Submitted on Wed, 04/01/2020 - 18:50

Submitted by: Anonymous

Submitted values are:

Name
Tetyana Krupiy

Address
███████████████████████████████████████
██████████

Email address
███████████████

Telephone number
████████████

Is this submission presented on behalf of an organisation:
No

I have read the information about this Project concerning publication, confidentiality, and privacy obligations at https://www.humanrights.gov.au/submission-policy
Yes

My submission is:
Public

Enter your response in the box below
I have over three years of experience of conducting research on the legal and social aspects of the use of artificial intelligence technology. As a researcher I find it important that society takes steps to protect human diversity and individuals who have historically experienced unjust treatment. I found that the Human Rights and Technology Discussion Paper does not examine in sufficient depth the fact that representing individuals who enjoy protection under the prohibition of discrimination using mathematical models within an artificial intelligence system is harmful for those groups. I published a blog entry with the Groningen Journal of International Law in which I discuss that the prohibition of discrimination includes a right for the protected groups to control how the computer scientists represent them in artificial intelligence models as well as to object to being represented using mathematical methods. The blog entry examines briefly the manner in which the representation of individuals in mathematical models injures groups and undermines human diversity. Such approach to representing individuals is steeped in colonial practices and threatens the protection of human diversity. Please find the text of the blog entry and the link to the website below. If you would like to read more information about the manner in which representing individuals in mathematical models harms the protected groups then I can share the content of a forthcoming academic article I wrote with your organisation at no cost. The Dutch government provides funding for researchers to make their research open access. While I arrange open access I would request that the content of my research not be widely circulated.

https://grojil.org/2020/03/16/the-ascendance-of-artificial-intelligence-how-international-human-rights-law-protects-human-diversity/

The ascendance of artificial intelligence: how international human rights law protects human diversity

March 16, 2020 by GroJIL-blog Leave a comment

Dr. Tetyana (Tanya) Krupiy

Artificial intelligence ("AI") technologies are ushering in the fourth industrial revolution. Indeed, countries are in a race for dominance in the field of AI and companies are pouring billions of dollars into such research. The Special Rapporteur on Extreme Poverty and Human Rights, Philip Alston released a report in October 2019 in which he analysed some of the human rights concerns associated with using AI technology to predict the future behaviour of human beings. He expressed the view that the use of this technology is "highly likely" to replicate existing biases (par 77). There is evidence emerging from the United States, that historically disadvantaged groups including people of colour, women and individuals with a health impairment are particularly adversely affected by the use of AI systems for a variety of different purposes. Diane Desierto expresses the concern that the employment of AI systems instrumentalises individuals. It undermines their human dignity and impairs their self-determination. This stems from the fact that the use of AI technology leads to human skills becoming obsolete, to the displacement of human decision-making, to the curation of human experiences through online platforms and to the impairment of the ability of human beings to create genuine interpersonal connections. This intervention engages with the concerns Daine Desierto voices from a different angle. It investigates whether the prohibition of discrimination in international human rights law protects human diversity through conferring rights on individuals with protected characteristics to object to being represented in mathematical space of an algorithmic model.

The centrality of the prohibition of discrimination can be seen in the fact that the Inter-American Court in an Advisory Opinion Juridical Condition and Rights of Undocumented Migrants expressed the view that this norm is peremptory in its character (jus cogens) from which no derogation is permitted (par 101).The prohibition of discrimination protects human diversity. According to Richard Crisp, diversity relates to the fact that there are a multitude of ways in which human beings can define their identity and that of other individuals (p 1). He maintains that identities need to be conceived of in terms of complex representations, rather than in terms of particular categories, like ethnicity (p 1). The prohibition of discrimination protects human diversity through framing the legal protection around the possession of a protected characteristic. It incorporates a right of individuals to define their identity and control inappropriate representations of their relevant identity. Discrimination is defined in international treaties. The treaties and treaty bodies define discrimination as any distinction, exclusion, restriction or preference based on a protected characteristic which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise, on an equal footing, of human rights and fundamental freedoms in the political, economic, social, cultural or any other field of public life. While the prohibition of discrimination has customary international law status, the specialist treaties that states concluded provide evidence of how states understand the nature of protected characteristics and the type of protection attached to the possession of the characteristics in question. The specialist treaties states concluded include the Convention on the Elimination of all Forms of Racial Discrimination ("CERD") and the Convention on the Elimination of all Forms of Discrimination Against Women ("CEDAW").

Numerous sources support the fact that the prohibition of discrimination confers a right on groups to define their identity. The CERD Committee explained in General Recommendation 8 that in establishing whether an individual belongs to a protected racial, ethnic, national or indigenous group, regard shall be had to how individuals define themselves. Further evidence for the argument that individuals have a right to define their identity stems from the fact that discriminatory practices are inextricably linked to structural societal processes. The CERD Committee in its General Recommendation 34 explains that racism is structural and cultural in nature (par 5). In its report on Cuba, the CERD Committee highlighted that race is a social construct and that the classification of individuals based on race is an arbitrary practice(par 5). The background to this is that scientists created the idea of human difference and of a hierarchy of races in the 18thcentury (p 393). The CERD Committee's conceptualisation of race as a social construct illustrates that the prohibition of discrimination is designed to protect individuals from other individuals, and from the operations of institutional processes ascribing disadvantageous arbitrary meanings to their identities.

