Adopting AI in Australia

Australian Human Rights Commission

Submission to Select Committee on Adopting Artificial Intelligence

15 May 2024

1   Introduction

1. The Australian Human Rights Commission (Commission) welcomes the opportunity to make this submission to the Select Committee on Adopting Artificial Intelligence (Committee) in response to the inquiry into the opportunities and impacts for Australia arising out of the uptake of artificial intelligence (AI) technologies.

2. The role of the Commission is to work towards a world in which human rights are respected, protected and fulfilled. The Commission has expertise on the intersection between technologies and human rights.

3. The Commission has previously provided submissions on the human rights impacts of AI, including:

   • National inquiry into AI regulation submission

   • Global Digital Compact submission.

4. The Commission welcomes further opportunities to engage with the Committee.

2   Definitions

2.1   Misinformation and disinformation

5. Throughout this submission we have adopted the same definitions for these terms as provided by the Electoral Integrity Assurance Taskforce, namely:

   • 'Misinformation' is false information that is spread due to ignorance, or by error or mistake, without the intent to deceive.[1]

   • 'Disinformation' is knowingly false information designed to deliberately mislead and influence public opinion or obscure the truth for malicious or deceptive purposes.[2]

2.2   Deepfakes

6. This submission defines 'deepfakes' as referring to:

A digital photo, video or sound file of a real person that has been edited to create an extremely realistic but false depiction of them doing or saying something that they did not actually do or say.[3]

3 Overseas approaches to AI

7. There is much discourse in Australia, and overseas, about how best to regulate AI to mitigate harms. In its [submission](#) to the Department of Industry, Science and Resources in response to the [Safe and Responsible AI in Australia: Discussion Paper,](#) the Commission calls for urgent action to address these risks and ensure human rights-centred design in the deployment of new and emerging technologies (including AI).

8. The Commission welcomes the review of the existing regulatory framework to determine where reform is most needed, and where existing legislation may be adapted to respond to the emergence of AI. However, it is possible that gaps may be identified which do not fall within the remit of existing legislation. Where such challenges are identified, Australia should adopt AI-specific legislation (provided it does not lead to unnecessary duplication).

9. This approach would strengthen Australia's existing technology-neutral approach to regulation while considering experiences in relation to the European Union's AI Act.

    Recommendation 1: Australian should strengthen existing legislation and then introduce AI-specific legislation, if necessary, to address risks that are not currently within the scope of the existing regulatory framework.

4 Harms of AI

10. Unchecked proliferation of powerful AI technologies may radically reshape economies and societies over the coming decade – for both better and for worse.[4] As noted by the United Nations General Assembly:

    > The improper or malicious design, development, deployment and use of artificial intelligence systems, such as without adequate safeguards or in a manner inconsistent with international law … [could] undermine sustainable development in its three dimensions – economic, social and environmental; widen digital divides between and within countries; reinforce structural inequalities and biases; lead to discrimination; undermine information integrity and access to information; undercut the protection, promotion and enjoyment of human rights and fundamental freedoms, including the right

not to be subject to unlawful or arbitrary interference with one's privacy; and increase the potential risk for accidents and compound threats from malicious actors.[5]

11. The General Assembly resolution goes on to emphasise that 'human rights and fundamental freedoms must be respected, protected and promoted throughout the life cycle of artificial intelligence systems'.[6] As Australia continues its work to develop appropriate regulatory and policy responses to AI, the protection of human rights should be a central priority.

Recommendation 2: Australia should adopt a human rights-centred approach to AI development and deployment.

4.1   Bias and algorithmic discrimination

12. One of the advantages of AI is to process and analyse vast amounts of information quickly. However, the use of these tools can produce biased and unfair outcomes, otherwise known as algorithmic bias.[7]

13. Despite a perception that AI-based decision-making is free from human intervention and thus objective in nature, algorithmic bias can perpetuate existing biases, thus replicating human errors and entrenching issues such as unlawful discrimination.[8]

14. For example, a 2019 US study found that a clinical program used in hospitals to determine which patients required further medical attention produced racial bias. Black patients needed to exhibit significantly more severe symptoms than their white counterparts to be recommended for the same care. This happened as a result of the clinical algorithm being trained on past data on healthcare spending which reflected a history of black patients having less income to spend on their health – an outcome of longstanding wealth and income disparities.[9]

15. Ultimately, such risks emphasise the need for greater regulation of AI to limit the effects of algorithmic bias, especially in cases of decision-making which have legal or similarly significant implications for individuals. This need is exacerbated by the difficulty of applying anti-discrimination laws to existing AI systems, and issues when determining liability when unlawful discrimination occurs.[10]

16. The Commission's 2020 technical paper, [Using artificial intelligence to make decisions: Addressing the problem of algorithmic bias](#), considers algorithmic bias in greater detail.

