Secretariat of the Human Rights Council Advisory Committee
OHCHR - United Nations Office at Geneva CH-1211
Geneva 10, Switzerland Fax: +41 22 917 9011
OHCHR-hrcadvisorycommittee@un.org

03rd October 2022

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

Dr. Monika Zalnieriute
Faculty of Law and Justice UNSW Sydney, Australia;
ARC Centre of Excellence 'Automated Decision-Making and Society';
Allens Hub for Technology, Law and Innovation;
Australian Human Rights Institute, UNSW;
m.zalnieriute@unsw.edu.au

Associate Professor Tatiana Cutts
Melbourne Law School, University of Melbourne, Australia;
Centre for Artificial Intelligence and Digital Ethics;
London School of Economics Law, Technology and Society;
tatiana.cutts@unimelb.edu.au

Dear Members of the Human Rights Council Advisory Committee,

Thank you for the opportunity to make a submission to this inquiry. We do so in a private capacity as scholars of human rights law, legal theory, and technology at UNSW Sydney and the University of Melbourne, Australia. The views expressed are our own, not those of our institutions. In line with our expertise, this submission addresses questions 19, 20, and 22 in *Questionnaire on patterns, policies, and processes leading to incidents of racial discrimination and on advancing racial justice and equality*.

Systemic, structural, and institutional racism is prominent across many areas of public life in national and international settings, such as access to justice, enjoyment of political and social rights, and access to key public services. New emerging technologies, such as Artificial Intelligence (AI) and facial recognition, play a significant role in sustaining systemic, structural, and institutional racism. In this submission, we would like to draw attention to several issues in particular:

1) discriminatory effects of AI algorithms;

2) racist implications of facial recognition technology;

3) racist digital profiling and targeting on Internet platforms; and

4) the need for legally binding obligation for private actors to eradicate systemic, structural, and institutional racism.

1. AI, Machine Learning Algorithms and Systemic Racism (Question No 20)

First, new technologies such as AI and machine learning algorithms reinforce structural racism, whereby certain peoples and communities are treated as having a lower status in society. In 1896, statistician Frederick Ludwig Hoffman invoked brute evidence of a higher proportion of black men amongst prison populations as evidence of higher criminality.[1] That conclusion was used to justify the racist 'Jim Crow' laws that developed across the US over the next 50 years. In the same way, systems of predictive policing are based on past arrest and conviction data that embed racist decision-making. For instance, the Suspect Target Management Plan (STMP) is a New South Wales (Australia) Police Force initiative designed to reduce crime among high-risk individuals

1 Frederick L Hoffman, *The Race Traits and Tendencies of the American Negro* (1896).

through proactive policing. Data shows the STMP disproportionately targets young people, particularly Aboriginal and Torres Strait Islander people: of the 73 children under the age of 16 identified as targets, 73% were indigenous, compared with national census data of 3.2%.[2]

AI systems in public health also entail implications for people of colour. For instance, healthcare providers in the US routinely attempt to limit their exposure to high medical costs by adopting 'complex care management'[3] systems that channel resources to so-called 'high-cost beneficiaries'.[4] Optum, an algorithmic service owned by UnitedHealth Group, was developed to streamline the process of identifying these beneficiaries, and is now applied to more than 200 million people across the US each year.[5] Recent evidence suggests that Optum systematically fails to refer people of colour to the support programmes at the same level of healthcare need as white people.[6] The reason for this failure is that the algorithm was trained to predict spending behaviours rather than hospitalizations, and people of colour are less inclined than white people to seek medical care when they are equally ill.[7]

The concern is not simply with *overall* rates of predictive accuracy; it is with how the burden of predictive inaccuracy is borne – what has been called 'predictive parity'.[8] The same concern can arise from a lack of representative data, which feeds into poor decision-making. For instance, the anti-coagulant medication warfarin is regularly prescribed to patients on the basis of dosing algorithms, which incorporate race as a predictor along with clinical and genetic factors.[9] Yet, most of the studies used to develop those algorithms were conducted in cohorts with >95% white

---

2 Vicki Sentas and Camilla Pandolfini, *Policing Young People in New South Wales*: A study of the Suspect Targeting Management Plan. A Report of the Youth Justice Coalition *NSW* (Sydney: Youth Justice Coalition NSW, 2017).