The CEDAW Committee in General Recommendation 28 similarly regards social processes as operating in a complex manner to produce a set of associations with a particular gender identity (par 5). The way in which society constructs the definition of gender identity gives rise to relationships between individuals characterised by hierarchy and power disparity (par 5). It follows from the general recommendations of the CERD and CEDAW Committees that discriminatory practices are inextricably linked to dominant groups defining membership of their group, membership of other groups and the relationship between different groups. What is more, the prohibition of discrimination protects groups from the operation of social processes which ascribe meanings to their identities by providing them with a right to define their identity. Of equal significance is that

the CEDAW Committee specifies in the General Recommendation 28 that equality between the sexes includes the freedom to make choices not limited by gender stereotypes (par 22). Implicit in the conceptualisation of equality as freedom from coercion associated with society ascribing identity to groups, is that groups have a right to define their identity. The consistency between how the CERD and CEDAW Committees envisage the nature of the protected characteristic and the associated protections provide support for the assertion that the prohibition of discrimination incorporates the right of individuals with a protected characteristic to define their identity. Along with a right to define their identity comes a right to control how other individuals represent their identity.

Diane Otto critiques the interpretation of the prohibition of discrimination by the United Nations treaty bodies as defining identity categories corresponding to protected characteristics in a dualistic manner (p 2). Otto argues that international human rights law should reconceptualise the identity of sex/gender as encompassing "multiple forms of expression and identification"(p 14). Otto's argument relating to the plural and fluid nature of gender identity is equally applicable to all categories covered by the prohibition of discrimination. Conceiving of individuals in terms of a fixed or binary identity category dehumanises them. Individuals lack an opportunity to understand and express the multidimensionality of their self-perception and experiences. The writings of womanists (feminists of colour) reveal the undesirable nature of making overgeneralisations about how all individuals who belong to a particular protected group experience and perceive life.

Where I depart from Diane Otto is that the prohibition of discrimination already enshrines the protection of human diversity in terms of individuals having a right to define their identity in a multidimensional manner and the experiences associated with possessing a protected characteristic. To illustrate, the purpose of the CEDAW is to challenge stereotypes associated with having a particular gender identity. CEDAW places the combat of stereotypes at its heart because society constructs institutional relationships around the definitions of a protected characteristic, such as gender identity. For instance, Article 5 of CEDAW states that in order to eliminate prejudices and stereotypes relating to possessing a particular gender identity, states should modify the social and cultural patterns of conduct of individuals. CEDAW illustrates that the conduct the prohibition of discrimination proscribes is closely linked to the meaning individuals give to a particular identity category. The prohibition of discrimination would be deprived of its purpose if groups with a protected characteristic lacked the right to define their identity in a multifaceted manner. This ensures that individuals and institutions pattern their conduct in a manner which reflects the actual lived experienced of individuals with a protected characteristic.

The representation of individuals in geometric space using numerical values as part of creating a model within the AI system violates the right of individuals with protected characteristics to define their identity in a multifaceted manner. The purpose of AI systems is to enable the user to optimise the achievement of a desired outcome. For instance, an employer could use the system to identify whether an applicant for the position of a sales person possesses characteristics associated with delivering high sales volumes. In order to produce predictions about future performance, the model uses information which is linked to the variable to be predicted. The model allocates individuals into groups based on the presence of correlations between the data. As a result, the model allocates individuals into groups based on the similarities the operation of the software detects. Anya Prins and Daniel Schwarcz explain that even when programmers exclude protected characteristics from being put into the model, the AI process will locate characteristics which serve as proxies for the protected characteristic (p 7-8). It will associate such characteristics with the predicted future performance. This stems from the inherent purpose of AI systems to seek out characteristics which correspond to the desired goal to be predicted (p 7-8). Due to the existence of inequality and income disparities, there is a likely link between the protected characteristic and the predicted score for future performance (p 16). Of course, depending on what the program is designed to predict, it will utilise different pieces of information and produce vastly different correlations between the relevant pieces of information. Presumably, the possible permutations of how various characteristics are linked to group membership are infinite. This is because it is possible to think of many purposes for which one can use AI systems to make predictions about future performance.

The representation of individuals in a model for the purpose of predicting their future performance is problematic from the standpoint of protecting human diversity. Since the model contains only that information which it treats as relevant for making a prediction, it does not display how contextual factors are linked to the possession of a particular characteristic or the value the model attaches to the characteristic. The representation of individuals in the model may diverge vastly from how a group perceives its identity. For instance, the program predicting the likelihood of the child succeeding in school does not account for whether a Roma child scored poorly on a school admission test because the child did not speak the native language. Neither does it take into account whether the test may reflect the cultural assumptions of the dominant group. It is known that

Roma children in the past obtained low scores on school admission tests for these reasons. The authorities placed them in special educational institutions. The model could designate the Roma children as poor performers due to grouping them together based on low scores. This example illustrates how the model can generate an inaccurate representation of Roma children as having low aptitude. Equally problematic, the program will seek out correlations between numerous characteristics of Roma children which it detects as being linked to their performance. The logic of AI systems differs from that of human beings. The system can generate false or spurious correlations. Such correlations within a model can produce misleading representations of groups. The modelling of individuals as part of algorithmic processes mirrors the past trend of society artificially constructing social categories of groups and defining groups. The construction of racial difference is just one example of such unjustifiable practices. The existence of a parallel between colonial practices and the modelling of individuals in geometrical space demonstrates that the process of representing individuals in a mathematical model can violate the right inherent in the prohibition of discrimination of groups to define their identity in a multidimensional manner and to control how they are represented.

The results of this submission may be viewed at:
http://www.humanrights.gov.au/admin/structure/webform/manage/submission_form_tech_2019/submission/8527