The Toronto Declaration: Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems is a declaration that advocates responsible practices for machine learning practitioners and governing bodies. It is a joint statement issued by groups including Amnesty International and Access Now , with other notable signatories including Human Rights Watch and The Wikimedia Foundation . [1] It was published at RightsCon on May 16, 2018. [2] [3]

The Declaration focuses on concerns of algorithmic bias and the potential for discrimination that arises from the use of machine learning and artificial intelligence in applications that may affect people's lives, "from policing, to welfare systems, to healthcare provision, to platforms for online discourse." [4] A secondary concern of the document is the potential for violations of information privacy .

The goal of the Declaration is to outline "tangible and actionable standards for states and the private sector." [5] The Declaration calls for tangible solutions, such as reparations for the victims of algorithmic discrimination. [6]

Contents

The Toronto Declaration consists of 59 articles, broken into six sections, concerning international human rights law , duties of states , responsibilities of private sector actors, and the right to an effective remedy .

Preamble

The document begins by asking the question, "In a world of machine learning systems, who will bear accountability for harming human rights ?" [4] It argues that all practitioners, whether in the public or private sector, should be aware of the risks to human rights and approach their work with human rights in mind – conscious of the existing international laws, standards, and principles. The document defines human rights to include "the right to privacy and data protection, the right to freedom of expression and association, to participation in cultural life, equality before the law, and access to effective remedy"; [4] but it states that the Declaration is most concerned with equality and non-discrimination.

Using the framework of international human rights law

The framework of international human rights law enumerates various rights, provides mechanisms to hold violators to account, and ensures remedy for the violated. The document cites the United Nations Human Rights Committee 's definition of discrimination as "any distinction, exclusion, restriction or preference which is based on any ground [including but not limited to] race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status, and which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise by all persons, on an equal footing, of all rights and freedoms." [7]

Governments should proactively create binding measures, and private entities should create internal policies, to protect against discrimination. Measures may include protections for sensitive data, especially for vulnerable populations. Systems should be designed in collaboration with a diverse community in order to prevent discrimination in design.

Duties of states: human rights obligations

Governments today are deploying machine learning systems, often in collaboration with private entities. Even when development is contracted to such third parties, governments retain their obligation to protect human rights. Before implementation, and on an ongoing basis thereafter, they should identify risks and conduct regular audits, then take all necessary measures to mitigate these risks. They should be transparent about how machine learning is implemented and used, avoiding black box systems whose logic cannot be easily explained. Systems should be subject to strict oversight from diverse internal committees and independent judicial authorities.

Governments must also protect citizens from discrimination by private entities. In addition to oversight, they should pass binding laws against discrimination, as well as for data protection and privacy, and they should provide effective means to remedy for affected individuals. It is important for national and regional governments to expand on and contextualize international law.

Responsibilities of private sector actors: human rights due diligence

Private entities are responsible for conducting "human rights due diligence." Just like governments, private entities should identify risks before development by considering common risks and consulting stakeholders, "including affected groups, organizations that work on human rights, equality and discrimination, as well as independent human rights and machine learning experts." [4] They should design systems that mitigate risks, subject systems to regular audits, and forego projects that carry too high of risks. They should be transparent about assumed risks, including details of the technical implementation where necessary, and should provide a mechanism for affected individuals to dispute any decisions that affect them.

The right to an effective remedy

"The right to justice is a vital element of international human rights law." [4] Private entities should create processes for affected individuals to seek remedy, and they should designate roles for who will oversee these processes. Governments must be especially cautious when deploying machine learning systems in the justice sector. Transparency, accountability, and remedy can help.

References