Humanity by "Code"

Rohit Kumar
St. Architect & Manager - IBM

ABSTRACT
Code is an interesting word, especially in relation to other words or subjects. It may be -"Code of Law/Conduct" or "Code of Life " or even "Code of a Computer Program ". This paper echoes with the main subject of discussion and includes all the above three codes and their relations. The topic "Impact ofAl on Human Rights " is an amalgamation of all these three codes. The term Code of Life normally is considered in biological terms as the genetic code (DNA) but could cover in all sense the Code of Dignity of Life. The codes by which the life of a living organism, primarily Hitman Life is protected and regulated. In simple terms known as Codes of Human Rights. The Code of Law and Code in a Machine should be in good synchronization, if the Code of the machine. is capable in any way, to manipulate and control the Human Rights, even the minutest way. The Al-driven machines are imperative to many walks of life and have a deep impact in the way we live now be it medical sciences, education, law and order, travel, marriages, divorces, and myriad areas. In simple terms, Al is slowly becoming a major influencer in the areas mentioned and many others. There had been concerns raised, that these AI algorithms should he free from biases and discrimination and should not he instrumental in any way, snatching or distorting basic rights of people. Famous American Mathematician, Cryptographer and "Father of Information Theory!" Claude Shannon, had once quoted very well in this context that, "I visualize there will so much indulgence with computers and AI in walks of life, that humans might be to robots what dogs are to humans". Conceptual super intelligent machines with full-blown human cognitive capabilities are still a dream. But we are by now already have reached a level where the Al-driven applications are influencing various walks of life and could be already having biases ingrained in their algorithms and may easily take away human rights in ways even not known to people using them. A very famous and recent example in the case ofFacebook, where it was found responsible to aggravate genocides in Myanmar. Through this paper, we will walk through some facts and details to elaborate on the influence and risks of AI on Human Rights. This paper will conclude trying to answer statements like - Is this really a risk? How seri-

ous is this? Is there a way or ways to mitigate this?

AI and Human Rights - What are these and the crossroad?

Artificial Intelligence (AI) is intelligence demonstrated by machines, which helps execute intelligent activities, by using them. Intelligent activities could be simple action-reaction pair or learning based results, e.g. in case of strong AI. AI was started as an academic research subject, at Dartmouth College, Hanover, New Hampshire, USA in 1956. At that point of time, it was predicted that the AI driven machines, as intelligent as humans, could be available vvithin two decades or less. Reality, we are still nowhere close to it. Many subdisciplines and related subjects were invented, like - Machine Learning (ML), Natural Language Processing (NLP), Neural Networks, Deep Learning and recently Cognitive Computing. All this is being done under the pursuit to develop machines, bearing cognitive capabilities alike human or more than that. The first AI machine could demonstrate ability of playing game of checkers with average human capabilities. With time now, it has progressed to execute far more complex activities which might even need multiple discipline of knowledge e.g. Robotic Process Automation (RPA), Digital HR based Recruitment, Resource Management, Irrigation Management, Driverless Cars and many more. Although, all these machines are still not comparable to human cognitive capabilities.

But, it has certainly reached a level, where it performs complex intelligent activities requiring complex considerations and interactions e.g. Recruitment requiring at least — Psychology, Market Research, Resource demand Planning knowledge; Population Analytics requiring at least -Culture, Weather, Demography, Infectious Disease patterns, Crime patterns knowledge.; Bank Fraud Prediction requiring at least - Banking, Crime Pattern knowledge; Predictive maintenance requiring at least - Machine, Breakdown and Maintenance, Inventory of parts, Production Schedule knowledge; Medical diagnostics e.g. Cancer , requiring at least - Disease, Treatment, Case study, Research knowledge; and many more, further to name few - Planning a movie cast by Netffix, Playing Game of Jeopardy, Crime detection. AI has become integrated in various aspects of our daily life. It does impacts in lot of areas, decision making. But a very important fact to bear in mind is - "the machine intelligence is mere reflection of thoughts and concept of world,of its makers and does includes their biases. The biases do sit at the core, where the machine's algorithm resides. Apart from these biases, lack of adequate training may also generate machine generated biases."

Moral and Social sensitivity, even if built in these machines will

also be just rules and logic for them. There is nothing like moral or Ethical machine. What makes it more complicated is-

"Morality, criteria of considering something good or bad, acceptance or Non-acceptance have no global baseline, it may vary very much in each demography.