4.2　Automation bias

17. Algorithmic decision-making may also impede on independent decision-making due to tendency to over rely on the outcomes produced by AI. This overreliance is known as 'automation bias', which is the:

> tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing.[11]

18. Automation bias can have consequences for individuals from a consumer to government level. One example of automation bias occurring in practice can be seen in stories about individuals driving their cars into the ocean while following GPS systems like Google Maps.[12]

19. Automation bias may have also been a factor in the harms caused by the 'robodebt' scheme which involved overreliance on automated data matching to calculate Centrelink debt.

20. When integrating AI into both government and private sector business models and service, the risk of both automation and algorithmic bias should be mitigated. The establishment of a national AI Commissioner, an independent statutory body tasked with assisting the broad adoption of AI in Australia, would support organisations in their efforts to mitigate the risks associated with these biases.

Recommendation 3: The Australian Government establish a national AI Commissioner as an independent statutory office.

5　AI-generated misinformation and disinformation

21. AI-generated misinformation and disinformation may harm several human rights. The Commission has previously emphasised that misinformation and disinformation can have devastating effects on human rights, social cohesion and democratic processes. Indeed, this can be the very purpose intended by the release of disinformation.[13]

22. While there is a clear need to combat misinformation and disinformation, there is also a real risk of different perspectives and opinions being targeted when doing so. Robust safeguards for freedom of expression must form part of any measures taken to combat misinformation and disinformation in order to ensure that Australia's democratic values are not undermined.[14]

23. Misinformation and disinformation can pose a particular threat to democracy and trust in institutions. Some recent examples that highlight the potential for synthetic content to undermine democratic processes include:

    - A deepfake video purporting to show the President of Ukraine, Volodymyr Zelenskyy, calling on Ukrainian troops to lay down their arms and surrender circulated on social media and was briefly placed on a Ukrainian news website in March 2022 before being removed.[15]

    - A video of Muhammad Basharat Raja, a Pakistani politician and candidate, that was altered to tell voters they should boycott the 2024 general election started circulating on Facebook the day before the elections.[16]

    - More than 100 deepfake video advertisements impersonating the Prime Minister of the United Kingdom, Rishi Sunak, that were paid to be promoted on Meta between December 2023 – January 2024.[17]

24. In 2024, the World Economic Forum declared that misinformation and disinformation would be the 'most severe global risk anticipated over the next two years'.[18] The Organisation for Economic Co-operation and Development (OECD) has also stressed that the use of Large Language Models (LLMs) to spread disinformation can damage public trust in democratic institutions.[19]

25. The risk profile of synthetic content is exacerbated by its ease of use and general efficiency.[20] AI models can be used to generate cheap, persuasive and personalised content for harmful purposes.[21]

26. It is also becoming increasingly difficult to identify synthetic content. For example, in February 2024 Open AI announced Sora AI – an 'AI model that can create realistic and imaginative scenes from text instructions'.[22] The Sora AI demonstration shows simplistic prompts generating hyper-realistic footage that appears to be genuine.

27. Sora AI is not yet available to the public, with Open AI saying it is 'taking several important safety steps', including working with experts in 'misinformation, hateful content, and bias' and 'building tools to help detect misleading content'.[23]

28. Synthetic content is capable of generating images, videos and audios of people doing or saying anything. These deepfakes can have real consequences. For example, a deepfake image of an explosion at the Pentagon resulted in the Dow Jones Industrial Index dropping 85 points (0.3 per cent) in four minutes.[24]

29. In Australia, Nine News recently [claimed](#) that an automation tool in Photoshop led to the cropping of clothing and enlarging of breasts of a photo of Animal Justice Party MP Georgie Purcell.

30. There is growing concern about the harm caused by sexualised deepfakes, with Sensity AI finding that 90–95% of online deepfake videos are non-consensual pornography.[25]

31. There have been a number of examples of legislative responses to the problems posed by deepfakes being introduced. Examples include the *Deepfakes Accountability Act* being introduced in the US House of Representatives in 2019 (which required deepfakes to be digitally watermarked) and South Korea banning political content which involves deepfakes within 90 days before an election.[26]

32. The Australian Government is also considering voluntary watermarking.[27] Given the increasing evidence of the harms that can be caused by deepfakes, the Commission would recommend that digital watermarks for AI-generated content should be adopted as a priority. Additional consideration should also be given to how synthetic content can affect elections and the role that watermarking could play in mitigating any adverse impacts.