3 Clemens S Hong, Allison L Siegel, and Timothy G Ferris, "Caring for high-need, high-cost patients: what makes for a successful care management program?" (2014) 19 Issue Brief (Commonwealth Fund) 1.

4 Carol Urato, Nancy McCall, Jerry Cromwell, Nancy Lenfestey, Kevin Smith, and Douglas Raeder, "Evaluation of the Extended Medicare Care Management for High Cost Beneficiaries (CMHCB) Demonstration: Massachusetts General Hospital (MGH)" Final Report (2013) Available at http://docplayer.net/18477639-Evaluation-of-the-extended-medicare-care-management-for-high-cost-beneficiaries-cmhcb-demonstration-massachusetts-general-hospital-mgh.html

5 Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations" (2019) 25 Science 447.

6 ibid.

7 ibid.

8 William Dieterich, Christina Mendoza, and Tim Brennan, "COMPAS risk scales: Demonstrating accuracy equity and predictive parity" (Northpoint, 2016).

9 Ash Clinical News, 'Race-Specific Dosing Guidelines Urged for Warfarin', Feburary 2017: https://ashpublications.org/ashclinicalnews/news/2145/Race-Specific-Dosing-Guidelines-Urged-for-Warfarin (accessed 25th August 2022).

European ancestry, and there is now robust evidence that these algorithms assign a 'lower-than-needed dose' to black patients, putting them at serious risk of heart attack, stroke, and pulmonary embolism.[10] If data are not representative, overall predictive accuracy may be good (or better than human decision-making alone), but the risks of inaccuracy will be borne unevenly.

Perhaps most worrying of all, predictive algorithms are used across the US to impose criminal sentences on the basis of factors such as economic hardship and underemployment.[11] These factors are themselves an objectionable basis for the imposition of criminal sanctions, but they also correlate strongly to race. As Associate Professor Tatiana Cutts has highlighted in her research (available freely on SSRN),[12] the result is that people of colour are treated as *predestined* for a life of crime – as if they cannot but recidivate. This is objectionable both from an instrumental point of view, and because of what is expresses about the rational agency of people of colour.

Moreover, there is a concern that these algorithms concentrate the risk of error with people of colour.[13] The quality of the data has thus far been insufficient to determine the extent of this concern and the steps needed to ameliorate it, and it is very difficult to fashion studies that can. If we rely upon evidence about rearrest or reconviction data as a proxy for recidivism, we risk embedded pre-existing disparities in the way in which society has treated people of colour.

Thus, AI and machine learning technologies can reinforce racism in many areas of life, including healthcare and criminal justice. We invite the HRC Advisory Committee to call for a ban on the use of predictive algorithms in criminal sentencing. These practices not only impose burdens upon people of colour without justification; they also reinforce unjust views that are liable to result in the systemic denial of opportunities to people of colour across many aspects of public and private life.

---

10 Nita A Limdi, Todd M Brown, Qi Yan, Jonathan L Thigpen, Aditi Shendre, Nianjun Liu, Charles E Hill, Donna K Arnett, and T Mark Beasley, Race influences warfarin dose changes associated with genetic factors (2015) 126 *Blood* 539, 544.

11 See e.g. Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica, "Machine Bias, There's software used across the country to predict future criminals. And it's biased against blacks", May 23, 2016: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

12 Tatiana Cutts, 'Supervising Automated Decisions' in Zofia Bednarz and Monika Zalnieriute (eds), *Money, Power and AI: From Automated Banks to Automated States,* Cambridge University Press, 2023, *forthcoming*, available at http://ssrn.com/abstract=4215108.

