Responding to discrimination arising from the use of artificial intelligence as part of the decision-making process

Authors

Dr Tetyana (Tanya) Krupiy, lecturer, Newcastle University

Professor Martin Scheinin, British Academy Global Professor, University of Oxford

For more information, please contact Tanya at tanya.krupiy@newcastle.ac.uk

Context

Tetyana (Tanya) Krupiy demonstrated that the Convention on the Elimination of All Forms of Discrimination against Women and the Convention on the Rights of Persons with Disabilities may be interpreted so as to include an additional test prohibiting discrimination in a context where actors use artificial intelligence as part of the decision-making process. Martin Scheinin showed that the International Convention on the Elimination of All Forms of Racial Discrimination and the International Covenant on Civil and Political Rights may be interpreted so as to incorporate this test. The findings appear in two academic journals: Human Rights Law Journal (Oxford University Press) and International Human Rights Law Review (Brill).

Test for algorithmic discrimination which is included in numerous existing international human rights law treaties

The employment of the decision-making process relying on artificial intelligence should be treated as amounting to discrimination whenever it places the subjects of the decision-making in a situation which:[1]

i) is inequitable;[2] [An inequitable situation may emerge because the organisation deploying artificial intelligence places an individual in an unequal situation.[3] Additionally, there is an inequitable situation when a state fails to regulate social and institutional relationships in which individuals find themselves, in a manner which would acknowledge the universal vulnerability of all human beings.[4]] OR
ii) impedes the ability of candidates or other subjects of decision-making to define and to express their identities;[5] OR

[1] Krupiy, ‘Meeting the Chimera: How the CEDAW Can Address Digital Discrimination’, (2021) 10 *International Human Rights Law Review* 1 at 26.

[2] Fineman, ‘Vulnerability and Inevitable Inequality’ (2017) 4 *Oslo Law Review* 133 at 143.

[3] Krupiy and Scheinin, ‘Disability Discrimination in the Digital Realm: How the ICRPD Applies to Artificial Intelligence Decision-Making Processes and Helps in Determining the State of International Human Rights Law’ (2023) 23 (3) *Human Rights Law Review* https://doi.org/10.1093/hrlr/ngad019.

[4] Fineman, ‘Vulnerability and Inevitable Inequality’ (2017) 4 *Oslo Law Review* 133 at 147; Fineman, ‘Equality and Difference – The Restrained State’ (2015) 66 *Alabama Law Review* 609 at 624-25.

[5] Krupiy, ‘Meeting the Chimera: How the CEDAW Can Address Digital Discrimination’, (2021) 10 *International Human Rights Law Review* 1 at 26.

iii) impedes the ability of the subjects to access opportunities which are crucial to their self-actualisation and which enable them to overcome setbacks in the face of adversity on an equal basis;[6] OR
iv) places the subject at a disadvantage; OR
v) does not account for how structural inequality in society affects the ability of the subject to access an opportunity;[7] OR
vi) compounds a pre-existing disadvantage by making the harm caused by prior injustice worse or by creating a new type of harm in another area of life.[8]

The test is disjunctive in nature, i.e., the presence of any single one of the criteria listed as i) to vi) entails that the application of artificial intelligence constitutes discrimination. As the authors demonstrate, each of the six criteria has a solid basis in treaty law and/or moral theories of discrimination. As comprehensive and contemporary instruments against discrimination, a number of international human rights law treaties are capable of addressing the full scope of factors that make the use of artificial intelligence discriminatory and hence prohibited.

The proposed test for algorithmic discrimination allows for the identification of a set of rights that flow from the contours of discrimination. Irrespective of whether the use of algorithms in decision-making directly violates other human rights, such as the right to privacy, the right to work, the right to health or the right to social security, any discrimination inherent in its use gives rise to a range of legislative and other obligations of the state, as well as to rights and justified claims by persons subjected to such discrimination. The scope of these entitlements, rights and state obligations are explained in more detail in published articles. The proposed test provides a method for assessing when the use of algorithms is discriminatory and therefore triggers the entitlements, rights and obligations in question.