33. The Australian Government's interim response on AI has also indicated a risk-based approach where 'high-risk' uses of AI will be subject to mandatory safeguards.[28] Notably the EU (which has also adopted a risks-based approach) has not included deepfakes as a high-risk usage of AI. While it is unclear what constitutes 'high-risk' according to the Australian Government, it should include synthetic content such as deepfakes – particularly when used in pornography or to influence democratic processes.

5.1   Foreign interference

34. While social media has many positives, these platforms are increasingly being used in foreign interference operations to disseminate misinformation and disinformation.[29] Interference during elections and referendums have increased significantly in the online environment in recent years.[30] As noted in the Commission's 2023 [submission to the Senate Select commission on Foreign Interference through Social Media](#), the rise of AI-generated misinformation and disinformation will likely increase in prevalence during democratic processes.

35. A key risk is that synthetic content may be disseminated on social media platforms in a persuasive and authoritative manner. This often leads it to be viewed as legitimate news by end users.

36. The number of people using social media for news is only growing. Compared to last year, Gen Z users are increasingly using social media as a new source:

    - 23% use YouTube (up 4%)

    - 10% use WhatsApp (up 4%)

    - 26% use Instagram (up 6%)

    - 17% use TikTok (up 4%).[31]

37. In consequence, there is a greater risk that synthetic content produced as part of foreign interference operations may be accepted as trustworthy or reliable by users.

38. Propaganda and disinformation will likely increase as deepfakes lower the cost to entry, while also expanding the reach of content shared online.[32] In recent times propaganda has been generated by individuals in places such as China's '50-centres' and Russia's 'troll farms'.[33] However AI-generated content will likely increase the prevalence and reach of misinformation and disinformation.

39. Foreign actors are able to use social media to skew public debate and undermine trust in democratic processes. The rise of AI only makes this process cheaper and more efficient.[34] More needs to be done to address foreign interference utilising synthetic misinformation disinformation. The Commission would echo recommendation 11 of the Senate Select Committee on Foreign Interference through Social Media's Final Report.[35]

Recommendation 4: The Australian Government investigate options to identify, prevent and disrupt AI-generated disinformation and foreign interference campaigns.

40. There is still a clear need for further research about the prevalence and impact of synthetic content spread by foreign actors and how that adversely impacts Australian democracy. Accordingly, the Commission also emphasises recommendation 13 of the Senate Select Committee on Foreign Interference through Social Media's Final Report.[36]

Recommendation 5: The Australian Government build capacity to counter social media interference campaigns by supporting independent research.

6   Environmental impact

41. The international community is increasingly recognising the human right to a healthy environment. The first formal recognition at a global level was by the UN Human Rights Council in October 2021[37] and has continued with the adoption of Resolution A/76/L.75 by the UN General Assembly in July 2022.

42. AI has the potential to have a positive impact on the environment in a range of ways, including by improving energy efficiency and enhancing sustainable practices.[38]

43. However, AI also poses significant environmental risks, particularly due to the large amount of computational power and energy involved in developing and training an AI model.[39]

44. A broad societal uptake of AI, especially LLMs, poses a danger that the environment will be negatively impacted through the additional consumption of electricity (provided the necessary energy is not sourced from renewables).

45. In 2019, researchers from the University of Massachusetts Amherst estimated that the carbon footprint of training a single LLM equals around 300,000 kg of carbon dioxide emissions, or 125 round trip flights between New York and Beijing.[40]

46. By increasing transparency around the potential environmental impacts of AI, the risks may be mitigated. Initiatives such as 'FAIR Forward – Artificial Intelligence for All' enables the sharing of knowledge and environmental best practices, and is a valuable tool in developing AI with environmental impacts in mind.[41]

47. Other researchers have suggested that AI research should include mandatory reporting on the computational costs of training algorithms.[42] However, reporting is also needed on the environmental cost of updating and regularly using AI models.[43]

48. To best understand the risks that AI poses to climate change the Australian Government should act upon previous calls to work with the private sector to better understand and mitigate the climate impact of AI.[44]

Recommendation 6: The Australian Government increase its engagement with the private sector to better understand and mitigate the climate impact of AI.

7 Recommendations

49. The Commission makes the following recommendations.

Recommendation 1

Australian should strengthen existing legislation and then introduce AI-specific legislation, if necessary, to address risks that are not currently within the scope of the existing regulatory framework.