13 See generally Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica, "Machine Bias, There's software used across the country to predict future criminals. And it's biased against blacks", May 23, 2016: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

2. Facial Recognition Technology Reinforces Systemic Racism (Question No 20)

Second, many public places such as cities and airports, as well as personal electronic devices, are increasingly equipped with facial recognition technology. As Dr. Zalnieriute explains in a recent article (available freely on SSRN),[14] the surveillance of public spaces not only has a 'chilling' effect on the rights to freedom of expression, peaceful association, and assembly; it also entails systemically racist effects, which have been demonstrated in an increasing body of academic research.[15] The emerging consensus is that facial recognition technologies are not 'neutral',[16] but instead reinforce historical inequalities.[17] For example, studies have shown that facial recognition technology performs poorly in relation to women, children, and people of colour.[18] The discrimination can be introduced into the facial recognition technology software in three technical ways: first, through the machine learning process through the training data set and system design; second, through technical bias incidental to the simplification necessary to translate reality into code; and third, through emergent bias which arises from users' interaction with specific populations.[19] Because the training data for facial recognition technologies in law enforcement

---

14 Monika Zalnieriute, 'Burning Bridges: The Automated Facial Recognition Technology and Public Space Surveillance in the Modern State' (2021) 22 Columbia Science and Technology Review 314. Available at https://papers.ssrn.com/abstract=3805494.

15 ibid.

16 Clare Garvie, Alvaro Bedoya and Jonathan Frankle, 'The Perpetual Line-Up' (Centre on Privacy and Technology 2016) <https://www.perpetuallineup.org/> accessed 11 November 2019; BF Klare and others, 'Face Recognition Performance: Role of Demographic Information' (2012) 6 Information Forensics and Security, IEEE Transactions On 7 1789; Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', *Proceedings of Machine Learning research* (2018) <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf> accessed 17 June 2020.

17 Matthew Schwartz, 'Color-Blind Biometrics? Facial Recognition and Arrest Rates of African-Americans in Maryland and the United States' (Thesis in partial fulfilment of a Master of Public Policy, Georgetown University 2019) 15.

18 Salem Hamed Abdurrahim, Salina Abdul Samad and Aqilah Baseri Huddin, 'Review on the Effects of Age, Gender, and Race Demographics on Automatic Face Recognition' <https://link-springer-com.wwwproxy1.library.unsw.edu.au/content/pdf/10.1007/s00371-017-1428-z.pdf> accessed 2 June 2020; 'Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots' (*American Civil Liberties Union*) <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28> accessed 2 June 2020.

19 Rebecca Crootof, "'Cyborg Justice" and the Risk of Technological–Legal Lock-In' (2019) 119 Columbia Law Review 1, 8; Batya Friedman and Helen Fay Nissenbaum, 'Bias in Computer Systems' (1996) 14 ACM Transactions on Information Systems 330, 333–36.

context comes from photos relating to past criminal activity, [20] people of colour are overrepresented in facial recognition technology training systems.[21] In some jurisdictions, such as the United States, people of colour are at a much higher risk of being *pulled over,*[22] *searched,*[23] *arrested,*[24] *incarcerated,*[25] *and wrongfully convicted*[26] than whites. Therefore, facial recognition technology produces many false positives because it is already functioning in a highly discriminatory environment.

Law and border enforcement agencies around the world are experimenting with automated facial recognition technology with complete discretion and on *ad hoc* basis, without appropriate legal frameworks to govern their use nor sufficient oversight or public awareness.[27] We invite the HRC Advisory Committee to call for a ban on the use of facial recognition technology[28] for its disproportionate impact on people of colour.

---

[20] Henriette Ruhrmann, 'Facing the Future: Protecting Human Rights in Policy Strategies for Facial Recognition Technology in Law Enforcement' (May 2019) 46 <https://citrispolicylab.org/wp-content/uploads/2019/09/Facing-the-Future_Ruhrmann_CITRIS-Policy-Lab.pdf> accessed 1 June 2020; Garvie, Bedoya and Frankle (n 11).