A person may be able to sense these emotional, social and cultural variations even through unspoken language say a glance, a gesture etc. So, he will certainly acknowledge them, even if he may not decide to accept it partially or completely. This is still impossible for smartest of available AI machines. A machine cannot be sensitive this way and nor it can decide so flexibly. AI is still more wire, algorithms and code. It is still incapable to understand or project delicate and complicated human expressions and feelings. It may detect expressions based on mathematical and statistical algorithm by identifying face point distances, but still not perfectly. A person may ciy seeing his son returning to him after years, but is it happy or sad expression? It may be trained to do so, to an extent by training and adding rules, but that is what also creates the limitations. Machines still do not feel but respond. They provide outputs not by becoming sensitive to these but by able to detect them and converting them to designated inputs. Primary job of machine is to maximize goal benefits. So, there is always a chance of humane factor-less decisions. The ones based on results for best optimized benefits and not emotions etc. Human rights are defined as rights to which every human is entitled on this planet, by virtue of being a human. The core of it is respect, security and dignity for each human, irrespective of any racial or physiological variations. Concept of Human Rights are recognized worldwide. Even if a country does not adopt it fully or partially it is aware of this concept. Most important and most ratified treaty in this regard is The UDHR i.e. Universal Declaration of Human Rights adopted in 1948. The two Covenants adopted further in 1966 were International Covenants 1CCPR i.e. International Covenant on Civil and Political Rights and ICESRC i.e. The International Covenant on Economic, Social and Cultural Rights.

The UDHR stands as common agreed goal and basic principles and rules towards dignified, safe and good life of all human and based on principles of natural law. Within the thirty articles of UDHR - Civil, Political, Economic and Social rights for people are defined. In last 70 years of its adoption, it is ratified by majority of the countries barring a few. This treaty was not created to put any legal obligation over any country, but to just prescribe a uniform recommendation globally. Many countries have adopted components from this treaty into their Constitutions, Including India. Such is the effect and accept-

ance that 198 countries have by now changed voting laws to allows female suffrage. More than 100 countries have abandoned capital punishment laws in their states, while it was just 4 when UDHR was first adopted. These declarations and treaties are the most adopted ones and they are longest standing treaties in this area. Despite not having a legally enforceable status, these could be interpreted as legally accepted principles based on the overwhelming majority of acceptance in General assembly or in general.

Abuse of Human Rights had been conventionally observed between human Vs human only. Now, additionally the parties in conflict could be AI driven "decision taking machines" and human. But, is it really a "Machine Vs Human" conflict? The answer is no. It's a conflict between provisioning, programming and application of such intelligent machines by human and their implications on human. So still Human Vs Human. The only difference here is this is happening indirectly by using machine which could abuse.

In general, people even in most complicated situations, would come up with very simple and straight solutions. It is not avoiding the "paralysis by analysis" like scenario, the primary objective is to resolve the larger issue in hand asap. Time to analyze all possible solution combinations could be huge. While machines with huge computational capabilities and vast databases, could analyze and weigh millions of such combinations in fractions of second. It could generate most optimized results and to achieve best possible results and maximizing benefits where possible. The humane aspects like mercy, morality, ethics, human rights, are not essential components towards calculating such results. That's why a most optimized result with maximum benefit to one could be very unfair to someone.

Let us see one quick example. An organization ABC traditionally managed workforce through Resource Managers. The managers used reporting tools to have real time visibility of resource Vs project and planned future assignments. Manger M decided to match Peter with one project X.-

Peter was an excellent match for project from various skills and experience prerequisite. In discussion with Peter, it comes up that Peter has some extremely bad and hurtful experiences with the location. In shadow of such bad and hurtful experiences tied with location, he might not be able to perform optimally. Now these are not kind of details which are reduced as system records against employees. Maximum detail possible in this regard is storing location preferences but not normally possible to break it up to sublocations. These are purely left upon human interactions. In view of empathy and people management principles, M decides to deploy Peter elsewhere. This also helped gain his loyalty and motivate

him further.

Let's see the same scenario when a machine is assigned the same task. This could execute 100% of such allocations in list in a day, earlier when done by M this was about two weeks task. It's an extremely simple and repetitive task correlating project required skills and employee's skill along with their next availability. Additionally, to further reach a potential result, the machine could run predictive to assign weightage on their chances of success in such a project. This brings Peter's name.

