ISBN 978-93-5445-447-9

Legal Dimensions of Human Rights in the Technological Realm

— Edited by —
Wasantha Seneviratne
Sourabh Ubale

— PUBLISHED JOINTLY BY —

THE CENTRE FOR THE STUDY OF HUMAN RIGHTS (CSHR), FACULTY OF LAW, UNIVERSITY OF COLOMBO (SRI LANKA)

MARATHWADA MITRA MANDAL'S SHANKARRAO CHAVAN LAW COLLEGE, PUNE (AFFILIATED TO SAVITRIBAI PHULE PUNE UNIVERSITY, & ACCREDITED 'A' GRADE BY NAAC) (INDIA)

THE TORONTO DECLARATION ON PROTECTING THE RIGHTS TO EQUALITY AND NON-DISCRIMINATION IN MACHINE LEARNING SYSTEMS - AN ANALYSIS

Dr. Vikas Bhatnagar*

ABSTRACT : With the advancement in technology and artificial intelligence, it has not only globalized the world but also highlighted certain issues. One such problem is violation of human rights amidst this digital age which needs to be safeguarded. To address this, human rights organizations launched a framework of international human rights law to be applied to machine learning systems. With an aim that it will be adopted by civil society, private sector, and government stakeholders, "The Toronto Declaration on Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems" was initiated. The basic aim is that the technological advances must not weaken our human rights. The research paper deals with human rights framework related to artificial intelligence. It tries to analyze the draft thoroughly. An emerging concept of Private sector's responsibility towards human rights due diligence has also been highlighted that suggests three core steps to the process of human rights due diligence as well.

KEYWORDS: *Artificial Intelligence, Digital Age, Toronto Declaration, Human Rights, Due Diligence.*

A. INTRODUCTION

Today's era is considered as the era of science and technology. In fact, we are surrounded by machines . Initially scientific research and inventions were introduced to reduce our efforts and to save our energy but, day by day we are completely dependent on technology . Nowadays, situation is different, we are one step ahead regarding the use of technology, this era is considered as era of Artificial Intelligence. The meaning of Artificial Intelligence is the development of computer systems that are able to perform tasks that would require human intelligence. Examples of these tasks are visual perception, speech recognition, decision-making, and translation between languages.

In India for the purpose of making use of the Government scheme and other benefits, for the purpose of Census, economic analysis etc. the Government is collecting data. For

* Assistant Professor, Marathwada Mitra Mandal's Shankarrao Chavan Law College, Pune.

data collection along with Government agencies some private agencies are also involved in analysis. These agencies collect and analyze data through AI's like computer soft ware's and for that they have access to personal information of an Individual and his family. For the purpose of internet banking we provide information to bank software, to keep in touch we have account on Face book or WhatsApp, there is possibility of leakage of data by these agencies, which is a violation of Right to privacy. One big example is of Facebook–Cambridge Analytica data scandal.[1] Thus, Privacy per se is an undisputed right of an individual, comprehensively bonded with his/her life. Privacy becomes matter of controversy when it is mishandled/intruded illegitimately by erroneous individuals.

Article 41 of the Constitution provides that "the State shall within the limits of its economic capacity and development, make effective provision for securing the right to work, to education and to public assistance in cases of unemployment, old age, sickness and disablement, and in other cases of undeserved want". No doubt, technology is playing important role in reducing human efforts but dropping employment of people who are engage in different employment whether related to professional or labor. For an example:

1. Use of robots in surgery replacing Doctor.
2. Use of ATM, Printing machine of Net Banking replaces man power of bank.
3. Use of computers and internet connection by students replacing traditional teachers.
4. Use of techniques in replacing labor of e. g. Agro/ automobile industries.

We can't say that, because of AI there is clearly violation of freedom of speech and expression but up to some extent it's effects. e .g on a click I can connect myself to any part of the world. I can find anything, I can search solution of my problem, although it's good but it would disconnect me from my friends, relatives, mentor, teachers etc. from whom I can get all these things after discussion and with a different view.