Recommendation 2

Australia should adopt a human rights-centred approach to AI development and deployment.

Recommendation 3

The Australian Government establish a national AI Commissioner as an independent statutory office.

Recommendation 4

The Australian Government investigate options to identify, prevent and disrupt AI-generated disinformation and foreign interference campaigns.

Recommendation 5

The Australian Government build capacity to counter social media interference campaigns by supporting independent research.

Recommendation 6

The Australian Government increase its engagement with the private sector to better understand and mitigate the climate impact of AI.

Endnotes

[1] Electoral Integrity Assurance Taskforce, *Disinformation and Misinformation Factsheet* <eiat-disinformation-factsheet.pdf (aec.gov.au)>.

[2] Electoral Integrity Assurance Taskforce, *Disinformation and Misinformation Factsheet* <eiat-disinformation-factsheet.pdf (aec.gov.au)>.

[3] eSafety Commissioner, '*Deepfake trends and challenges — position statement*' (website) <https://www.esafety.gov.au/industry/tech-trends-and-challenges/deepfakes>.

[4] World Economic Forum, '*The Global Risks Report 2024*' (Report, January 2024) 50.

[5] United Nations General Assembly, Resolution A/78/L.49 2-3.

[6] United Nations General Assembly, Resolution A/78/L.49 4-5.

[7] Australian Human Rights Commission ('AHRC'), '*Final Report on Technology*' (Report, 2021) 13.

[8] AHRC, '*Final Report on Technology*' (Report, 2021) 13.

[9] Crystal Grant, 'Algorithms are Making Decisions About Health Care, Which May Only Worsen Medical Racism' *American Civil Liberties Union* (Article, 3 October 2022) <https://www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism>.

[10] AHRC, '*Final Report on Technology*' (Report, 2021) 108.

[11] Max Schemmer, et al., '*On the Influence of Explainable AI on Automation Bias*' (Working Paper, 2022) 1 quoting Kathleen Mosier & Linda Skitka, 'Automation Use and Automation Bias' in Proceedings of the Human Factors and Ergonomics Society Annual Meeting (1999) 43(3) *SAGE Publications* 344–348.

[12] See e.g. Hilary Hanson, 'GPS Leads Japanese Tourists To Drive Into Australian Bay' *Huffpost* (Article, 19 March 2012) <https://www.huffingtonpost.co.uk/entry/gps-tourists-australia_n_1363823>.

[13] Australian Human Rights Commission, *Inquiry into the risk posed to Australia's democracy by foreign interference through social media* (Submission to the Senate Select Committee on Foreign Interference through Social Media, 16 February 2023), [16] <https://humanrights.gov.au/our-work/legal/submission/foreign-interference-through-social-media>; Australian Human Rights Commission, *Finding balance: combatting misinformation and disinformation without threatening free expression* (Submission to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts, 18 August 2023), [5]

<https://humanrights.gov.au/our-work/legal/submission/finding-balance-fighting-disinformation-without-threatening-free>.

[14] Australian Human Rights Commission, *Finding balance: combatting misinformation and disinformation without threatening free expression* (Submission to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts, 18 August 2023), [7].

[15] John Joseph Twomey, Conor Linehan and Gillian Murphy, 'Deepfakes in warfare: new concerns emerge from their use around the Russian invasion of Ukraine', *The Conversation* (27 October 2023) <https://theconversation.com/deepfakes-in-warfare-new-concerns-emerge-from-their-use-around-the-russian-invasion-of-ukraine-216393>.

[16] Sameen Aziz, *Video of PTI's Basharat Raja boycotting elections is a deepfake* (Soch Fact Check, 8 February 2024) <https://www.sochfactcheck.com/video-of-ptis-basharat-raja-boycotting-general-elections-2024-is-a-deepfake/>.

[17] Fenimore Harper Communications, *Report: Over 100 Deep-Faked Rishi Sunak Ads Found on Meta's Advertising Platform* (2024) <https://www.fenimoreharper.com/research/deepfake-political-advertising>.

[18] World Economic Forum, '*The Global Risks Report 2024*' (Report, January 2024) 8.

[19] Organisation for Economic Co-operation and Development, '*AI Language Models Technological, socio-economic and policy considerations*' (OECD Digital Economy Papers, April 2023) 10.

[20] World Economic Forum, '*The Global Risks Report 2024*' (Report, January 2024) 18.

[21] G. Bell et al., '*Rapid Response Information Report: Generative AI - language models (LLMs) and multimodal foundation models (MFMs)*' (Australian Council of Learned Academies, Report, 24 March 2023) 12.

