United Nations Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance
Call for submissions: Thematic report on new information technologies, racial equality and non-discrimination

13 December 2019

Introduction

1. Access Now welcomes this opportunity to submit input and provide relevant information to the United Nations Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance on new information technologies, racial equality and non-discrimination.[1] As artificial intelligence (AI) continues to find its way into our daily lives, its propensity to interfere with human rights only gets more severe. In this submission, Access Now seeks to provide relevant information on the topic. We particularly offer insight on three main areas: (1) international human rights law approaches to regulating new technologies and AI; (2) how new information technologies may entrench bias, including racial bias and; (3) how information technologies have affected the enjoyment of human rights (drawing on the example of racial bias in machine learning used in the United States (US) criminal justice system).

About Access Now

2. Through representation in 14 countries around the world, Access Now provides thought leadership and policy recommendations to the public and private sectors to ensure the continued openness of the internet and the protection of fundamental rights. We engage with an action-focused global community, convene stakeholders through the RightsCon Summit Series, and operate a 24/7 Digital Security Helpline that provides real-time direct technical assistance to at-risk individuals and communities worldwide. As an ECOSOC accredited organization, Access Now routinely engages with the United Nations in support of our mission to extend and defend human rights in the digital age.[2]

3. In recent years, Access Now has developed a position on AI governance, advocating for a human rights centric approach to AI. Our work includes a launch of the Toronto Declaration on protecting the rights to equality and non-discrimination in machine learning systems (the Toronto Declaration) -- a statement on the role and accountability of states and the private

---

[1] OHCHR, Call for submissions: Thematic report on new information technologies, racial equality and non-discrimination

[2] Access Now, Access Now About Us (2019).

sector where human rights harms arise.[3] The Toronto Declaration was spearheaded by Access Now and Amnesty International at RightsCon 2018 held in Toronto, Canada, and is endorsed by several organizations and individuals.[4]

4. Access Now's Europe Policy Manager, Fanny Hidvegi, was selected to join the European Union's High-level Expert Group on Artificial Intelligence (AI HLEG).[5] We published our preliminary recommendations to improve the Ethics Guidelines on Trustworthy AI and the positive and negatives of the Policy and Investment Recommendations for Trustworthy AI.[6]

5. Finally, Access Now has published two papers on AI. The first report, "Mapping Regulatory Proposals for Artificial Intelligence in Europe" covers regional strategies from the European Union and the Council of Europe as well as national plans from several member states including France, Finland, Germany and Italy.[7] In the report, Access Now lays out a criteria to assess AI strategies to make sure that the development and deployment of AI is individual-centric and human rights-respecting. The second report, "Human Rights in the Age of Artificial Intelligence," provides a comprehensive analysis on the potential pitfalls of AI and how to address AI-related human rights harms.[8]

Call for Submission

(1) International human rights law approaches to regulating new technologies and AI

6. Several lenses exist through which experts examine AI. Access Now recognizes that ethical principles, such as respect for human autonomy, prevention of harm, fairness, and explicability must be at the forefront of any AI guidelines, but they can only be a first step.[9] While ethical reasoning and individual conscience certainly have a role to play, we believe that it is important to stress the role of human rights. Specifically, whereas the violation of an ethical principle can perhaps be written off as collateral damage, human rights offer us a set of principles which command respect and adherence in all circumstances, as well as a framework for contesting norms and ensuring accountable implementation.[10]

---

[3] Moreover, in 2018, Access Now became a member of the Partnership on Artificial Intelligence *See* Access Now, Artificial Intelligence: We just became a member of the Partnership on AI! (15 May 2019).
[4] Access Now and Amnesty International, The Toronto Declaration Protecting the right to equality in machine learning (2019), Access Now, Access Now and Amnesty International launch Toronto Declaration on human rights and artificial intelligence (16 May 2018), Rich Haridy, Human rights groups call for protections against discriminatory and biased artificial intelligence, *New Atlas* (17 May 2018).
[5] European Commission, High-Level Expert Group on Artificial Intelligence (4 October 2019).
[6] Access Now, Laying down the law on AI: ethics done, now the EU must focus on human rights (8 April 2019); Access Now, European Union: more big words on AI, but where are the actions? (26 June 2019).
[7] Access Now, Mapping Regulatory Proposals for Artificial Intelligence in Europe (November 2018).
[8] Access Now, Human Rights in the Age of Artificial Intelligence (November 2018).
[9] Access Now, Artificial Intelligence: What role for the European Union? (9 May 2018), Access Now, AI Ethics guidance a first step but needs to be transformed into tangible rights for people (8 April 2019).
[10] Access Now, Laying down the law on AI ethics done Now the EU must focus on human rights (8 April 2019).

