THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

CREATE CHANGE

Submission to the Senate Select Committee on Adopting Artificial Intelligence (AI)

Senate Select Committee

20 May 2024

Author:
Dr Caitlin Curtis
Research Fellow, The University of Queensland Centre for Policy Futures and School of Business.

Email: c.curtis@uq.edu.au
Phone: +61 7 3443 3148
UQ Researcher Profile: https://researchers.uq.edu.au/researcher/14369

This submission represents the opinions of the contributing author listed in this document. It does not necessarily represent an official position of The University of Queensland.

Terms of Reference

The committee's terms of reference include:

a. recent trends and opportunities in the development and adoption of AI technologies in Australia and overseas, in particular regarding generative AI;
b. risks and harms arising from the adoption of AI technologies, including bias, discrimination and error;
c. emerging international approaches to mitigating AI risks;
d. opportunities to adopt AI in ways that benefit citizens, the environment and/or economic growth, for example in health and climate management;
e. opportunities to foster a responsible AI industry in Australia;
f. potential threats to democracy and trust in institutions from generative AI; and
g. environmental impacts of AI technologies and opportunities for limiting and mitigating impacts.

Background and recommendations

Points a – f: Opportunities and threats, and governance approaches

Background: benefits, risks and governance of AI

Artificial Intelligence (AI) offers transformative potential across various sectors, with the potential to enhance efficiency, productivity, advance healthcare, and enable innovative solutions to complex problems. AI-driven diagnostics can lead to early disease detection, and intelligent automation can streamline industrial processes, resulting in significant economic benefits. However, these advantages come with substantial risks, including job displacement, privacy concerns, the potential for unfair treatment and discrimination, deep fakes and misinformation, and all these can cause potential harms to individuals, groups, and Australian and global society. The misuse of AI can lead to biases in decision-making processes, exacerbating social inequalities and potentially infringing on individual rights.

Globally, different approaches to AI governance highlight the need for balanced regulation. In the United States, the AI Bill of Rights aims to protect individuals from AI-induced harms by ensuring transparency, privacy, and fairness in AI applications ("Blueprint for an AI Bill of Rights" 2022) (See Fig 1), with some overlap with Australia's AI Framework (2019) (See Fig 2). Similarly, the European Union's AI Act seeks to classify AI systems based on risk and enforce stringent requirements for high-risk applications ("Artificial Intelligence Act: MEPs Adopt Landmark Law" 2024). In addition, the Council of Europe has adopted the first-ever international legally binding treaty aimed at ensuring the respect of human rights, democracy, the rule of law and legal standards in the use of artificial intelligence (AI) systems (Council of Europe 2024). China, on the other hand, emphasizes strict state control and security measures in its AI regulations. These diverse strategies underscore the importance of crafting governance frameworks that safeguard human rights while fostering innovation.

Earlier this week, youth activists issued a broad call to action (May 16th, 2024) demanding stronger protections from the impacts of artificial intelligence (*The Washington Post* 2024). They urged global leaders to take swift measures to harness AI's potential while mitigating its dangers, emphasizing the significant impact these tools could have on their generation – with specific reference to some of the potential harms from generative AI ("AI 2030" 2024).

It is crucial to ensure that AI systems don't exacerbate existing inequalities during their development and implementation. Nearly 1 in 10 Australians (9.4%) are classified as 'highly excluded' from digital technologies, according to the Australian Digital Inclusion Index of 2023 ("Australian Digital Inclusion Index" 2023), many of which belong to vulnerable groups.

Background: public trust and expectations about AI governance

Concerns about risks posed by AI systems (Curtis, Gillespie, and Lockey 2023) have also permeated public consciousness: 3 out of 4 people surveyed across 17 countries (incl. Australia) expressed concern about potential AI risks, including security risks and job loss due to automation (Gillespie et al. 2023). Most participants (73%) believe in the need for some form of AI regulation, yet only 39% believe current safeguards are enough to make it safe (Gillespie et al. 2023). We examined how attitudes towards AI have evolved since 2020 in Australia, the UK, USA, Canada, and Germany. In all these countries, trust in AI and awareness of its common applications have increased. However, perceptions of the adequacy of regulations, laws, and safeguards to protect against AI risks, as well as confidence in entities to develop, use, and govern AI, have not changed. This is not surprising given that not a week goes by without significant developments and new areas of applications with potentially far-reaching consequences and profound impact.

Recommendations

A dual approach: Creating a human rights framework alongside regulations to keep AI use safe.

Stronger governance of AI is recommended to make the use of AI safe, fair, and equitable: to protect the public and meet public expectations (Gillespie et al. 2023; Curtis, Gillespie, and Lockey 2023). Ensuring that AI use is safe, to meet public expectations, will involve the creation of regulations and safeguards and/or the application or strengthening of existing regulations and safeguards.