Rights which this test confers on individuals and respective obligations on the state to ensure the enjoyment of these rights

- An inclusive right to be consulted and a right to have significant input[9] into decisions relating to the funding, development, design, sale, procurement and employment of artificial intelligence technology.
- Prohibition of the full automation of the decision-making process using artificial intelligence technology.[10]
- A right to request decision-making by a human decision-maker.
- A right to object to the employment of decision-making processes which either rely on artificial intelligence technology in part or which have a human decision-maker who can veto the decision, or which involve a human decision-maker reaching the decision using the outputs generated by the artificial intelligence software.

[6] Fineman, 'Vulnerability and Inevitable Inequality' (2017) 4 *Oslo Law Review* 133 at 146-47.

[7] Khaitan, *A Theory of Discrimination Law* (2015) at 31.

[8] Hellman, 'Indirect Discrimination and the Duty to Avoid Compounding Injustice' in Collins and Khaitan (eds), *Foundations of Indirect Discrimination* (2018) 105 at 113.

[9] Equal Rights Trust, *Equality by Design in Algorithmic Decision-making Systems* (2022) at p 10-11.

[10] Krupiy, 'Meeting the Chimera: How the CEDAW Can Address Digital Discrimination', (2021) 10 *International Human Rights Law Review* 1 at 34.

- A right to veto decisions regarding in what contexts and for what purpose to employ decision-making processes relying on artificial intelligence.
- A right to veto decisions regarding what goals the decision-making process which relies on artificial intelligence technology is designed to achieve,[11] what decision-making criteria may be used[12] and what type of inputs the decision-making process relying on artificial intelligence software utilises.[13]
- A right to object to the collection and labelling of data relating to oneself without obtaining prior informed consent.[14] The developers need to obtain informed consent prior to labelling the individuals' data in a particular way.
- A state obligation to require developers to use data which is representative of the entire population in all its diversity.[15]
- A state obligation to prohibit the use of data which incorporates systemic societal bias.[16]
- A right to contest how the system models the external environment, allocates data about a person to a particular data cluster and clusters data based on similarity.[17]
- A right to effective channels for contesting what inferences the artificial intelligence technology drew about the subject of the decision-making, where such inferences would be incompatible with one of the limbs of the legal test for algorithmic discrimination.
- A right to contest the logic which the artificial intelligence system used to produce the decision, "including but not limited to weightings of different inputs and parameters."[18]
- A right to object to the prediction and to the final decision.
- A right to object to a decision where the criterion for reaching a decision constitutes a proxy for a protected characteristic.[19]

[11] Barocas and Selbst, 'Big Data's Disparate Impact' (2016) 104 *California Law Review* 671 at 678–80.

[12] Barocas and Selbst, 'Big Data's Disparate Impact' (2016) 104 *California Law Review* 671 at 722.

[13] Barocas and Selbst, 'Big Data's Disparate Impact' (2016) 104 *California Law Review* 671 at 688.

[14] The right to health in Art 25 CRPD includes the right to health care on the basis of free and informed consent. Committee on the Rights of Persons with Disabilities, General Comment No. 1 (2014) Article 12: Equal Recognition Before the Law, 19 May 2014 at para 41. The CRPD Committee "recommends that States parties ensure that decisions relating to a person's physical or mental integrity can only be taken with the free and informed consent of the person concerned." Ibid., para 42. Such duty should be extended to other contexts including the collection and labelling of data.

[15] Engler, 'For Some Employment Algorithms, Disability Discrimination by Default', 31 October 2019, available at: www.brookings.edu/blog/techtank/2019/10/31/for-some-employment-algorithms-disability-discrimination-by-default [last accessed: 28 November 2022].