[21] Ruhrmann (n 20) 63; Garvie, Bedoya and Frankle (n 20).

[22] 'New Data Reveals Milwaukee Police Stops Are About Race and Ethnicity' (*American Civil Liberties Union*) <https://www.aclu.org/blog/criminal-law-reform/reforming-police/new-data-reveals-milwaukee-police-stops-are-about-race-and> accessed 2 June 2020; Frank R Baumgartner, Derek A Epp and Kelsey Shoub, *Suspect Citizens What 20 Million Traffic Stops Tell Us About Policing and Race* (Cambridge University Press 2018).

[23] 'New Data Reveals Milwaukee Police Stops Are About Race and Ethnicity' (n 22); Camelia Simoiu, Sam Corbett-Davies and Sharad Goel, 'The Problem of Infra-Marginality in Outcome Tests for Discrimination' (2017) 11 The Annals of Applied Statistics 1193; Lynn Lanton, 'Police Behavior during Traffic and Street Stops, 2011' <https://www.bjs.gov/content/pub/pdf/pbtss11.pdf> accessed 2 June 2020.

[24] 'NAACP | Criminal Justice Fact Sheet' (*NAACP*) <https://www.naacp.org/criminal-justice-fact-sheet/> accessed 2 June 2020; Megan Stevenson and Sandra Mayson, 'The Scale of Misdemeanor Justice' (2018) 98 Boston University Law Review 371.

[25] 'The Color of Justice: Racial and Ethnic Disparity in State Prisons' (*The Sentencing Project*) <https://www.sentencingproject.org/publications/color-of-justice-racial-and-ethnic-disparity-in-state-prisons/> accessed 2 June 2020.

[26] Samuel Gross, Maurice Possley and Klara Stephens, 'Race and Wrongful Convictions in the United States' (National Registry of Exonerations 2017) <http://www.law.umich.edu/special/exoneration/Documents/Race_and_Wrongful_Convictions.pdf> accessed 2 June 2020.

[27] Monika Zalnieriute, 'Burning Bridges: The Automated Facial Recognition Technology and Public Space Surveillance in the Modern State' (2021) 22 Columbia Science and Technology Review 314. Available at https://papers.ssrn.com/abstract=3805494.

[28] For example, the UK Equality and Human Rights Commission had, in March 2020, called on suspension, see 'Facial Recognition Technology and Predictive Policing Algorithms Out-Pacing the Law' (*Equality and Human*

3. Racist Profiling & Targeting on Internet Platforms (Question 20)

Third, people of colour from marginalized groups, such as LGBTI communities, face additional threats and challenges in the digital environment and on media platforms. As Dr. Monika Zalnieriute explains in her research, the rise of large-scale data collection and algorithm-driven analysis targeting sensitive information poses many threats for people of colour, especially from LGBTI communities, who are especially vulnerable to privacy intrusion due to their often hostile social, political, and even legal environments.[29] A great deal of publicly available data, such as Facebook friend information or individual music playlists on YouTube, can be processed effectively to infer individual sexual preferences with high levels of accuracy.[30] Indeed, this predictions may be more accurate than those of our human friends.[31] If widely-traded advertising information 'correctly discriminates between homosexual and heterosexual men in 88% of cases',[32] most Internet users should assume that the companies advertising to them can predict their sexual orientation with a high degree of accuracy — and are incentivised to do so in order to sell them products. The issues go well beyond simple product advertising. Amongst many other examples, they include different treatment in health and life insurance policies (as discussed above),[33] as well as lead to arrests in certain countries based on sexual orientation. Such ready access to personal information can get even more complicated with the 'real name' policies of social platforms, such as Facebook,[34] which may place people of colour, especially women of colour from LGBTI communities, in danger of physical assaults.

Rights Commission, 12 March 2020) <https://www.equalityhumanrights.com/en/our-work/news/facial-recognition-technology-and-predictive-policing-algorithms-out-pacing-law> accessed 16 September 2020.

