IMPLICATIONS OF AI BIAS IN THE OPERATIONS OF ADJUDICATION BY AI

Abstract

Artificial Intelligence (AI), though had started as a tool of minimizing human effort has now grown into an organism which is thought as a replacement option to human intelligence in multiple sectors, where it is on the verge of being entrusted with multiple grave responsibilities, including legislation and adjudication. The emerging trends, for example, an algorithmic tool is currently in use in the American justice system which assesses bail decisions, the AI legislator model being tested for real world applications in China, reveal a global trend, that in future the world will see further uses of AI in administration, adjudication and even legislation. The primary arguments against the use of AI in the fields of adjudication and legislation regarding a right to free trial are as follows – (i) there is a right to a reasoned judgement, (ii) there is a right to be informed of the reason and (iii) a right to rebut the reasoning. In a scenario where a judgement is solely formulated by the use of AI all the above-mentioned rights are violated. The roots of the right to be tried by a natural judge or the right of a human to be adjudicated by another human is enshrined in the ethics of the morality of the global north and it has been further confirmed in the annex I of the European Ethical Charter on the use of Artificial Intelligence in judicial systems, where it is mentioned in the 8th paragraph that the fundamental principles of a fair trial are inclusive of the right to be tried by a natural judge (human judge). This paper aims to elucidate upon the major legal problems of application of AI in adjudication and legislation, and what could be a possible way forward from it. In the following sections a brief explanation of AI bias has been provided in which the main research question of the paper has been further contextualized.

Authors

Sunavo Ray Chowdhury
Student
Department of School of Law
VIT-AP University, VIT-AP School of Law
Amaravati, Andhra Pradesh, India.
Sunavo.20bal7010@vitap.ac.in

Rishav Sinha
Student
Department of School of Law
VIT-AP University, VIT-AP School of Law
Amaravati, Andhra Pradesh, India.
rishav.21bbl7008@vitapstudent.ac.in

I. INTRODUCTION

Artificial Intelligence (AI), though had started as a tool of minimizing human effort has now grown into an organism which is thought as a replacement option to human intelligence in multiple sectors, where it is on the verge of being entrusted with multiple grave responsibilities, including legislation and adjudication. Even though in the free world, there hasn't been yet a case which has completely been adjudicated by the use of AI, but the emerging trends and recent developments, for example, an algorithmic tool is currently in use in the American justice system which assesses bail decisions, the AI legislator model being tested for real world applications in China, reveal a global trend, that in future the world will see further uses of AI in administration, adjudication and even legislation.

However, there are fundamental differences amongst the global south and the global north regarding their approach towards AI. The main problem regarding, implementation of AI revolves around the issue of 'Black Box Paradox' and the question of AI Bias, stemming from it. In the same light it is also contended that the humans have a right to be adjudicated by humans[1] and AI, which hasn't yet been attributed a legal personality, should definitely not be used in the adjudication process of human beings let alone, legislate[2].

This paper aims to elucidate upon the major legal problems of application of AI in adjudication and legislation, and what could be a possible way forward from it. In the following sections a brief explanation of AI bias has been provided in which the main research question of the paper has been further contextualized.

II. HYPOTHESIS AND METHODOLOGY

It is the primary hypothesis, conceived in the paper due to absence of legal reasoning created by humanity in the working algorithms of the AI, it would never be able to impart impartial, truly neutral and justiciable judgements and, due to 'Black Box Paradox'[3], it would further lead to AI Bias. Another area of focus in this paper is associated to the Data Privacy related concerns, as the AI models are trained on data environments created for their training, so often when the data that is to be accessed by the AI remains unrestricted the model trains on unreliable and undesirable data which then further down the line leads to the problem of AI Bias. Furthermore, the absence of human adjudicator, opaque process of decision making and limited scopes of intellectual review of the reasoning of the AI behind particular decisions violate the basic principles of fair trial.

The main research question of the paper being a qualitative question, the qualitative methodology of research has been employed in the paper and the paper has been formulated in a purely doctrinal methodology. The true novelty of the paper lies in the fact that this paper adds to the building literature on use of AI in the fields of adjudication and legislation.

---

[1] GORSKI, "Why a human judge? (2023)", EUCRIM, Issue 1/2023, pp. 83 - 88
[2] CHESTERMAN, "Artificial Intelligence and the limits of Legal Personality (2020)", International and Comparative Law Quarterly, Cambridge University Press, 69(4), pp. 819-844

III. AI BIAS AND BLACK BOX PARADOX

Black Box Paradox in simple terms means the opaqueness of the machine learning process through which any AI model comes up to a decision and also the fact that the process in question cannot be known. The process in true sense can never be known which is why, the way an AI model decides upon an outcome can never be truly deciphered the way a decision made by a human can be. This problem has been a perpetual concern for the developers which is why it has given title of paradox and this paradox happens to be the biggest drawback of the AI technologies. Furthermore, AI Bias[4], which is also known as the algorithmic bias or machine learning bias is the undesired existence of a human bias in an AI system. The most common argument raised against this says that a careful dealing of the algorithm will in all likelihood preserve an AI system from any biases.

Due to Black Box Paradox, it can never be truly predicted, exactly what will an AI model learn from a data environment hence, even if a human bias makes its way into the AI thought process it will not be known until it has become too obvious. In the curious case of AI adjudicators, it would mean that by the time a bias in an AI is discovered multiple cases would already be decided upon a biased reasoning which violates the very fundamental foundations of the conception of fair trial.

