DE JURE NEXUS LAW JOURNAL

Author:

Ashwin Panicker

NLU, Kochi, 3rd Year.

BRIDGING THE GAP BETWEEN A.I. AND HUMAN RIGHTS

I. INTRODUCTION

In the span of 70 years, nine major human rights treaties, regional rights instruments in Europe, Africa and America have culminated[1], with the enforcement of international human rights law differing in variable magnitudes.[2] Thus, a human rights substructure has been put in place to defend human rights as a direct result. International human rights have come to represent shared global values even though the development of human rights has had its own share of politics in the past.

It can be argued that the foundation of Artificial Intelligence (hereinafter, A.I.) governance were laid down in the form of Human Rights Law.The information imbibed from humungous amounts of historical training data in order to identify patterns and probabilities transform into effective decision making.[3]The core of A.I. debates lie in the emergence of statistical miscalculations

[1] For an overview of regional human rights implementation in the Americas, Europe, and Africa, see David C. Baluarte and Christian De Vos, From Judgment to Justice: Implementing International and Regional Human Rights Decisions, Open Society Justice Initiative (November 2010), https:// www.opensocietyfoundations.org/sites/default/files/from-judgment-to-justice-20101122.pdf

[2] For an overview of the challenges of implementation, see International Institutions and Global Governance Program, "The Global Human Rights Regime," Council on Foreign Relations, May 11, 2012, Web, August 31, 2018.

[3] Machine Learning: The Power and Promise of Computers That Learn by Example, The Royal Society (April 2017): 19, https://royalsociety.org/~/media/policy/ projects/machine-learning/publications/machine-learning-report.pdf.

which can be adjudged as erroneous at the first blush.[4]However, in the event of incomplete historical data, these biases can quickly blur the lines and enter the domain of being discriminatory in the form of statistical biases. Such systems can further entrench prejudiced outcomes in people's lives. For instance, women with darker skin were denied recognition due to a lack of adequate training data which puts a huge cloud on the efficacy of facial recognition system as they regenerate historically instilled biases against people of color.[5]So the union of A.I. and humans climaxing into a happy marriage required the addressal of ethical and legal implications that data science entails.

The purpose of this article is to elucidate the impediment in the applicability of Human Rights Law into the convoluted world of A.I. and this article is an attempt to understand the extent to which the humans and A.I. can be held accountable under the same laws.

II. ALIGNING THE MORALITY OF A.I. WITH HUMANS

The potential moral consequences of A.I. can be subsumed under the age-old opposing ideologies of Immanuel Kant and David Hume regarding *Rationality* and *Morality*. Kant's theory of morality being derived out of rationality (Categorical Imperative) rests on a simple test of generalization.[6]For instance, stealing and lying could not be generalized, and not permitted as there would be no property to begin with, if everybody stole, and no communication, if everybody reserved the right to lie.So if a certain action does not hold up in the event everybody chose to do it, such action would not be allowed. In essence, any intelligent being would fall into a contradiction with itself by violating other rational beings. It is only our rational choosing that gives any value to anything in the first place and human reason is the sole source of any value. If this theory is applied to A.I.,it can be a true role-model for ethical behavior. A.I. might bridge the

[4] A vibrant community of academic researchers and practitioners are focused on fairness, accountability, and transparency. See, e.g., Proceedings of Machine Learning Research, vol. 81 (February 2018), http://proceedings.mlr.press/v81 .

[5] See, e.g., Joy Buolamwini and TimnitGebru "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of Machine Learning Research 81:1–15, 2018

[6]IMMANUEL KANT, GROUNDWORK FOR THE METAPHYSICS OF MORALS (Arnulf Zweig trans., Thomas E. Hill, Jr. &Arnulf Zweig eds., (2002) (1785).

gap that opens when humans with their Stone-Age, small-group-oriented DNA try to operate in a global context since human nature intensely parochial in its judgements.[7]

But on the other hand, on a careful perusal of the David Hume's theory of rationality and morality operating in stark contrasts, it can be presumed that the A.I. could have just about any type of value commitment in that scenario, including ones that would strike humans as rather absurd.[8]The possibility of such values being misguided or detrimental to humankind is not preposterous after all. But in any case, the existence of some sense of morality among A.I. cannot be contested whether it is derived from rationality or not.