[16] Bach, Gerdon, Kern and Kreuter, 'Social Impacts of Algorithmic Decision-making: A Research Agenda for the Social Sciences' (2022) 9 *Big Data & Society* 1 at 3; Peters, 'Algorithmic Political Bias in Artificial Intelligence Systems' (2022) 35 *Philosophy & Technology* 24 at 25.

[17] Provost and Fawcett, *Data Science for Business* (2013) 24.

[18] Adams-Prassl, Abraha, Kelly-Lyth, Silberman, and Rakshita, 'Regulating Algorithmic Management: A Blueprint' (Forthcoming) 14 *European Labour Law Journal* at 13.

[19] Prince and Schwarcz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big

- A right to object to the decision if the artificial intelligence uses information which is correlated with membership of a protected group,[20] or if it uses data which encodes membership of a protected group,[21] or if it can predict membership of a protected group based on the data.[22]
- A right to object to the decision if the use of artificial intelligence in the decision-making process leads to unequal degrees of accuracy of the decisions for different persons or groups.[23]
- A right to challenge the design and the decision outputs of artificial intelligence systems which fail to take into account how the intersectional identities of individuals influence their access to opportunities.[24] For instance, the applicant cannot use transferrable skills to evidence the suitability for the position.[25]
- A right to challenge the use of artificial intelligence in the decision-making process if the employment of this technology creates new types of disadvantaged groups[26] through detecting correlations in the data.[27] Such groups can be counterintuitive and therefore challenging to comprehend for human beings.[28]
- A right to challenge the employment of decision-making processes relying on artificial intelligence which fail to meet accessibility requirements.[29] The accessibility requirements should anticipate that users with a disability will interact with the system in different and unforeseen ways.[30] The system should be responsive and adapt to the needs of the user with a disability.
- A right to request access to detailed information relating to the design, operation and decision output generated by the artificial intelligence technology.[31] The individuals must also be able to understand the basis for the decision[32] and what they could have done to obtain a positive result.

Data' (2020) 105 *Iowa Law Review* 1257 at 1261.

[20] Prince and Schwarcz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data' (2020) 105 *Iowa Law Review* 1257 at 1261.

[21] Dwork et al., *Fairness Through Awareness* (Proceedings of the 3d Innovations in Theoretical Computer Science Conference, 2012) 214, 226.

[22] Prince and Schwarcz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data' (2020) 105 *Iowa Law Review* 1257 at 1264.

[23] Gajane, 'On Formalising Fairness in Prediction with Machine Learning' (2017) *arXiv:1710.03184v1* [cs.LG] 1 at 4.

[24] Krupiy, 'Meeting the Chimera: How the CEDAW Can Address Digital Discrimination', (2021) 10 *International Human Rights Law Review* 1 at 30.

[25] Krupiy, 'Meeting the Chimera: How the CEDAW Can Address Digital Discrimination', (2021) 10 International *Human Rights Law Review* 1 at 30.

[26] Wachter, 'The Theory of Artificial Immutability: Protecting Algorithmic Groups Under Anti-Discrimination Law' (Forthcoming) *Tulane Law Review* 1 at 5.

[27] Ibid., p 9.

[28] Ibid.

[29] Committee on the Rights of Persons with Disabilities, General Comment No. 2 on Article 9: Accessibility, 31 March–11 April 2014 at para 13.

[30] Humphry and Park, 'Exclusion by Design: Intersections of Social, Digital and Data Exclusion' (2019) 22 *Communication & Society* 934 at p 936.

[31] Adams-Prassl, Abraha, Kelly-Lyth, Silberman, and Rakshita, 'Regulating Algorithmic Management: A Blueprint' (Forthcoming) 14 *European Labour Law Journal* at 13.

[32] OECD AI Policy Observatory, 'Transparency and Explainability (Principle 1.3)', 2023, available at: https://oecd.ai/en/dashboards/ai-principles/P7 [last accessed: 5 July 2023].