Australian Human Rights Commission

Thematic Report on Artificial Intelligence (AI) and Racial Discrimination

Australian Human Rights Commission

Submission to Special Rapporteur on Contemporary Forms of Racism

31 March 2024

ABN 47 996 232 602
Level 3, 175 Pitt Street, Sydney NSW 2000
GPO Box 5218, Sydney NSW 2001
General enquiries 1300 369 711
National Information Service 1300 656 419
TTY 1800 620 241

Australian Human Rights Commission
www.humanrights.gov.au

ABN 47 996 232 602
Level 3, 175 Pitt Street, Sydney NSW 2000
GPO Box 5218, Sydney NSW 2001
General enquiries 1300 369 711
National Information Service 1300 656 419
TTY 1800 620 241

1 Introduction

1. The Australian Human Rights Commission (Commission) welcomes the opportunity to make this submission to the Special Rapporteur on Contemporary Forms of Racism (Special Rapporteur) in response to the call for input on artificial intelligence (AI) and racial discrimination.
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 to the United Nations (UN) on the human rights impacts of AI, including:
   - Global Digital Compact submission; and
   - UN High-Level Advisory Body on AI submission.
4. Equally the Commission has provided numerous UN-level submissions on racial discrimination. Most recently, the Commission provided input to the Committee on the Elimination of Racial Discrimination on General Recommendation No. 37 (2023) on racial discrimination in the enjoyment of the right to health.
5. The Commission welcomes further opportunities to engage with the Special Rapporteur.

2 AI impact on human rights

6. Recent developments in AI have affected a broad range of civil and political rights, and economic, social and cultural rights.[1] For example, the previous Special Rapporteur on Contemporary Forms of Racism noted the growing use of digital technologies, including AI, to:

   > determine everyday outcomes in employment, education, health care and criminal justice, which introduces the risk of systemised discrimination on an unprecedented scale.[2]

7. Examples of how AI can negatively affect human rights are commonplace. High-profile examples include AI tools producing racist statements,[3] and, at the extreme end, a facial-recognition application that mislabelled some people with dark skin as 'gorillas'.[4] The rise of generative AI, has created additional human rights concerns.[5]

2.1 Algorithmic bias

8. Algorithmic bias arises where an AI product produces an output that results in unfairness.[6] Algorithmic bias can entrench unfairness, or even result in unlawful discrimination.[7]
9. Algorithmic bias arises most often due to:
    - problems with the training data
    - the AI system or tool itself
    - existing prejudice and bias built in by human designers and deployers.[8]
10. Such flaws can arise, and lead to algorithmic bias, for a variety of reasons. Examples include where the AI-informed decision-making system is designed in a way that:
    - gives undue weight to a particular data set
    - relies on a data set that is incomplete, out of date or incorrect
    - uses a data set that is affected by selection bias – that is, where the data set is not representative of a population so may ultimately favour one group over another.[9]
11. The use of historical training data creates problems as it may include prejudice or bias information which can have adverse impacts for certain groups.
12. For example in 2019, a study discovered that a clinical algorithm, used by many hospitals to determine which patients required extra medical care, produced racial bias.[10] The algorithm was trained on past data on healthcare spending, which reflected a trend of black patients having less income to spend on their healthcare than white patients – a result of systemic wealth and income disparities.[11] The algorithm's outputs therefore reflected a discriminatory result whereby white patients 'required' more medical care than black patients.[12]
13. The Commission's 2020 technical paper, 'Using artificial intelligence to make decisions: Addressing the problem of algorithmic bias', considers algorithmic bias in detail. It provides insights and strategies on how to mitigate algorithmic bias across several case studies which may assist the Special Rapporteur in their work.