Peter is assigned to the project and informed via system generated electronic mail. This most optimized and potential result missed the most essential humane aspect. Even Peter was deprived of the opportunity to discuss best employment option for him. The company had no reconsideration of appeal to the assignment for him as the next eligible person had huge difference of weightage score of success and the project was of high priority. As a result, Peter faired very bad. Machine took not of this and identified as a bad resource. This impacted Peter's growth opportunity and all similar cases in future were categorized as risky and weighed less. A machine generated bias or magnified bias formed.

More scenarios could be there, where due to bias people could be impacted, e.g. a crime profiling machine identifying some subjects as most probable criminal, due to scar on face and dark skin and weighing more on probability due to these factors. This could be due to firm belief and bias in makers mind or just even due to insufficient training data. So, biases originating from code or training data. There is no guarantee of having a data set wide enough to suffice all possible scenario, there could be infinite mathematical possibilities. The code must conclude one definite result and that is the most weighted response primarily via code and then training and on top of it due to experiences from previous interactions and success or failure. The biases with human decision change with time or experiences. But machines might keep holding them and amplify them until it is removed or updated by human intervention. Recently, Facebook analytics redirected pages to posts related to hate messages, just based on trend analysis and to maximize page hits in Myanmar, That lead to huge genocides. Faeebook did not intentionally make such logics in code, but the biases were considered opportunity by leading to unfortunate situation.

AI is not really bad for mankind. Its development, growth and applications will increase and must increase quality of life and provide more security and comfort too. The growth of AI is helping us, Some examples could be drone based precise nozzle fertilizer sprinkling in China, defect tracking on factory lines, bank fraud detection etc. But the fact until now is, these are made

as machines with a primary purpose to maximize benefits. This is where the conflict of result and expectations comes. If we examine from what we discussed until now, it's the missing focus and regulations in design of such machine reason for such biased machines. No stress was originally towards humane response from these machines. This is where AI and Human Right issues are bom and discussed now. Again, there is no term as ethical AI. These machines are high capability and data thirsty machines. Humanity to be sought from these machines must be achieved through Humanity by code.

Impact on Human Rights

UHDR articles are well-thought-out baseline human rights and ratified by most of the nations. Taking reference of some UDH.R articles, we will try to highlight where Human Rights listed in those articles could be obstructed.

Will it be wise and harmless to deliver criminal justice largely based on predictive and statistical methods? Is it safe to decide if a person is criminal or not just based on mathematical weightage? Undisputed response is "no". It might destroy a life or even result to death sentence. COMPAS [Correctional Offender Management Profiling for Alternative Sanctions] used in United States is one example. The application is largely used by judiciary to decide sentences, parole etc. There are many examples of human right violations due to its use. To count few, it has bias towards gender where female criminals are dealt leniently, also there are cases where dark skin suffers more as compared to white. Clear case of bias and bias leading to human right violations. Maybe the code is biased or based on historical data of policing, which for long time was biased towards one skin color. This violates the very first and second article of UDHR about the freedom, dignity and equal rights and no relation of race, color etc.

Article 6 of UDHR says, that everyone must be looked from same eye of law. But legal predictions to normalize and optimize justice to have a bias free and accurate justice might actually add biases into it and violate their equality in front of law. The same example of COMPAS could be valid, as mentioned in previous paragraph.

Article 12 of UDHR talks about holding right to prevent arbitrary interferences in personal space and privacy. The volume ad variety of data being collected without one's consent and knowledge; via web foot prints, cookies, web trails, transactions etc. is taking away this right, e.g. Google collects location details of every individual Android user or users using Google products on their devices. This might have severe privacy and security concerns too

Article 16 of UDHR talks about to have right of marriage and safely stay in society. But the analyt-

ics on marriage portals could provide weightage points based on what an individual write or watches on web. which educational or demographic background he or she belongs to, there Credit score etc. Visiting an adult magazine sites by a matured adult, which even does not have any immoral content, could be detected and flagged by marriage agency AI programs. It may surreptitiously categorize and mark him into a category of low morality and trust. Which may hamper one's chance of fair match with wider subjects there. One may argument this is for good, but chances of false negatives could be potentially hazardous.

Article 23 of UDHR talks about the basic right to work, free choice of employment. This should be without any discrimination and alt to be paid equal and at par with others. This is a much-exploited area, now thanks to AI based analytics. The algorithms scan the profiles over social media sites like Linkedln and might rate them based on there bias filled algorithms and instead of helping anyone they end up harming. Salary increments are not function of internal factors and KPI of an employee, but additionally, AI based comparison in relation with external salary listing sites e.g. Glassdoor, future prediction of volume of work in that area, availability of such skill in market and that demography etc. and may recommend a lower wage hike. The actual effort put by the employee could take a back seat in such measurements. This may even create a high attrition scenario, not only for that organization but that industry too.