Hobbes does not believe in the Kantian picture and had reasons to believe that these individuals would not act on shared values just by thinking clear-mindedly and that they would quickly experience the nastiness of life without a shared authority.[9]Individuals would feel compelled to strike against each other out of sheer necessity, in anticipation of future wrongs. The idea here is that even if one side is ready to give benefit of doubt to the other side, the lack of assurance of the other side offering them the same benefit is a compelling reason to strike first.

Another notable view is of Prof. T.M. Scanlon, who believes the morality of A.I. is limited to responses to the social setting it would be subjected to.[10] In case of a joint existence of animals and humans, the social setting has transmuted the behaviour of animals towards human in an appropriately respectful way.[11] The same analogy can be applied as an A.I. outperforming

[7] Steve Petersen, *Superintelligence as Superethical*, *in* ROBOT ETHICS 2.0: FROM AUTONOMOUS CARS TO ARTIFICIAL INTELLIGENCE 322 (Patrick Lin, Keith Abney, & Ryan Jenkins eds., 2017); Chalmers, *supra* note 9. *See also* Kahneman, *supra* note 2.

[8] First apparently in Nick Bostrom, Ethical Issues in Advanced Artificial Intelligence, in COGNITIVE, EMOTIVE AND ETHICAL ASPECT S OF DECISION MAKING IN HUMANS AND IN ARTIFICIAL INTELLIGENCE (George Eric Lasker, Wendell Wallach, Iva Smit, eds., 2003).

[9] THOMAS HOBBES, LEVIATHAN (1651).

[10] T. M. Scanlon, *What is Morality?*, *in* THE HARVARD SAMPLER: LIBERAL EDUCATION FOR THE TWENTY- FIRST CENTURY (Jennifer M Shephard, Stephen Michael Kosslyn, & Evelynn Maxine Hammonds eds., 2011).

[11] For speculation on what such mixed societies could be like, see MAX TEGMARK, LIFE 3.0: BEING HUMAN IN THE AGE OF ARTIFICIAL INTELLIGENCE 161 (2017).

humans in all fundamental functions of a social life is a conceivable future and the possibility of humans getting protections from A.I. entirely depends on the social conditions an A.I. is put in.

III. A.I. BASED DISCRIMINATION AROUND THE WORLD

Discriminatory algorithms for predicting the plausibility of a criminal being habitual offender are already used as a tool by the judges in some countries.[12]Classification on the basis of social characteristics is already in place in China.[13]The social trustworthiness of a citizen is calibrated on the basis of the data collected on each citizen. Displaying the faces of faulty debtors in public places or denying them to boo flights or trains is all based on the information originating from A.I.[14] Furthermore, in South Africa, A.I. based classification was exploited to bring into effect the inhumane policies of apartheid reign. Al these incidents serve as an important cautionary tale for any widespread deployment of AI social scoring systems.[15]AI systems built for mundane bureaucratic operations can be very well re-engineered to enact discriminatory policies of control, in the absence of proper firewalls to prevent such abuse.

IV. PROTECTION AGAINST BIASED DECISION MAKING OF A.I. ALGORITHMS

Discrimination in A.I. algorithms were raised as a human rights crisis in a World Economic Forum (WEF) report and it suggested viable solutions for the same in the form of

[12] Jeff Larson, Surya Mattu, Lauren Kirchner, and Julia Angwin, "How We Analyzed the COMPAS Recidivism Algorithm," ProPublica, May 23, 2016, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

[13] "Big Brother Is Watching: How China Is Compiling Computer Ratings on All Its Citizens," South China Morning Post International Edition, November 24, 2015, https://www.scmp.com/news/china/ policies-politics/article/1882533/big-brother-watching-how-china-compiling-computer

[14] Meg Jing Zeng, "China's Social Credit System Puts Its People Under Pressure to Be Model Citizens," The Conversation, January 23, 2018, https://theconversation.com/chinas-social-creditsystem-puts-its-people-under-pressure-to-be-model-citizens-89963.