2.2 Criminal justice

14. In the late 1800s Frederick Ludwig Hoffman incorrectly claimed that the higher proportion of black men in prison populations was evidence of higher criminality.[13] This notion was used over the next 50 years to justify the racist 'Jim Crow' laws.[14] Analogously, predictive policing tools which use historical arrest and conviction data now risk embedding racist decision-making in our contemporary criminal justice systems.
15. Data-driven risk assessment tools are being used to predict the likelihood of future criminal behaviour. These tools are being used in a number of countries to assist decision making in the criminal justice system, including decisions regarding sentencing, bail and post-sentence restrictions on people assessed as being likely to commit further crime.[15]
16. The use of AI in the criminal justice system engages several human rights, including the right to equality and non-discrimination, the right to equality before the law, personal security and liberty, the right to privacy, the right to a fair and public hearing, the right to procedural fairness, and the presumption of innocence.[16]
17. A common problem with these tools is that a particular racial or ethnic group may have experienced disproportionate scrutiny by police over a long period of time which, in turn, contributes to members of that group being convicted at a disproportionate rate for particular crimes. Where those statistics are used in training data to create risk assessment tools, these tools are likely to incorrectly 'learn' that members of that group are more likely to commit crime,[17] rather than recognising this data as being a result of past discrimination.
18. Even where this problem is identified, it can be difficult to address, because the risk assessment tool could still rely on an apparently unrelated factor which is, in reality, a close proxy for racial or ethnic origin. For example, it is common for particular areas or postcodes to be populated by people of a similar racial or ethnic background.[18]
19. If a risk assessment tool takes account of postcode, this could have the indirect consequence of considering race or ethnicity.[19] Accordingly, while some risk assessment and policing tools give the appearance of neutrality, they can reinforce discriminatory outcomes at scale.[20]
20. In Australia, there had been a disproportionately high number of Aboriginal and Torres Strait Islander peoples subject to the Suspect Target Management Program, an algorithm-based tool that NSW Police used to

identify individuals at risk of committing criminal offences.[21] The program was discontinued in 2023.

21. These issues have led to global calls for the banning of predictive algorithms in criminal sentencing.[22]

2.3 Transparency and explainability

22. There is often little publicly available information regarding how these risk assessment tools operate, and how they are used to make decisions. This can make it difficult, if not impossible, for individuals subject to these tools (especially those from racial or ethnic groups who are disproportionately impacted by their use in the criminal justice system) to challenge decisions, which in turn can limit human rights including the rights to liberty and freedom of movement.[23]

23. The UN Committee on the Elimination of Racial Discrimination (CERD) recently drew the link between black box (or opaque) AI and an inability to challenge unlawful decision making:

> Given the opacity of algorithmic analytics and decision-making, in particular when employing artificial intelligence methods, discriminatory outcomes of algorithmic profiling can often be less obvious and more difficult to detect than those of human decisions and thus more difficult to contest.[24]

24. Individuals must have the right to request meaningful information about substantially automated decisions. A right to reasons in this context will assist in promoting fairness and transparency in the use of AI in decision-making.

3 Facial recognition technologies

25. Broadly speaking, facial recognition technology (FRT) uses a combination of biometric and AI technologies to match one face with another set face (one-to-one matching) or match one face with another amongst a variety of others (one-to-many matching).

26. In late 2020, CERD published a new general recommendation on the International *Convention on the Elimination of All Forms of Racial Discrimination*. General Recommendation 36 observed:

> The increasing use of facial recognition and surveillance technologies to track and control specific demographics raises concerns with respect to many human rights, including the right to privacy, freedom of peaceful assembly

> and association; freedom of expression and freedom of movement. It is designed to automatically identify individuals based on their facial geometry, potentially profiling people based on grounds of discrimination such as race, colour, national or ethnic origin or gender.[25]

27. Currently, FRT is generally less accurate when identifying women, or people from minority racial groups, as compared with other people.[26] To this end, CERD recently concluded:

> It has been demonstrated that the accuracy of facial recognition technology may differ depending on colour, ethnicity or gender of the persons assessed, which may lead to discrimination.[27]

28. These inaccuracies can mean that certain racial or ethnic groups are more frequently misidentified and subjected to increased police scrutiny.[28]

29. Handling the risks of FRT requires a prudent approach, as the potential benefits of the technology must be measured against its potential harms. While it has the potential to improve public services and law enforcement (i.e. traffic congestion, pollution controls and public security), it can also be used for mass surveillance, ethnic profiling, targeted repression and privacy violations.[29]

30. In 2018 Delhi police used FRT to reunite nearly 3,000 children with their parents in just four days.[30] This pilot FRT programme had later reunited 10,561 missing children with their families after only 15 months in operation.[31] However, the Indian government has also used FRT technology in 2020 to facilitate the arrest of protesters of a citizenship law which critics contend marginalises Muslims.[32] This highlights the potential for technologies such as FRT to both strengthen and diminish human rights, depending on the particular use to which the technology is put.

31. The use of FRT against marginalised populations, such as the Muslim Uyghur minorities in China's Xinjiang Uyghur Autonomous Region, is also well known.[33] The 2022 report by the Office of the United Nations High Commissioner for Human Rights that focused on human rights concerns in this region described 'an ever-present network of surveillance cameras, including deploying facial recognition capabilities' as one element of 'what has been alleged to be a sophisticated, large-scale and systematized surveillance system in practice'.[34]

3.1 Model law

32. Due to the impact that FRT can have on all people (especially those who are racially or ethnically marginalised), Australian civil society has called for regulation via the University of Technology Sydney's Model Law.[35]

33. While it is generally undesirable to regulate a specific technology, there are exceptions. Governments tend to regulate technology deemed high-risk, which helps to explain the comparatively strict laws which govern fields such as gene technology, aviation, healthcare and the energy industry.[36] It is in these areas that regulation is often applied to both the technologies themselves as well as their use. In relation to FRT, the greater the risk to human rights, the greater the need for regulation.