B. A NEW HUMAN RIGHTS FRAMEWORK[2]

With the wider use of technologies like AI, experts call for ethical standards and new legislative frameworks ensuring citizen's protection from potential threats. Hence, new specific regulatory frameworks are in development:

---

1 The Facebook–Cambridge Analytica data scandal was a major political scandal in early 2018 when it was revealed that Cambridge Analytica had harvested the personal data of millions of people's Facebook profiles without their consent and used it for political advertising purposes.

2 Europe and Central Asia (Aug.14, 2020, 1:10 PM), https://www.eurasia.undp.org

1. The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems, launched on May 16, 2018 at Rights on, reaffirms the role of human rights law and standards in protecting individuals and groups from discrimination and non-equality in developing ethical frameworks for machine learning.
2. Following personal data protection concerns, the European Union has adopted a new data protection framework called General Data Protection Regulation (GDPR) to improve data privacy regulation across every sector including healthcare, banking and many others.
3. UN's principles on Business and Human Rights provide a set of standards, which also needs to be applied to AI, for preventing and addressing human rights violations linked to business activities.
4. Council of Europe European Commission for the efficiency of justice (CEPEJ) has recently set up the multidisciplinary team of experts who prepare the guidelines for the ethical use of algorithms within justice systems, including predictive justice.

C. THE TORONTO DECLARATION ON PROTECTING THE RIGHTS TO EQUALITY AND NON-DISCRIMINATION IN MACHINE LEARNING SYSTEMS (EDITED) [3]

We all have access to large data sets and powerful computers, machine learning and artificial intelligence which may be beneficial for society but at the same time can pose serious risks to human rights .To keep in mind all these things a declaration has been launched in 'Toronto' on May 16, 2018 by Human right group Amnesty International and Access Now *and it has been endorsed by Human Rights Watch and Wikimedia Foundation. It is also known as "Toronto declaration* : Protecting the rights to equality and non-discrimination in machine learning systems". The Declaration emphasizes the centrality and applicability of human rights law, which is designed to protect rights and provide remedies where human beings are harmed. The declaration also urge to the government, Public and private companies that - "We must keep our focus on how these technologies will affect individual human beings and human rights and in a world of machine learning systems, who will bear accountability for harming human rights?".

[3] The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems (Aug. 14, 2020, 1:30 PM), https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/

a. PREAMBLE

1. As machine learning systems advance in capability and increase in use, we must examine the impact of this technology on human rights. We acknowledge the potential for machine but we must urgently address how these technologies will affect people and their rights. In a world of machine learning systems, who will bear accountability for harming human rights?
2. As discourse around ethics and artificial intelligence continues, this Declaration aims to draw attention to the relevant and well-established framework of international human rights law and standards. Human rights are "universal, indivisible and interdependent and interrelated."
3. This Declaration aims to build on existing discussions, principles and papers exploring the harms to Human rights arising from this technology. The human rights law and standards referenced in this Declaration provide solid foundations for developing ethical frameworks for machine learning, including provisions for accountability and means for remedy.
4. There is a substantive and growing body of evidence to show that machine learning systems, which can be opaque and include unexplainable processes, can contribute to discriminatory or otherwise repressive practices if adopted and implemented without necessary safeguards.
5. States and private sector actors should promote the development and use of machine learning and related technologies where they help people exercise and enjoy their human rights. States should promote the positive right to the enjoyment of developments in science and technology as an affirmation of economic, social and cultural rights.[4]
6. There are numerous other human rights that may be adversely affected through the use and misuse of machine learning systems, including the right to privacy and data protection, the right to freedom of expression and association but we focus in this Declaration on the right to equality and non-discrimination
7. While this Declaration is focused on machine learning technologies, many of the norms and principles included here are equally applicable to technologies housed under the broader term of artificial intelligence, as well as to related data systems.

[4] The International Covenant on Economic, Social and Cultural Rights (ICESCR), art.15.

b. USING THE FRAMEWORK OF INTERNATIONAL HUMAN RIGHTS LAW

8. States have obligations to promote, protect and respect human rights; private sector actors, including companies, have a responsibility to respect human rights at all times. We put forward this Declaration to affirm these obligations and responsibilities.

9. There are many discussions taking place focusing on the ethics of artificial intelligence and how to make technology in this field human-centric.