29 Monika Zalnieriute, 'Digital Rights of LGBTI Communities: A Roadmap For A Dual Human Rights Framework' in Ben Wagner, Matthias C Kettlemann and Kilian Vieth (eds), *Research Handbook on Human Rights and Digital Technologies* (Edward Elgar 2019); Monika Zalnieriute, 'The Anatomy of Neoliberal Internet Governance: A Queer Critical Political Economy Perspective' in Dianne Otto (ed), *Queering International Law: Possibilities, Alliances, Complicities and Risks* (1st edn, Routledge 2017). Available at https://papers.ssrn.com/abstract=3332133 and https://papers.ssrn.com/abstract=2894136.

30 Michal Kosinski, David Stillwell and Thore Graepel, 'Private Traits and Attributes Are Predictable from Digital Records of Human Behavior' (2013) 110 *Proceedings of the National Academy of Sciences* 5802.

31 Wu Youyou, Michal Kosinski and David Stillwell, 'Computer-Based Personality Judgments Are More Accurate than Those Made by Humans' (2015) 112 *Proceedings of the National Academy of Sciences* 1036.

32 Kosinski, Stillwell and Graepel (n 35).

33 Angela Daly, 'The Law and Ethics of "Self Quantified" Health Information: An Australian Perspective' (2015) 5 International Data Privacy Law 144.

34 Andrew Griffin, 'Facebook to Tweak "Real Name" Policy after Backlash' (*The Independent*, 1 November 2015) <http://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-to-tweak-real-name-policy-after-backlash-from-lgbt-groups-and-native-americans-a6717061.html> accessed 6 November 2016.

4. The Need for Binding Human Rights Obligations for Private Actors (Question No 19)

Finally, to address systemic racism, we need to develop binding international human rights law for private actors to remedy the violations right to freedom from discrimination of people of colour. Because private actors such as digital platforms, and AI companies providing algorithms to police and law enforcement agencies, hold a lot of power in the design and running of technologies, the basic tools of accountability and governance — public and legal transparency and pressure – are very limited. As Dr. Monika Zalnieriute has argued in detail, existing efforts focused on voluntary 'social and corporate responsibility' and ethical obligations of technology companies are insufficient and incapable to tackle structural racism.[35] The existing international human rights framework is not adequate to safeguard human rights in the age of AI and new technologies because its obligations are limited to states, and not such private actors. Binding obligations for private actors under international human rights framework are needed to ensure protection of fundamental rights in the digital age for three main reasons:

- First, to rectify an imbalance between hard legal commercial obligations and human rights soft law.
- Second, to ensure that people of colour, whose rights to freedom of expression, association and anti-discrimination have been affected, can access an effective remedy.
- Finally, private actors are themselves engaging in the balancing exercise around fundamental rights, therefore, an explicit recognition of their human rights obligations is crucial for the future development of access to justice in the digital age.

Therefore, we invite the UN HRC Advisory Committee to call for the development of binding international human rights law for private actors to remedy the violations of right to freedom from discrimination, especially in a transnational context. The development of such obligations is crucial for eradicating structural, systemic and institutional racism.

---

35 Monika Zalnieriute, 'From Human Rights Aspirations to Enforceable Obligations by Non-State Actors in the Digital Age: The Case of Internet Governance and ICANN' [2019] Yale Journal of Law & Technology 278. Available at https://papers.ssrn.com/abstract=3333532.

5. A Call for Action & Recommendations (Question 22)

We invite the HRC Advisory Committee to:

1) Call for a ban on the use of predictive algorithms in criminal sentencing.
2) Call for a ban on the use of facial recognition technology in public city spaces.
3) Call for the development of binding international human rights law for private actors to remedy the violations of the right to freedom from discrimination of people of colour, especially in a transnational context.

These steps are crucial for eradicating structural, systemic, and institutional racism.

Acknowledgments

Dr. Monika Zalnieriute's work and research for this submission has been funded by Australian Research Council Discovery Early Career Research Award ('Artificial Intelligence Decision-Making, Privacy and Discrimination Laws', project number DE210101183).