[22] Open AI '*Open AI*' (webpage) <https://openai.com/sora>.

[23] Open AI '*Open AI*' (webpage) <https://openai.com/sora>.

[24] Paul Smith, 'Deepfakes spell deep trouble for markets' *The Australian Financial Review* (Article, 23 March 2023) <https://www.afr.com/technology/deepfakes-spell-deep-trouble-for-markets-20230523-p5daih>.

[25] Anastasia Powell, Adrian J. Scott et al, 'Whether of politicians, pop stars or teenage girls, sexualised deepfakes are on the rise. They hold a mirror to our sexist world', *The Conversation* (8 February 2024). <https://theconversation.com/whether-of-politicians-pop-stars-or-teenage-girls-sexualised-deepfakes-are-on-the-rise-they-hold-a-mirror-to-our-sexist-world-222491>.

[26] OECD AI, '90-day ban on deepfake political ads passes parliamentary special committee' *Policy Observer* (online, 05 December 2023) <https://oecd.ai/en/incidents/55571>.

[27] See e.g. Department of Industry, Science and Resources, '*Safe and Responsible AI in Australian Consultation: Australian Government's Interim Response*' (interim response, 17 January 2024) 20.

[28] See e.g. Department of Industry, Science and Resources, '*Safe and Responsible AI in Australian Consultation: Australian Government's Interim Response*' (interim response, 17 January 2024) 20.

[29] See generally Senate Select Committee on Foreign Interference through Social Media, '*Final Report*' (report, August 2023).

[30] Sarah O'Connor, et al., *Cyber-enabled Foreign Interference in Elections and Referendums* (ASPI, 2020) 6.

[31] News and Media Research Centre, '*Digital News Report: Australia 2023*' (report, June 2023) 11.

[32] Hannah Smith & Katherine Mansted, '*Weaponised Deep Fakes*' (ASPI, Report No. 28, April 2020) 11-12.

[33] Hannah Smith & Katherine Mansted, '*Weaponised Deep Fakes*' (ASPI, Report No. 28, April 2020) 11-12.

[34] Senate Select Committee on Foreign Interference through Social Media, '*Final Report*' (report, August 2023) 11.

[35] Senate Select Committee on Foreign Interference through Social Media, '*Final Report*' (report, August 2023) 17.

[36] Senate Select Committee on Foreign Interference through Social Media, '*Final Report*' (report, August 2023) 17.

[37] United Nations Human Rights Council, 48th sess, UN Doc A/HRC/RES/48/13 (18 October 2021).

[38] Rita Li, 'The Environmental Impact of AI', *GRC Insights* (Blog Post, 8 May 2023) <https://insights.grcglobalgroup.com/the-environmental-impact-of-ai/#:~:text=Training%20an%20AI%20model%20can,to%20exacerbate%20existing%20environmental%20problems.>.

[39] Rita Li, 'The Environmental Impact of AI', *GRC Insights* (Blog Post, 8 May 2023) <https://insights.grcglobalgroup.com/the-environmental-impact-of-ai/#:~:text=Training%20an%20AI%20model%20can,to%20exacerbate%20existing%20environmental%20problems.>.

[40] Payal Dhar, 'The carbon impact of artificial intelligence' (2020) 2 *Nature Machine Intelligence* 423.

[41] Organisation for Economic Co-operation and Development (OECD) 'Measuring the environmental impacts of artificial intelligence compute and applications: the AI footprint' (2022) 341 *OECD Digital Economy Papers* 7.

[42] Annette Ekin, 'AI can help us fight climate change. But it has an energy problem, too', *Horizon the EU Research & Innovation Magazine* (Article, 12 September 2019) <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/ai-can-help-us-fight-climate-change-it-has-energy-problem-too>.

[43] Kate Saenko, 'Is generative AI bad for the environment? A computer scientist explains the carbon footprint of ChatGPT and its cousins', *The Conversation* (Blog Post, 23 May 2023) <https://theconversation.com/is-generative-ai-bad-for-the-environment-a-computer-scientist-explains-the-carbon-footprint-of-chatgpt-and-its-cousins-204096>.

[44] See especially Gordon Noble, Alison Atherton & Fiona Berry, '*IT and Data Centre Sustainability in Australia*' (University of Technology Sydney, July 2023). <https://www.uts.edu.au/sites/default/files/2023-07/Pure%20Storage%20and%20Institute%20of%20Sustainable%20Futures%20Report.pdf>.