Many such scenarios could be mentioned against each of the UDHR articles and are a risk towards Human Rights for people in general.

AI, is it a necessary evil?

AI is necessary and not an evil. We will now discuss about this aspect and thought. The natural question might come to mind, why not evil when we discussed so many demerits in previous section. Let us discuss further to clarify. It is necessary to discuss all the aspects to conclude. So, before reaching any conclusion, let us discuss the other aspects as well.

in global markets of world, businesses compete, push elbow-to-elbow for survival and growth. The world is now flat-and-even ground for multinational players, even in similar lines of business, and competing locally. This is known as fourth industrial revolution or Industry 4.0,

The diverse customer base has opened-up for businesses. Marginal profit business is reality and automation is key: to predict, plan and succeed. Customer journeys are to be manipulated not followed. Managing client loyalty is very important and AI based analytics are very important. Also, the operating costs should be under tight control and business

processes reinvented continuously.

Thus, repetitive tasks should be automated using intelligent systems, making workforce focus on next level of intense quality delivery, e.g. A law firm of present time understands that, in most of the cases where they are working on various contracts etc., the core of these documents is similar. A vast part is always generic but needs to be just verified for sanity. So, do they really need to read through whole of it to compare every time or focus into key areas. AI based solution could take care of this and the lawyers in firm could spend more time is dealing with law and representing in court rather than reading unnecessary and generic documents. There are many such examples where using AI. productivity could be reclaimed, and workforce could focus on specialized areas of work, instead of doing repetitive jobs. This is one of the important applications of AI, "repetitive job automation". Other very important use case is precise and rapid decision making. Due to huge computational capabilities and historical data sets, the machines are in a better position to rapidly take fact-based decisions.

In this era of reinventing again and again to sustain and succeed it is very important that the utilization of work time is as much more productive and focused towards redefining goals constantly and achieving the same too.

What is considered evil then: AI will take away jobs and may interfere or manipulate human rights. Is this really a fault and evil of AI? Let us discuss this further in following section.

Does AI decide?

AI takes decision is as phantom analogy, as it would be, if we say a speeding truck knocked a person on road, while it was actually driven by the driver sitting behind the wheels. So, in reality the phrase should be, the driver driving a speeding truck knocked a person. If there had been any rights court for trucks, they would have sued humankind and news agencies, to say this about them, "a truck killed a person". Where it is a human driver who made the truck drive over a person and made both the victim and truck suffer.

This was just an imaginary commentary in lighter vain, but not completely away from the point of discussion here. Think.

AI applications are also bare machines and takes decisions as programmed and trained by human. It is more than a normal mechanical machine e.g. a simple truck because it does participate in generating decisions, based on - the rules coded in it, historical data sets fed into it, and learning based on success & failures. This still does not make it a thinking machine, it mimics and synthesizes decision based on previous planned course of action and training. All the capabilities are around the written

code, to give it a way of thinking - The Machine thinking capabilities. The machine thinking capabilities till date are largely single dimensioned or subject area specific or a combination of few. It can mimic thinking, but only the way it is coded or designed to think and only limited to scope of subjects it is made for- at core a bare Algorithmic arrangement. The code and Algorithm are designed by people and does include their view of solution or world, for a to be problem or even approach refinement. Thoughts transformed and framed into machines via Algorithms. So, it is joint view of world by those set of people, which becomes the eye and brain through which the AI machine sees and think. Which also means it would be inclusive of their biases, conscious or subconscious. Facebook translation blunder - one Palestinian man was arrested and grilled for hours by police in Israel as he posted a picture with a bulldozer and Hebrew captions saying Good morning. But the AI translation in Facebook re-translated it to "Attack them" in Arabic, as the text looks similar to that in Arabic. Many Crime detection applications have biases in them e.g. COMPAS mentioned earlier have also proved to be rigged with biases.

Are these AI biases or Algorithm induced biases?