3.2 Moratorium

34. Due to its high-risk nature, there should be a moratorium on the use of FRT in decision making that has a legal, or similarly significant, effect for individuals, or where there is a high risk to human rights, such as in policing and law enforcement.[37] The moratorium should only be in place until specific legislation is introduced with effective human rights safeguards.

4 National regulation

35. Currently there is no AI-specific legislation addressing racial discrimination in Australia. However, the Australian Government has recently released its Interim Response to the *Safe and responsible AI in Australia* discussion paper as Australia seeks to develop its regulatory and policy responses to AI.

36. The Commission has previously recommended that it be resourced to produce guidance on complying with anti-discrimination laws in the use of AI-informed decision-making.[38] The Commission has recently produced a Guidance Resource on Artificial Intelligence (AI) and Discrimination in Insurance Pricing and Underwriting and a Human Rights Impact Assessment Tool for Artificial Intelligence-informed Decision-making Systems in Banking (which directly considered anti-discrimination).

37. National Human Rights Institutions (NHRIs) play an important role in safeguarding human rights, and with the emergence of generative AI there

is an increased need for all NHRIs to engage with these new and emerging technologies.

Endnotes

[1] Australian Human Rights Commission ('AHRC'), 'H*uman Rights and Technology Final Report 2021*' (Final Report, 2021) ('Final Report') 41.
[2] E Tendayi Achiume, '*Report of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*' (Report, A/HRC/44/57, 18 June 2020) [6].
[3] Davey Alba, 'OpenAI Chatbot Spits Out Biased Musings, Despite Guardrails', *Bloomberg* (online, 08 December 2022) <https://www.bloomberg.com/news/newsletters/2022-12-08/chatgpt-open-ai-s-chatbot-is-spitting-out-biased-sexist-results>.
[4] Jana Kasperkevic, 'Google Says Sorry for Racist Auto-tag in Photo App', *The Guardian* (online, 02 July 2015) <https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app>.
[5] Lorraine Finlay & Patrick Hooton, 'Chatbot Race Must not be Run with Blinkers on', *The Australian* (online, 09 February 2023) <https://humanrights.gov.au/about/news/opinions/chatbot-race-must-not-be-run-blinkers>; Lorraine Finlay, Patrick Hooton & Catriona Wallace, 'Big tech is Ignoring the Human Cost Behind the Rise of ChatGPT', *The Australian Financial Review* (online, 13 February 2023) <https://humanrights.gov.au/about/news/opinions/big-tech-ignoring-human-cost-behind-rise-chatgpt>.
[6] AHRC, *Final Report* (Final Report, 2021) 13.
[7] AHRC, *Final Report* (Final Report, 2021) 13.
[8] AHRC, *Final Report* (Final Report, 2021) 106.
[9] AHRC, *Final Report* (Final Report, 2021) 106.
[10] Crystal Grant, 'Algorithms are Making Decisions About Health Care, Which May Only Worsen Medical Racism' *American Civil Liberties Union* (online, 3 October 2022) <https://www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism>.
[11] Crystal Grant, 'Algorithms are Making Decisions About Health Care, Which May Only Worsen Medical Racism' *American Civil Liberties Union* (online, 3 October 2022) <https://www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism>.
[12] Crystal Grant, 'Algorithms are Making Decisions About Health Care, Which May Only Worsen Medical Racism' *American Civil Liberties Union* (online, 3 October 2022) <https://www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism>.
[13] Monika Zalnieriute & Tatiana Cutts, '*How AI and New Technologies Reinforce Systemic Racism*' Submission to the Study on Patterns, Policies and Processes Leading Racial Discrimination and on Advancing Racial Justice and Equality for the 54th Session of the United Nations Human Rights Council 2.
[14] Monika Zalnieriute & Tatiana Cutts, '*How AI and New Technologies Reinforce Systemic Racism*' Submission to the Study on Patterns, Policies and Processes Leading Racial Discrimination and on Advancing Racial Justice and Equality for the 54th Session of the United Nations Human Rights Council 2.
[15] AHRC, *Final Report* (Final Report, 2021) 43; See Committee of Experts on Internet Intermediaries (MSI-NET), Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated

Data Processing Techniques and Possible Regulatory Implications (Council of Europe Study, DGI/2017/12, March 2018) 11.