7. The use of international human rights law and its well-developed standards and institutions to examine AI systems can contribute to the conversations already happening, and provide a universal vocabulary and forums established to address power differentials. Beyond their embodiment in specific laws, human rights offer us a broad and well-defined set of principles to cover all instances in which our dignity and integrity are threatened.[11] Simply put, using ethics as a policy tool is not sufficient and it must be matched with the enforcement of existing and the development of new, binding safeguards.[12]

8. Universal human rights frameworks, such as the Universal Declaration of Human Rights are a cornerstone to protect individuals online and off.[13] In fact, the United Nations has formally affirmed that the same rights that people have offline must also be protected online.[14] Further, international guiding principles have been developed for the implementation of human rights in relation to business activities and should be applied in the context of AI. For example, the United Nations Guiding Principles on Business and Human Rights reiterates the government's duty to protect human rights, the corporate responsibility to respect human rights, and the need to guarantee access to remedy for victims of business-related abuses.[15] It is imperative to embed these human rights frameworks in every aspect of the deployment of AI.

(a) The Toronto Declaration: Calling on states and companies to uphold equality and non-discrimination in machine learning

9. The Toronto Declaration is a tangible means to uphold equality and non-discrimination and clarify the obligations of states and responsibilities of companies to protect human rights in the development and use of machine learning.[16] Based on binding international human rights law, the Toronto Declaration provides an established and widely accepted global ethics framework

---

[11] *Id.*

[12] Other experts and international organizations, such as the United Nations Educational, Scientific and Cultural Organization (UNESCO), have also highlighted the need for human rights and ethical standards on AI. During RightsCon 2018 in Toronto, Canada, UNESCO held an interactive session where participants stressed the urgent need to set ethical standards alongside human rights for these evolving technologies. *See* UNESCO, UNESCO highlights the need to set human rights and ethical standards for Artificial Intelligence at the RightsCon 2018 (29 May 2018).

[13] Universal Declaration of Human Rights (10 Dec 1948) UNGA Res 214 A (III) (1948).

[14] UN General Assembly, *The promotion, protection and enjoyment of human rights on the Internet*, UN Doc A/HRC/RES/38/7 (18 June - 6 July 2018), available online: https://undocs.org/A/HRC/RES/38/7 at para 1; *See also* UN General Assembly, *Promotion and protection of human rights and fundamental freedoms, including the rights to peaceful assembly and freedom of association*, UN Doc A/RES/73/173 available online: https://undocs.org/en/A/RES/73/173 at para 4.

[15] John Ruggie (2011) *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*. Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises, U.N. Doc. A/HRC/17/31 (Geneva, Switzerland: United Nations) at https://www.ohchr.org/Documents/Issues/Business/A-HRC-17-31_AEV.pdf, Access Now, Artificial Intelligence: What role for the European Union? (9 May 2018).

[16] Access Now, Access Now and Amnesty International launch Toronto Declaration on human rights and artificial intelligence (16 May 2018).

that is legally binding. It proposes that human rights law and standards are put front and center in existing and emerging conversations and methods analyzing the impact of machine learning and related technologies.

10. The Toronto Declaration focuses particularly on the right to equality and non-discrimination, a critical human right underpinning all others. It asserts that we must adhere to principles of inclusion, diversity and equity to ensure that machine learning systems do not create or perpetuate discrimination, particularly against already marginalized groups.[17] This is an urgent call in reaction to growing evidence of the risk of discriminatory harms associated with the use of machine learning systems in public and private use across many sectors, including policing, criminal justice, immigration and asylum -- to name a few.[18]

11. The Toronto Declaration makes specific recommendations to both private companies and governments. First, the Toronto Declaration suggests that private companies take into account the risk of bias being introduced into a system through incomplete or flawed machine learning training data. Second, the Toronto Declaration calls on governments to assess potential discriminatory outcomes when acquiring and deploying these technologies.

(2) How new information technologies may entrench existing racial inequalities, including through the use of datasets or metrics that already reflect racial biases

12. AI can be biased both at the system and the data or input level. Bias at the system level involves developers building their own personal biases into the parameters they consider or the labels they define.[19] Although this rarely occurs intentionally, unintentional bias at the system level is common. This often occurs in two ways. First, when developers allow systems to conflate correlation with causation. For example, people with a low income tend to have lower credit scores, for a variety of reasons. If a machine learning system used to build credit scores includes the credit scores of one's Facebook friends as a parameter, it will result in lower scores among those with low-income backgrounds, even if they have otherwise strong financial indicators, simply because of the credit scores of their friends.[20] Second, when developers chose to include parameters that are proxies for known bias. For example, although developers of an algorithm may intentionally seek to avoid racial bias by not including race as parameter, the algorithm will still have racially biased results if it includes common proxies for race like income, education, or postal code.[21]

13. Bias at the data or input level occurs in various ways.[22] These include, but are not limited to: use of historical data that is biased, when the input data is not representative of the target

---

[17] Access Now, We want AI to come to everybody's life - AI and human rights at the MWC (2018).
[18] Access Now, Access Now and Amnesty International launch Toronto Declaration on human rights and artificial intelligence (16 May 2018).
[19] *See* Access Now, Human Rights in the Age of Artificial Intelligence (November 2018) at page 12.
[20] Access Now, Human Rights in the Age of Artificial Intelligence (November 2018) at page 12.
[21] Access Now, Human Rights in the Age of Artificial Intelligence (November 2018) at page 12; Cathy O'Neil, "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy" 9780553418811: Amazon.Com: Books 115-60, accessed 13 May 2018, Weapons of Math Destruction Increases Inequality.
[22] *See* Access Now, Human Rights in the Age of Artificial Intelligence (November 2018) at page 12.