10. Human rights law sets standards and provides mechanisms to hold public and private sector actors accountable where they fail to fulfill their respective obligations and responsibilities to protect and respect rights. It also requires that everyone must be able to obtain effective remedy and redress where their rights have been denied or violated.

11. The risks that machine learning systems pose must be urgently examined and addressed at governmental level and by private sector actors who are conceiving, developing and deploying these systems. It is critical that potential harms are identified and addressed and that mechanisms are put in place to hold those responsible for harms to account. Government measures should be binding and adequate to protect and promote rights.

c. THE RIGHT TO EQUALITY AND NON-DISCRIMINATION

12. This Declaration focuses on the right to equality and non-discrimination, a critical principle that underpins all human rights.

13. Discrimination is defined under international law as "any distinction, exclusion, restriction or preference which is based on any ground such as race, color, 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."

d. PREVENTING DISCRIMINATION

14. Governments have obligations and private sector actors have responsibilities to proactively prevent discrimination in order to comply with existing human rights law and standards. When prevention is not sufficient or satisfactory, and discrimination arises, a system should be interrogated and harms addressed immediately.

15. In employing new technologies, both state and private sector actors will likely need to find new ways to protect human rights, as new challenges to equality and representation of an impact on diverse individuals and groups arise.

16. All actors, public and private, must prevent and mitigate against discrimination risks in the design, development and application of machine learning technologies. They must also ensure that there are mechanisms allowing for access to effective remedy in place before deployment and throughout a system's lifecycle.

e. PROTECTING THE RIGHTS OF ALL INDIVIDUALS AND GROUPS: PROMOTING DIVERSITY AND INCLUSION

17. This Declaration underlines that inclusion, diversity and equity are key components of protecting and upholding the right to equality and non-discrimination.

18. While the collection of data can help mitigate discrimination, there are some groups for whom collecting data on discrimination poses particular difficulty. Additional protections must extend to those groups, including protections for sensitive data.

19. Implicit and inadvertent bias through design creates another means for discrimination, where the conception, development and end use of machine learning systems is largely overseen by a particular sector of society.

20. Inclusion, diversity and equity entails the active participation of, and meaningful consultation with, a diverse community, including end users, during the design and application of machine learning systems, to help ensure that systems are created and used in ways that respect rights – particularly the rights of marginalized groups who are vulnerable to discrimination.

f. DUTIES OF STATES: HUMAN RIGHTS OBLIGATIONS

21. States bear the primary duty to promote, protect, respect and fulfill human rights. Under international law, states must not engage in, or support discriminatory or otherwise rights-violating actions or practices when designing or implementing machine learning systems in a public context or through public-private partnerships.

22. States must adhere to relevant national and international laws and regulations that codify and implement human rights obligations protecting against discrimination and other related rights harms, for example data protection and privacy laws.

23. States have positive obligations to protect against discrimination by private sector actors and promote equality and other rights, including through binding laws.

24. The state obligations outlined in this section also apply to public use of machine learning in partnerships with private sector actors.

g. STATE USE OF MACHINE LEARNING SYSTEMS

25. States must ensure that existing measures to prevent against discrimination and other rights harms are updated to take into account and address the risks posed by machine learning technologies.

26. Machine learning systems are increasingly being deployed or implemented by public authorities in areas that are fundamental to the exercise and enjoyment of human rights, rule of law, due process, freedom of expression, criminal justice, healthcare, access to social welfare benefits, and housing. While this technology may offer benefits in such contexts, there may also be a high risk of discriminatory or other rights-harming outcomes. It is critical that states provide meaningful opportunities for effective remediation and redress of harms where they do occur.

27. As confirmed by the Human Rights Committee, Article 26 of the International Covenant on Civil and Political Rights "prohibits discrimination in law or in fact in any field regulated and protected by public authorities".[5] This is further set out in treaties dealing with specific forms of discrimination, in which states have committed to refrain from engaging in discrimination, and to ensure that public authorities and institutions "act in conformity with this obligation".

28. States must refrain altogether from using or requiring the private sector to use tools that discriminate, lead to discriminatory outcomes, or otherwise harm human rights.