It is also important to articulate what our rights and expectations are, as Australian citizens, with respect to AI. This could take the form of a human rights approach. Australia's AI framework can be a guide, although some additional rights have been suggested in the following section. These rights do not replace regulation, but rather they complement and guide it. We need laws articulating what people cannot do, to protect the public from harms and to keep AI use safe. The public has repeatedly indicated in our surveys that they do not feel that there are enough laws and regulations in place to make AI use safe in Australia. However, regulations may lead to a fractured approach, and may not be sufficient.

This dual approach could create a cohesive, human-centred framework to establish and articulate public expectations and rights with respect to AI systems. As a model, the U.S. AI Bill of Rights outlines key principles aimed at protecting individuals from potential harms associated with AI technologies.

I propose a dual approach, which involves regulations and safeguards, but also articulating what our rights are, in consultation with the public. Outlining our rights alongside the regulations will support a more holistic approach to ensure that we can adopt AI in ways that benefit citizens (Terms of Reference point D), ensure public trust, and foster a responsible AI industry in Australia (Terms of Reference point E).

Stating our rights in relation to AI in a proactive manner could also support Australian citizens in interacting safely and fairly with AI systems. Even the act of creating and naming a right can significantly influence people's ability to advocate for its provision.

This would establish a framework that balances innovation in AI with essential protections for individuals, ensuring AI technologies serve the public good while safeguarding fundamental rights and freedoms. Australia also has developed its voluntary and non-binding AI Ethics framework (Department of Industry Science and Resources 2019 (See Fig 2). The principles are voluntary and intended to be aspirational and complement – not substitute – existing regulations and practices relevant to AI. This AI Ethics framework, along with the U.S. AI Bill of Rights (2022), could be drawn upon for an outline of Australian AI Human Rights.

Right	Summary
Right to be Protected from Unsafe or Ineffective Systems	You should be protected from unsafe or ineffective systems.
Right to Privacy	You should be protected from abusive data practices via built-in protections and you should have agency over how data about you is used.
Right to Notice and Explanation	You should know that an automated system is being used and understand how and why it contributes to outcomes that impact you.
Right to Avoid Discrimination	You should not face discrimination by algorithms and systems should be used and designed in an equitable way.
Right to Human Alternatives, Consideration, and Fallback	You should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems you encounter.

Figure 1: Rights in the U.S. Blueprint for an AI Bill of Rights (figure drawn from U.S. Blueprint for an AI Bill of Rights).

Principle	Summary
Privacy Protection and Security	AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.
Reliability and Safety	AI systems should reliably operate in the context of their intended purpose throughout their lifecycle.
Transparency and Explainability	There should be transparency and responsible disclosure to ensure people know when they are being significantly impacted by an AI system, and can find out when an AI system is engaging with them. Explainability includes what the AI system is doing and why, and may include the system's processes and input data.
Fairness	AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals or groups.
Contestability	When an AI system significantly impacts a person, group or environment, there should be a timely process to allow people to challenge the use or output of the system.
Accountability	Those responsible for the various phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the system, and human oversight of AI systems should be enabled.
Human-centred Values	AI systems should respect human rights, diversity, and the autonomy of individuals.
Human, Social and Environmental Wellbeing	AI systems should benefit individuals, society, and the environment.

Figure 2: Principles of Australia's AI Ethics framework. Figure created from Australia's AI Ethics Framework (2019).

Potential rights and protections to consider with AI and generative AI.

Australian AI human rights could be articulated from Australia's AI Ethics framework and drawn from rights outlined in the U.S. Blueprint for AI Bill of Rights. They could also include specific rights such as the following:

I. **A right to a freedom from undue surveillance enabled by AI systems, including facial recognition systems.**

II. **A right to be free from manipulation by AI systems.**

III. **A "Right to Our Own Image and Self". [To protect individuals from the misuse of their likeness through technologies like deepfakes. This right would ensure individuals have control over how their image and personal attributes are used.]**. Calls for these types of rights and protections were also included in the Encode Justice Youth Platform ("AI 2030" 2024).

Elements that might be included under this right could include:

a. Right to Control One's Likeness: *Individuals should have the authority to consent to or refuse the use of their likeness in any form of AI-generated content, including deepfakes. This could include privacy protections to prevent unauthorized use and manipulation of an individual's image, voice, and other personal attributes by AI technologies. This could also require explicit notification and consent from individuals before their likeness can be used in AI-generated media, ensuring they are fully informed and agree to such use. Ideally this could help protect the integrity of an individual's likeness and identity, ensuring that AI technologies cannot be used to impersonate or misrepresent someone without their permission.*

b. Right to Redress, Removal and Legal Recourse: *Establish legal frameworks that allow individuals to take legal action against those who create and distribute deepfakes or other AI-generated content without their consent. Provide mechanisms for individuals to seek redress and demand the removal of unauthorized or harmful AI-generated content that uses their likeness.*

IV. A right to trustworthy information or a "Right to Not Be Deceived" or "Right to Reality" [PUBLIC EXPECTATIONS AROUND AI GENERATED CONTENT, INCLUDING DEEP FAKES, MISINFORMATION]

Articulating a right to trustworthy information particularly in the context of AI-generated content, deepfakes, and virtual reality. This would ensure that individuals can trust the authenticity of the information and experiences they encounter. Efforts to require watermarking of images are already underway, but this may be insufficient, as incidents of deceptive generative AI content have already extended outside the traditional screen (e.g. President Joe Biden 'robocalls' (Seitz-Wald 2024; Malleck 2024). This right would extend outside of the digital screen.

