implementing human-in-the-loop requirements commensurate with the potential risks.
 - Testing systems before they are deployed.
 - Ensuring that best efforts are applied where digital output is generated so the public can identify the output is AI generated.
 - Informing individuals when they are communicating or interacting with an AI system.
 - Establishing processes for people impacted by AI systems to challenge use or outcomes.
 - Keeping and disclose certain records to assist with compliance.
- **Options for regulatory mechanisms.** Options include strengthening and clarifying existing laws, through to whole of economy measures such as framework legislation or an Australian AI Act (similar to the approach taken in Canada).

Supporting and promoting best practice for safety and adoption

The government is committed to fostering a thriving AI industry in Australia and supporting broad adoption of AI across the economy, complementing efforts to ensure that Australia has the necessary guardrails in place to build trust and confidence in the use of AI.

Around the world, countries are investing in their industrial base, manufacturing capability and economic sovereignty. Canada, the United Kingdom, the Republic of Korea, Singapore and the United States are all investing in AI initiatives because they realise that investing in AI is a strategic, economic and industrial necessity. The ability of Australia to remain competitive will be underscored by the adoption of automation and AI. Considering the rapidly changing global environment, there is a role for government to play alongside industry and civil society to ensure the opportunities of AI are captured for all Australians.

In line with the government's objective to maximise the opportunities that AI presents, it acknowledges the importance of supporting industry to put in place appropriate governance arrangements to use AI safely and responsibly.

Budget 2024-25

In the 2024-25 Budget context, the government provided \$39.9 million over five years from 2023-24 for the development of policy and capability across Government to support the adoption and use of AI technology in a safe and responsible manner. Budget measures relating to the Department of Industry, Science and Resources include:

- The establishment of a permanent AI Advisory Body that will advance the role carried out by the temporary AI Expert Group. The AI Advisory Body will include expertise from civil society, industry and academia and will provide advice to policymakers and regulators on Australia's ongoing response to the opportunities and risks presented by AI.
- Repurposing \$21.6 million to bring the National AI Centre (NAIC) into the Department of Industry, Science and Resources. This support the NAIC's role as the Commonwealth's flagship organisation for enabling industry engagement and driving new models of collaboration among Government, researchers, academics and industry on AI.
- \$15.7 million over two years from 2024–25 to support industry analytical capability and coordination of AI policy development, regulation and engagement activities across government, including to review and strengthen existing regulations in the areas of health care, consumer and copyright law.

The government is supporting businesses through a range of other initiatives like the \$17 million **AI Adopt Program**¹⁰, **Next Generation Graduates Program**¹¹, **National AI Centre's Responsible AI Network, AI Sprint program**, and the existing **Industry Growth Program, Research and Development Tax Incentive** and the **National Reconstruction Fund**.

Voluntary Safety Standard

Recognising the need to support organisations using AI to promote best practice for safety, the government asked the National AI Centre (NAIC) to develop Australia's first **Voluntary AI Safety Standard**. The voluntary standard may have a role in supporting consistent advice to industry for best practice in support of government's wider online safety agenda.

In February 2024, the NAIC, supported by DISR, convened a meeting of leading AI specialists to develop the scope, design principles and subsequently, the core content of the voluntary standard. Roundtables were hosted by Responsible AI Network partners in March 2024 with industry representation including Australian Institute of Company Directors, Australian Information Industry Association, AI Group, Governance Institute of Australia, Committee for Economic Development of Australia, Choice, Tech Council of Australia, Responsible AI Thinktank, and Business Council of Australia. Key insights from the roundtables and early content were tested with a cross section of stakeholders including COSBOA, Centre for Inclusive Design, Social Policy Group and the Diversity Council of Australia.

Related work across Government

As DISR consults to develop a consolidated response to the opportunities and harms presented by AI, we are working across government to ensure related efforts are aligned and harmonised. In addition to the measures identified in the Online Safety Act Issues Paper, this includes:

Copyright reform

DISR engages with the Attorney General's Department (AGD) on the Copyright and AI Reference Group to address future copyright challenges emerging from AI.¹²

¹⁰ Department of Industry, Science and Resources, [\\$17 million to boost AI adoption by SMEs](#), Australian Government, 8 December 2023

¹¹ G Egan, [Building an emerging technology pipeline in regional Australia](#), CSIRO, 24 January 2024

¹² <https://www.ag.gov.au/rights-and-protections/copyright/copyright-and-artificial-intelligence-reference-group-cairg>

Privacy Act reform

DISR engages with the Attorney-General's Department (AGD) on the government's response to the implementation of the Privacy Act reforms, in particular proposals relating to automated decision making and data flows.¹³

Deepfake harms

DISR engages with AGD on proposed legislation to strengthen criminal offences targeting the creation and non-consensual distribution of deepfake sexual material.

Automated Decision Making

DISR engages with AGD on implementation of Robodebt inquiry recommendations relating to legislative reform ensuring government services can operate ethically, without bias and with appropriate safeguards.