D. STATES MUST TAKE THE FOLLOWING STEPS TO MITIGATE AND REDUCE THE HARMS OF DISCRIMINATION FROM MACHINE LEARNING IN PUBLIC SECTOR SYSTEMS

Any state deploying machine learning technologies must thoroughly investigate systems for discrimination and other rights risks prior to development or acquisition, where possible, prior to use, and on an ongoing basis throughout the lifecycle of the technologies, in the contexts in which they are deployed. This may include:

5 Un Human rights committee , General comment 18 (1989) Para 12.

1. Conducting regular impact assessments prior to public procurement, during development, at regular milestones and throughout the deployment and use of machine learning systems to identify potential sources of discriminatory or other rights-harming outcomes.
2. Taking appropriate measures to mitigate risks identified through impact assessments – for example, mitigating inadvertent discrimination or under representation in data or systems; conducting dynamic testing methods and pre-release trials; ensuring that potentially affected groups and field experts are included as actors with decision-making power in the design, testing and review phases; submitting systems for independent expert review where appropriate.
3. Subjecting systems to live, regular tests and audits; interrogating markers of success for bias and self-fulfilling feedback loops; and ensuring holistic independent reviews of systems in the context of human rights harms in a live environment.
4. Disclosing known limitations of the system in question - for example, noting measures of confidence, known failure scenarios and appropriate limitations of use.

Ensure Transparency and Accountability: States must ensure and require accountability and maximum possible transparency around public sector use of machine learning systems. This must include explain ability and intelligibility in the use of these technologies so that the impact on affected individuals and groups can be effectively scrutinized by independent entities, responsibilities established, and actors held to account. States should Publicly disclose where machine learning systems are used in the public sphere, provide information that explains in clear and accessible terms how automated and machine learning decision-making processes are reached, and document actions taken to identify, document and mitigate against discriminatory or other rights-harming impacts.

Enforce Oversight : States must take steps to ensure public officials are aware of and sensitive to the risks of discrimination and other rights harms in machine learning systems. States should:

1. Proactively adopt diverse hiring practices and engage in consultations to assure diverse perspectives so that those involved in the design, implementation, and review of machine learning represent a range of backgrounds and identities.
2. Ensure that public bodies carry out training in human rights and data analysis for officials involved in the procurement, development, use and review of machine learning tools.

As research and development of machine learning systems is largely driven by the private sector, in practice states often rely on private contractors to design and implement these technologies in a public context. In such cases, states must not relinquish their own obligations around preventing discrimination and ensuring accountability and redress for human rights harms in the delivery of services.

Any state authority procuring machine learning technologies from the private sector should maintain relevant oversight and control over the use of the system, and require the third party to carry out human rights due diligence to identify, prevent and mitigate against discrimination and other human rights harms, and publicly account for their efforts in this regard.

Promoting Equality: States have a duty to take proactive measures to eliminate discrimination[6]. In the context of machine learning and wider technology developments, one of the most important priorities for states is to promote programs that increase diversity, inclusion and equity in the science, technology. States should also invest in research into ways to mitigate human rights harms in machine learning systems. Holding private sector actors to account: International law clearly sets out the duty of states to protect human rights; this includes ensuring the right to non-discrimination by private sector actors. According to the UN Committee on Economic, Social and Cultural Rights, "States parties must therefore adopt measures, which should include legislation, to ensure that individuals and entities in the private sphere do not discriminate on prohibited grounds". States should put in place regulation compliant with human rights law for oversight of the use of machine learning by the private sector in contexts that present risk of discriminatory or other rights-harming outcomes, recognizing technical standards may be complementary to regulation. In addition, non-discrimination, data protection, privacy and other areas of law at national and regional levels may expand upon and reinforce international human rights obligations applicable to machine learning. States must guarantee access to effective remedy for all individuals whose rights are violated or abused through use of these technologies.