By incorporating a "right to trustworthy information," AI governance can address the growing challenges posed by AI's ability to blur the lines between reality and fiction, ensuring that individuals can trust the information and experiences they encounter in an increasingly digital world.

Elements that might be included under a right to trustworthy information could include some or all of the following:

a. Right to Authentic/Trustworthy Information and right to Identify AI-Generated Content: *Ensure that AI systems do not produce or disseminate false or misleading information, particularly in news, media, and public communications. Require clear labelling of AI-generated content, including text, images, videos, and audio, so individuals can distinguish between real and synthetic media.*

b. Right to Protection from Deepfakes: *Implement strict regulations against the creation and distribution of malicious deepfakes that can harm individuals' reputations, privacy, and safety.*

c. Right to Authentic Experiences: *Safeguard against the manipulation of virtual or augmented reality experiences in ways that could deceive, manipulate, or harm individuals, including emotional manipulation (including by generative AI chatbots and assistants). This is particularly important as AI system interfaces become more 'lifelike' (e.g. Replika AI, ChatGPT 4.o), and potentially increased risk for people to anthropomorphise them, emotionally bond or feel that the AI system 'cares' about them in a human way.*

V. Rights in the Context of AI and Work

Addressing how AI impacts work is crucial in AI governance. A comprehensive framework should include rights to protect workers from potential negative effects while promoting equitable benefits from AI advancements.

a. Right to Fair Employment, retaining and education: *Ensure that AI-driven automation and decision-making do not result in unfair job displacement or discriminatory hiring practices. Provide opportunities for workers to receive retraining and upskilling programs to adapt to changes brought by AI technologies in their industries.*

b. Right to Transparency in AI-Driven Decisions, and non-discrimination: *Guarantee that workers have access to explanations of how AI systems make decisions affecting their employment, such as hiring, promotions, and performance evaluations. Implement measures to prevent bias in AI systems that could lead to discriminatory practices in recruitment, pay, or job assignments.*

c. Right to Work-Life Balance: *Ensure that AI technologies are designed and implemented in ways that support, rather than undermine, workers' work-life balance.*

References

"AI 2030." 2024. AI 2030 | Encode Justice. March 30, 2024. https://ai2030.encodejustice.org/.

"Artificial Intelligence Act: MEPs Adopt Landmark Law." 2024. March 13, 2024. https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law.

"Australian Digital Inclusion Index." 2023. 2023. https://www.digitalinclusionindex.org.au/key-findings-and-next-steps/.

"Blueprint for an AI Bill of Rights." 2022. The White House. October 4, 2022. https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

Council of Europe. 2024. "Council of Europe Adopts First International Treaty on Artificial Intelligence." Council of Europe. May 17, 2024. https://www.coe.int/en/web/portal/-/council-of-europe-adopts-first-international-treaty-on-artificial-intelligence.

Curtis, Caitlin, Nicole Gillespie, and Steven Lockey. 2023. "AI-Deploying Organizations Are Key to Addressing 'perfect Storm' of AI Risks." *AI and Ethics* 3: 145–53.

Department of Industry Science, and Resources. 2019. "Australia's AI Ethics Principles." October 5, 2022. https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles.

Gillespie, Nicole, Steve Lockey, Caitlin Curtis, Javad Pool, and Ali Akbari. 2023. "Trust in AI: A Global Survey." 2023. https://doi.org/10.14264/00d3c94.

Malleck, Julie. 2024. "Why 'Right to Reality' Has Become a Rallying Cry in the Age of AI." *Digital Frontier*, May 8, 2024. https://digitalfrontier.com/articles/right-to-reality-ai-misinformation-deepfakes.

Seitz-Wald, Alex. 2024. "Democratic Operative Admits to Commissioning Fake Biden Robocall That Used AI." NBC News. February 25, 2024. https://www.nbcnews.com/politics/2024-election/democratic-operative-admits-commissioning-fake-biden-robocall-used-ai-rcna140402.

The Washington Post. 2024. "Youth Activists Call on World Leaders to Set AI Safeguards by 2030," May 16, 2024. https://www.washingtonpost.com/politics/2024/05/16/youth-activists-call-world-leaders-set-ai-safeguards-by-2030/.