population, when the input data are poorly selected, and when the data are incomplete, incorrect or outdated. Bias at the data or input level occurs because machine learning systems use an existing body of data to identify patterns, any bias in that data is naturally reproduced. For example, a system used to recommend admissions at a top university that uses the data of previously admitted students to train the model is likely to recommend upper class males over women and traditionally underrepresented groups.[23] Second, selection bias (*i.e* when the input data are not representative of the target population), results in recommendations that favour certain groups over another. For example, if a GPS-mapping app used only input data from smartphone users to estimate travel times and distances, it could be more accurate in wealthier areas of cities that have a higher concentration of smartphone users, and less accurate in poorer areas or informal settlements, where smartphone penetration is lower and there is sometimes no official mapping.[24]

14. Unfortunately, biased data and biased parameters are the rule rather than the exception. Because data are produced by humans, the information carries all the natural human bias within it. Researchers have begun trying to figure out how to best deal with and mitigate bias, including whether it is possible to teach machine learning systems to learn without bias;[25] however, this research is still in its nascent stages. For the time being, there is no cure for bias in AI systems.[26]

(3) Examples of racial bias in machine learning systems: the US criminal justice system

15. Evidence suggests that the use of algorithms in the US for criminal justice system contains distinct racial bias.[27] From predictive policing to predicting recidivism, law enforcement agencies and courts are relying on algorithmic tools that perpetuate racial discrimination. For example, predictive tools such as PredPol, use "historical data, including arrest records and electronic police reports, to forecast areas where crime will occur and to help shape public safety strategies."[28] These tools are marketed as a means to provide objective, data-driven decisions; however, they are often trained on "data produced during flawed, racially biased, and sometimes unlawful practices and policies."[29]

16. Tools that predict recidivism work similarly as well. Upon arrest in the US system, it is common for suspects to get attributed a score aimed at predicting the likelihood that they will commit a

---

[23] Access Now, Human Rights in the Age of Artificial Intelligence (November 2018) at page 12. For more information on the gender of surveillance *See* Access Now, The gender of surveillance: how the world can work together for a safer internet (6 February 2018).

[24] Access Now, Human Rights in the Age of Artificial Intelligence (November 2018) at page 12.

[25] This is broadly known as the FATML community, "Fairness, Accountability and Transparency for Machine Learning" *See* https://www.fatml.org/ (2018).

[26] Access Now, Human Rights in the Age of Artificial Intelligence (November 2018).

[27] *See* Julia Angwin, Jeff Larosn, Surya Mattu and Lauren Kirchner Machine Bias: THere's software used across the country to predict future criminals. And its biased against blacks *Pro-Publica* (23 May 2016).

[28] Karen Hao, Police across the US are training crime-predicting AIs on falsified data, *MIT Technology Review* (13 February 2019).

[29] Rashida Richardson, Jason Schultz, & Kate Crawford, Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice, *NYU Law Review Online* (26 April 2019).

crime in the future.[30] One of the systems that is most commonly used was developed by a company called COMPAS, the Correctional Offender Management Profiling for Alternative Sanctions.[31] The "risk assessment" score that the COMPAS system provides is used to inform decisions about whether a suspect can be set free at each stage of the criminal justice process, from setting bonds to assess the likelihood for acceptance of a plea bargain. An investigation by ProPublica revealed that the system has a distinct racial bias.[32] It has falsely flagged black defendants as future criminals at nearly twice the rate as white defendants. At the same time, white defendants were mislabeled as low risk more often than black defendants. Because of these scores, which reflect observable prejudice, it is not uncommon for suspects to plead guilty even if they are innocent, or to accept longer sentences in order to get out of pre-trial detention. The scoring method and the design of the system may have made the criminal court system more "productive" by limiting the number of cases going to trial, but that benefit seems to come at the expense of fairness, justice and equality.[33]

![accessnow logo]

Access Now (https://www.accessnow.org) defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots grantmaking, and convenings such as RightsCon, we fight for human rights in the digital age.

For more information, contact:
Peter Micek
General Counsel
Access Now | www.accessnow.org
Peter@accessnow.org

---

[30] Access Now, Artificial Intelligence: What role for the European Union? (9 May 2018); Julia Angwin, Jeff Larosn, Surya Mattu and Lauren Kirchner Machine Bias: There's software used across the country to predict future criminals. And its biased against blacks *Pro-Publica* (23 May 2016).

[31] Northpointe Inc., COMPAS Risk & Need Assessment System (2012); Access Now, Artificial Intelligence: What role for the European Union? (9 May 2018).

[32] Mattu and Lauren Kirchner, Pro-Publica, Machine Bias: There's software used across the country to predict future criminals. And its biased against blacks (23 May 2016).

[33] Dylan Walsh, Why U.S. Criminal Courts are So Dependent on Plea Bargaining *The Atlantic* (2 May 2017). *See* also Vincent Southerland, With AI and Criminal Justice, the Devil is in the Data *ACLU* (9 April 2018).