Mis and disinformation

DISR engages with DITRDCA on intersections between safe and responsible AI, and proposed powers to combat mis and disinformation.¹⁴

Digital regulation

DISR engages with the Digital Platform Regulators' Forum (DP-REG) on measures to regulate AI alongside other work on digital platforms.¹⁵

Cyber security

DISR engages regularly with the Department of Home Affairs on intersections between AI adoption and regulation, and national cyber security threats from AI and other emerging technologies.¹⁶

Under the 2023-30 Australian Cyber Security Strategy, government is identifying and protecting datasets of national significance and assessing whether existing data protections such as storage and governance settings are proportionate and effective.

Working with states and territories

Government is working with States and Territories to ensure cohesive, safe and transparent use of AI upholding public trust. At the February 2024 Data and Digital Ministers' Meeting, Ministers agreed an initial national framework for the assurance of AI used by governments. This initial framework aligns with the Australian AI Ethics Principles and includes common assurance processes.¹⁷

Leveraging international approaches

Like online harms, AI is a shared challenge, with harms occurring across borders. The same AI systems can be designed, developed and deployed across multiple countries, all of whom are responding to a similar range of risks and challenges. The Australian Government's focus on transparency, testing and accountability measures for AI in high-risk contexts is aligned with international approaches.

Some jurisdictions have already introduced voluntary commitments they expect companies to implement. The United States (US) announced voluntary commitments from a group of 28 healthcare provider and payer organizations to help move toward safe, secure and trustworthy purchasing and use

¹³ <https://www.ag.gov.au/rights-and-protections/publications/government-response-privacy-act-review-report>

¹⁴ <https://www.infrastructure.gov.au/have-your-say/new-acma-powers-combat-misinformation-and-disinformation>

¹⁵ eSafety Commissioner, *Digital Platform Regulators' Forum*, 23 November 2023

¹⁶ Department of Home Affairs, *2023-2030 Australian Cyber Security Strategy*, 21 December 2023

¹⁷ <https://www.finance.gov.au/publications/data-and-digital-ministers-meeting-outcomes/23-february-2024>

OFFICIAL

of AI technology. Singapore introduced standardised self-testing tools ('AI Verify') to enable businesses to check AI models against a set of principles.

Other jurisdictions, including Canada and the European Union (EU), are seeking to make commitments mandatory for higher risk AI systems through new legal frameworks. Both Canada and the EU have also sought voluntary commitments from companies ahead of the enactment and enforcement of these proposed legal frameworks.

This pace of advancements in AI was also a catalyst for the Bletchley Declaration, signed at the AI Safety Summit (November 2023, UK) which was followed by the AI Seoul Summit (May 2024, Korea); and will be progressed at the AI Action Summit (February 2025, France).

Australia recognises we cannot act alone. Responding to AI requires likeminded countries working in close partnership – international engagement is critical to Australia achieving our domestic objectives. Australia is an active player shaping these efforts, seeking to take forward our commitments under the Bletchley Declaration; promote and align regulatory approaches with likeminded countries; shape global AI governance consistent with our values and interests; showcase Australia's AI strengths to boost our local industry; and be a trusted partner on AI in the Indo-Pacific region.

At the AI Safety Summit in 2023, Australia joined the EU and 27 countries in signing the Bletchley Declaration, committing to international collaboration on AI safety testing and the building of risk-based frameworks across countries to ensure AI safety and transparency. Through the Seoul Declaration we joined with other like-minded countries to establish a global network of AI safety research. Australia has been asked to elevate the voices of First Nations people in the final iteration of International Scientific Report on Advanced AI Safety which will be considered at the AI Action Summit.

As part of our comprehensive international engagement on AI, Australia participates in dialogues to shape and influence global standards. The government recently became an observer to Standards Australia's National Mirror Committee relating to AI: IT-043 (Artificial Intelligence). The Committee supports Australia's representatives on the AI standards working groups at international standards settings bodies, such as the International Organisation for Standardisation (ISO) and the International Electrotechnical Commission (IEC).

Australia supports AI standards that align with the values outlined in the Quad Principles on Critical and Emerging Technology Standards to:

- Support industry led, consensus-based multi-stakeholder approaches.
- Support technology standards that promote interoperability, competition, inclusiveness and innovation.
- Foster technology standards that support safety, security and resilience.

Increasing Australian influence on the design and use of technology standards supports the delivery of the government's technology and industry strategies and helps to:

- Create markets for Australia's technological innovations.
- Maximise our economic benefits in technology areas where Australia is investing and has competitive advantages.
- Support our social cohesion by ensuring security, safety, and privacy are embedded in international standards.

Glossary

- Artificial Intelligence (AI) system: a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

OFFICIAL

- General purpose AI (GPAI): (or general AI) a type of AI system that addresses a broad range of tasks and uses, both intended and unintended by developers.
- Generative AI: a branch of AI that develops generative models with the capability of learning to generate content such as images, text, and other media with similar properties as their training data.
- Narrow AI: a type of AI system that is focused on defined tasks and uses to address a specific problem.