6 The UN Committee on Economic, Social and Cultural Rights affirms that in addition to refraining from discriminatory actions, "State parties should take concrete, deliberate and targeted measures to ensure that discrimination in the exercise of Covenant rights is eliminated." – UN Doc. E/C.12./GC/20(2009) Para 36

E. RESPONSIBILITIES OF PRIVATE SECTOR ACTORS: HUMAN RIGHTS DUE DILIGENCE

Private sector actors have a responsibility to respect human rights; this responsibility exists independently of state obligations.[7] As part of fulfilling this responsibility, private sector actors need to take ongoing proactive and reactive steps to ensure that they do not cause or contribute to human rights abuses – a process called 'human rights due diligence'.[8] Private sector actors that develop and deploy machine learning systems should follow a human rights due diligence framework to avoid fostering or entrenching discrimination and to respect human rights more broadly through the use of their systems. There are three core steps to the process of human rights due diligence:

1. Identify Potential Discriminatory Outcomes: During the development and deployment of any new machine learning technologies, non-state and private sector actors should assess the risk that the system will result in discrimination. The risk of discrimination and the harms will not be equal in all applications, and the actions required to address discrimination will depend on the context. Actors must be careful to identify not only direct discrimination, but also indirect forms of differential treatment which may appear neutral at face value, but lead to discrimination. When mapping risks, private sector actors should take into account risk commonly associated with machine learning systems – for example, training systems on incomplete or unrepresentative data, or datasets representing historic or systemic bias. Private actors should consult with relevant stakeholders in an inclusive manner, including affected groups, organizations that work on human rights, equality and discrimination, as well as independent human rights and machine learning experts.

2. Take Effective Action To Prevent and Mitigate Discrimination and Track Responses: After identifying human rights risks, the second step is to prevent those risks. For developers of machine learning systems, this requires:

a. Correcting for discrimination, both in the design of the model and the impact of the system and in deciding which training data to use.

b. Pursuing diversity, equity and other means of inclusion in machine learning development teams, with the aim of identifying bias by design and preventing inadvertent discrimination.

---

7 UN Guiding Principles on Business and Human Rights and additional supporting documents

8 Council of Europe's Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries.

c. Submitting systems that have a significant risk of resulting in human rights abuses to independent third-party audits.

Where the risk of discrimination or other rights violations has been assessed to be too high or impossible to mitigate, private sector actors should not deploy a machine learning system in that context. Another vital element of this step is for private sector actors to track their response to issues that emerge during implementation and over time, including evaluation of the effectiveness of responses.

3. Be Transparent About Efforts To Identify, Prevent And Mitigate Against Discrimination In Machine Learning Systems: Transparency is a key component of human rights due diligence, and involves "communication, providing a measure of transparency and accountability to individuals or groups who may be impacted and to other relevant stakeholders." Private sector actors that develop and implement machine learning systems should disclose the process of identifying risks, the risks that have been identified, and the concrete steps taken to prevent and mitigate identified human rights risks.

F. THE RIGHT TO AN EFFECTIVE REMEDY

The right to justice is a vital element of international human rights law. Under international law, victims of human rights violations or abuses must have access to prompt and effective remedies, and those responsible for the violations must be held to account. Companies and private sector actors designing and implementing machine learning systems should take action to ensure individuals and groups have access to meaningful, effective remedy and redress. This may include, for example, creating clear, independent, visible processes for redress following adverse individual or societal effects, and designating roles in the entity responsible for the timely remedy of such issues subject to accessible and effective appeal and judicial review. The use of machine learning systems where people's rights are at stake may pose challenges for ensuring the right to remedy. The opacity of some systems means individuals may be unaware how decisions which affect their rights were made, and whether the process was discriminatory. In some cases, the public body or private sector actors involved may itself be unable to explain the decision-making process. The challenges are particularly acute when machine learning systems that recommend, make or enforce decisions are used within the justice system, the very institutions which are responsible for guaranteeing rights, including the right to access to effective remedy. The measures already outlined around identifying, documenting, and responding to discrimination, and being transparent and accountable about these efforts, will help states to ensure that individuals have access to effective remedies.

G. CONCLUSION

The signatories of this Declaration call for public and private sector actors to uphold their obligations and responsibilities under human rights laws and standards to avoid discrimination in the use of machine learning systems where possible. Where discrimination arises, measures to deliver the right to effective remedy must be in place. States must take meaningful measures to promote accountability and human rights, including, but not limited to the right to equality and non-discrimination, as per their obligations and responsibilities under international human rights law and standards. Technological advances must not weaken our human rights. We are at a crossroads where those with the power must act now to protect human rights, and help safeguard the rights that we are all entitled to now, and for future generations.

□ □