[16] AHRC, *Final Report* (Final Report, 2021) 43; See Committee of Experts on Internet Intermediaries (MSI-NET), Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications (Council of Europe Study, DGI/2017/12, March 2018) 11.

[17] AHRC, *Final Report* (Final Report, 2021) 43.

[18] AHRC, *Final Report* (Final Report, 2021) 43.

[19] Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104(3) *California Law Review* 671, cited in Filippo A Raso, et al., *Artificial Intelligence & Human Rights: Opportunities & Risks* (Berkman Klein Center Research Publication No 2018-6, September 2018) 21.

[20] Testimony of Andrea Nill Sanchez, Executive Director, AI Now Institute, to the Public Hearing on 'Artificial Intelligence in Criminal Law and its use by the Police and Judicial Authorities in Criminal Matters': Evidence to Committee on Civil Liberties, Justice and Home Affairs, European Parliament, Brussels, 20 February 2020 (Andrea Nill Sanchez); Toby Walsh, et al., '*The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve our Wellbeing*' (Report for the Australian Council of Learned Academies, July 2019) 99.

[21] AHRC, *Final Report* (Final Report, 2021) 43

[22] Monika Zalnieriute & Tatiana Cutts, '*How AI and New Technologies Reinforce Systemic Racism*' Submission to the Study on Patterns, Policies and Processes Leading Racial Discrimination and on Advancing Racial Justice and Equality for the 54th Session of the United Nations Human Rights Council 4

[23] AHRC, *Final Report* (Final Report, 2021) 43

[24] UN Committee on the Elimination of Racial Discrimination, General Recommendation No 36: Preventing and Combating Racial Profiling by Law Enforcement Officials, CERD/C/GC/36 (24 November 2020) [31].

[25] UN Committee on the Elimination of Racial Discrimination, General Recommendation No 36: Preventing and Combating Racial Profiling by Law Enforcement Officials, CERD/C/GC/36 (24 November 2020) [35].

[26] See, for example, Joy Buolamwini and Timinit Guru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' (2018) 81 *Proceedings of Machine Learning Research* 1; KS Krishnapriya, et al., '*Characterizing the Variability in Face Recognition Accuracy Relative to Race*' (Conference Paper, IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019); Inioluwa Deborah Raji and Joy Buolamwini, '*Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products*' (Conference on Artificial Intelligence, Ethics, and Society, 2019).

[27] UN Committee on the Elimination of Racial Discrimination, General Recommendation No 36: Preventing and Combating Racial Profiling by Law Enforcement Officials, CERD/C/GC/36 (24 November 2020) [35].

[28] AHRC, *Final Report* (Final Report, 2021) 116.

[29] See generally Genia Kostka, Lea Steubacker & Miriam Meckel 'Under Big Brothers Watchful Eye: Cross-country Attitudes Towards Facial Recognition Technology' (2023) 40(1) *Government Information Quarterly*.

[30] Sanjeev Kumar, 'Delhi: Facial recognition system helps trace 3,000 missing children in 4 days' *The New India Express*, (online, 22 April 2018)

<https://www.newindianexpress.com/nation/2018/apr/22/3000-missing-children-traced-in-four-days-by-delhi-police-with-facial-recognition-system-software-1804955.html>.

[31] Julie Zaugg, 'India is trying to build the world's biggest facial recognition system' *CNN Business* (online, 18 Oct 2019) <https://edition.cnn.com/2019/10/17/tech/india-facial-recognition-intl-hnk/index.html>.

[32] Alexandra Ulmer & Zeba Siddiqui, 'India's use of facial recognition tech during protests causes stir' *Reuters* (online, 17 February 2020) <https://www.reuters.com/article/us-india-citizenship-protests-technology-idUSKBN20B0ZQ>.

[33] James Leibold, 'Surveillance in China's Xinjiang region: Ethnic sorting, coercion, and inducement' *Journal of Contemporary China* (2020) 29(121) 46-60.

[34] Office of the United Nations High Commissioner for Human Rights, *OHCHR Assessment of human rights concerns in the Xinjiang Uyghur Autonomous Region, People's Republic of China* (31 August 2022) [96] <https://www.ohchr.org/en/documents/country-reports/ohchr-assessment-human-rights-concerns-xinjiang-uyghur-autonomous-region>.

[35] See generally Human Technology Institute, '*Facial Recognition Technology Towards a Model Law*' ('Model Law') (University of Technology Sydney, Report, September 2022).

[36] AHRC, Final Report (Final Report, 2021) 119.

[37] AHRC, Final Report (Final Report, 2021) 116.

[38] AHRC, Final Report (Final Report, 2021) 108.