Just, because the machine mimics thinking will it be logical to consider it culprit of committing bias. At times it could be just a choice of training data set resulting to biases, but effect is same -violation of human rights. So just like the truck and driver analogy, the AFs primary and secondary learnt responses are actually as they are made and trained by human makers and driven by them. They are coded to bias, consciously or subconsciously or both. So, bias by code. We can safely say that, AI machines are not core problem, the problem is the way they are made or programmed. Otherwise they are just machines taking dotted course of actions to maximize benefits and which is on Algorithms and training data sets. There is no AI machine, including Watson from IBM, which can think like humans do. They mimic thinking. They mimic and take solution course, the way they are programmed and trained and in respective subject areafs). So, we have Watson for Cancer, Watson tor Marketing etc. But no Watson for all or General AI machine. There is nothing as ethical machines, they could be made to mimic and behave ethically or act, but again by rules and training. But for machine it still is an output fulfilling conditions and framework of its operation, and nothing more. Any morality is as much code and rule to machine as any other result methodologies to provide result output by these machines, But good possibility and ray of hope is that, it is possible to code such machines. If we can mimic thinking, we can also mimic

thinking "including humanity".
Let us discuss in next section the root causes and mitigations of such AI propagate biases.

Conclusion: Mitigating AI propagated biases and prevent Human Right issues?

We could appreciate from the discussion by now that, there is human right violation occurring due to AI and more is expected, as AI based applications are becoming more mainstream, in all walks of life. But we do understand, that the core problem is bias by Algorithm and not AI based bias. If we control the former, later is resolved to a large extent. The actions by AI machines are reflection of the machine thoughts, result of coded algorithms, and the training provided to them.

Presence or use of AI is not core of problem, the problem is how they are coded or trained and absence of any standards and regulations, to include and standardize human rights and ethic related validations in their decision processing.

The resolution should be pointed in this direction. Until now, the codes are mostly commercially oriented. It is there to maximize benefits via smart and quick decision making. It does take in consideration: mandatory regulations, Tax and Legal compliances. But, there is no standard or compliance with regard to inclusion of humanitarian or ethical aspects, as integral decision-making factors. New rules and regulations should be in place, to ensure the makers of such machines, in collaboration with relevant legal experts do human rights impact analysis of processes being automated. Such impact analysis and recommended mitigations validations mandated in code should go through review and certified before code design. There should be appellate authorities, to reach out in case of any violation later in time, and method to assess damages done by then to provide relief and also approve changes required for further usage.

The regulations should be there to enforce designing machine codes inclusive of human right and ethics considerations on top of all the considerations. There should be provisions for audit by sampling decisions and impact to people over time.

These basic frameworks should suffice to put in place basic provisions and controls.

The impact of human on human or state on citizens were the limited application for Human right treaties and referendums earlier, as they were sole decision makers and action takers. This remains true now as well. But in due course of making decisions or taking actions smart AI machines are also used, and the machines know no ethics or morals. So, they may violate it by "not recognizing" it. So, time has come to put a solid framework in place.

As mentioned until now in this section, the key elements should be:

Mandatory human rights impact analysis when such a system is proposed

From impact perspective highlight the probable impact areas

Review and Certify - Design and Training Data Set requirements

Design should be submitted with confirmation of inclusion of scenarios highlighted in impact analysis

Build application should provide test evidences for all such scenarios highlighted earlier and submitted with certifying agency

Appellate authorities to report discrepancies and violations detected later

Provision of damages assessment and relief to affected parties

Provision of re-certifying with amendments in case of issues detected later or post audit checks to remove machine generated biases

Audits with samples of results (audit with framework)

The key is as simple as adding up the missing elements in the code to gain humanity by "code" in these machines. The deficiency and omission of Human rights could be corrected by putting that in code. The well recognized referendums like UDHR could become abase for this framework.

The moral responsibilities of these human right violations will always be there with people and not AI based machines. The bias does not only come by coding but also by machine learning. Based on experience and deducing meaning out of it, machine generated bias may also build and that might need some updates time to time.

Involvement of Government and neutral bodies both are veiy important in this.

Toronto declaration in 2018 May is an important step in this regard. It is called as "Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems". In Sep 2018, The UN Secretary General also released a strategy on new technologies that seeks to align the use of technologies like AI with global values found in the UN Charter, the UDHR, and international law.

So, there is already comers of law makers, researchers and academia seeing the importance of these regulations and work has started in this area.

The sad thing about artificial intelligence is that it lacks artifice (clever or cunning devices) and therefore intelligence." Jean Baudrillard [French sociologist, philosopher, cultural theorist, political commentator, and photographer]

***