[15] Geoffrey C. Bowker and Susan Leigh Star, Sorting Things Out — Classification and Its Consequences (MIT Press: 1999).

recommendations.[16] Companies were asked to keep a check on the compliance of human rights on a consistent level and ensure the performance of rights-based due diligence.

The Toronto Declaration: Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems was organised byAmnesty International and Access Now in May 2018.[17]It put the focus on A.I. bias on a global platform and outlined the duties of both State and private players in respect to the use of machine learning system which included provision of effective remedies to its victims and ensuring more transparency in the entire process.The efficacy of the Declaration is still to be proven as parties involved are in the process of seeking endorsements from A.I. companies. Nevertheless, efforts for translating fundamental human rights for the AI space have already begun.

V. THE WAY FORWARD

Protecting and respecting fundamental human rights could open doors for broader social benefit and common good. The failure of which can easily mount chaos. Limitations of Human Rights cannot be denied and they are not equipped to address all the inconspicuous concerns pertaining to A.I. There may arise scenarios where the negative social impacts of A.I. technology would not be anticipated in terms of human rights. Though human rights have attained legitimacy over the years, the intrinsic political value of human rights is still controversial.[18] And scrutinizing the geopolitical environment of the recent years, it can be seen that the chauvinistic nationalism, promoting self-interest over common good is on the rise which is a defeat of universal rights system.[19]

[16] How to Prevent Discriminatory Outcomes in Machine Learning, World Economic Forum, March 12, 2018, http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf.

[17]

[18] Christopher McCrudden, Understanding Human Dignity (Oxford University Press: 2014)

[19] Opening Statement and Global Update of Human Rights Concerns by UN High Commissioner for Human Rights ZeidRa'ad Al Hussein at 38th Session of the Human Rights Council, United Nations Human Rights Council, June 18, 2018, https://www.ohchr.org/EN/HRBodies/HRC/Pages/NewsDetail.aspx?NewsID=23206&LangID=E

In essence, a human rights approach to A.I. requires to be fully integrated so that it could be practically implemented through policy, practice, and organizational change. To further this goal, the report offered some initial recommendations:

- In areas of high human rights concerns, effective routes of communication which have local civil society groups and researchers so that identification and response to risks related to AI deployments can be done timely.
- Human Rights Impact Assessments should be done throughout the life cycle of every A.I. system to ensure efficiency and it should be consistently re-evaluated to accommodate any recent developments in algorithm impact assessments. Development of Toolkits to assess needs of a specific industry will also go a long way.
- Acknowledgement of human rights obligation and incorporating a duty to protect fundamental rights in national A.I. policies, guidelines and possible regulations should be taken by the governments around the world. The development of A.I. should also channelized through a more active role in the dynamics of multilateral institutions like U.N.
- Business models, workflows and product design should be operationalized from human rights under the guidance of human rights lawyers, social scientists, policy makers, engineers and computer scientists especially in the light of the fact that human rights principles were not written as technical specifications.
- A further examination of the limitations, value and interactions between humanitarian law and ethics in relation to up and coming A.I. technologies, should be done by Academicians. Specific A.I. risks and harms should be tackled by all the stakeholders like human rights and legal scholars in tandem. Empirical investigation on ground-zero impact of A.I. on human rights should be conducted by social science researchers.
- Human Rights Impacts derived from A.I. systems should be constantly researched and publicised by UN human rights investigators and special rapporteurs. Evaluation of existing UN mechanisms to check for international rights monitoring, accountability and redressal mechanisms should be done by UN officials and participating governments to identify whether they are adequate as a response to A.I. and other up and coming technologies. Shared global values based on fundamental rights should be vehemently promoted by UN leadership adorning a central role in international technology debates.