IV. AI AND FAIR TRIAL

The primary arguments against the use of AI in the fields of adjudication and legislation lies in the fact that humans have a right to be adjudicated by humans, which is an inherent extension of the fundamental principles of fair trial. Substantial amount of authorship has been developed on this issue which suggests that in a free and fair trial – (i) there is a right to a reasoned judgement, (ii) there is a right to be informed of the reason and (iii) a right to rebut the reasoning[5]. In a scenario where a judgement is solely formulated by the use of AI all the above- mentioned rights are violated. The roots of the right to be tried by a natural judge or the right of a human to be adjudicated by another human is enshrined in the ethics of the morality of the global north and it has been further confirmed in the annex I of the European Ethical Charter on the use of Artificial Intelligence in judicial systems, where it is mentioned in the 8th paragraph that the fundamental principles of a fair trial are inclusive of the right to be tried by a natural judge (human judge)[6]. The arguments in favour of the use of AI adjudicators rely heavily on the cost-effectiveness of the procedure involving an AI, the speeding up of the whole pace of disposals regarding cases and a significant reduction for the scope of corruption. Bt clearly the cons, outweigh thepros as it still cannot be guaranteed that the AI is able to comprehend the human code of ethics which again varies depending upon multiple factors. Essentially an AI is an embodiment of Neural Networks

[3] RAWASHDEH, "AI's mysterious Black Box problem explained (2023)", University of Michigan Dearborn, < https://rb.gy/r82u9> accessed on 31 Aug 2023

[4] MANYIKA et al, "What Do We Do About the Biases in AI? (2019)", Harvard Business Review, < https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai> accessed on 31 Aug 2023

The arguments in favour of the use of AI adjudicators rely heavily on the cost-effectiveness of the procedure involving an AI, the speeding up of the whole pace of disposals regarding cases and a significant reduction for the scope of corruption. But clearly the cons, outweigh the pros as it still cannot be guaranteed that the AI is able to comprehend the human code of ethics which again varies depending upon multiple factors. Essentially an AI is an embodiment of Neural Networks created by computers which, most likely, if not often, will, run into computational errors.

There has been a significant rise in the usage of autonomous models for various assessment purposes in the American criminal jurisprudence, which is also the focus point of interaction of AI and common law and 'due process of law' there in (As, till date no significant interaction between 'procedure established by law' and AI has been noted). In the case of *Brady v. Maryland*[7], a US criminal court decided that AI tools can be used only in cases where the non- disclosure of the algorithmic scheme of the AI tool used, does not prejudice the rights of the parties. Furthermore, it was stated in one of the other cases that in 'due process of law' the non- disclosure of the algorithmic scheme of the AI in question is violative of the rights of a person to confront the adversarial witnesses (witnesses that have spoken against) which is protected by the Sixth Amendment of US Constitution[8].

Apart from the issues of the constitutional rights the other major issue that lie in this case is regarding data and data structures, the models on which these AI tools work are trained on data as discussed in the erstwhile section. Regarding theses data structures the most common apprehensions presented by the American and the European think tanks are – (i) there is scarcity of vital and reliable data, (ii) whether or not data privacy measures are being well taken care of, (iii) ethical limit to access of data available to an AI. For example, there are apprehensions for a racial bias being created in AIs based on facial data and other bio-metric data related to the criminal justice system[9]. With the increase of the AI systems in US alone there can be seen a rise of false positives with a ratio of 1 to 10 against people of color and a similar bias can also be observed in cases of women as well[10]. Simply put, the human logic associated with humanity that the commission of a crime by a person of a particular race doesn't increase the probabilities of a person of same race committing the same crime, cannot be hardwired into the AI logic system. The closest analogy to the thought process of the AIs is, the AIs try to make a guess for an outcome when they are given a problem with a non-unitary solution, in this scenario the AI will engage in a method of 'odds against and odds in favour' mathematics, in terms of guessing the AI will make the best possible guess, which many times might even correlate with the reality but it will always be a qualified guess and that is why even if there is significant accuracy in the decisions made by AI, their judgement will never be a judgement in the lexographical sense of the word[11].

[5] ULENAERS. "The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge? (2020)", Asian Journal of Law and Economics, Issue 11, No. 2
[6] DMITRUK, "Towards Explainable Artificial Intelligence (2019)", UH Bio-computation Group, University of Hertfordshire Higher Education Corporation (2016), < http://biocomputation.herts.ac.uk/2019/11/06/towards-explainable-artificial-intelligence.html> accessed on 31 Aug 2023
[7] *Brady v. Maryland,* 373 U.S. 83 (1963), Supra note 60 at 30

Comparative study between the approaches of the Global North and Global South towards Artificial Intelligence: Artificial Intelligence (AI) has almost become an integral part of human lives now and for some time and there have been significant attempts made to regulate it and legislate upon it. There is a difference of approach between the global south and the global north towards AI. The southern approach is one which is based more in the economics where attempts have been made to realize the importance of AI through the landscape of Intellectual Property and then regulate upon it. Whereas, the northern approach has been purely based in a bottom-up fundamental rights approach. The main reasons behind humanity's need for regulating AI is directly related to the need for establishing a liability regime for AI, especially now, when Human-AI interactions have increased to such a degree that one not only does impact the other but also has the capability to shape the life of the other for better or worse. From being used in the in the financial sectors to that of government surveillance and warfare, AI has assumed roles which now really raises the questions regarding the morality of AI and whether or not any morality of sorts could be attributed to it.

Considering India, as a representative case study of the collective approach of the global south towards AI, it can be stated that the root of legislation regarding AI stems from its legal approach to Computer Related Inventions (CRI)[12]. Just the way AI inventions have been thought to be out of the scope of patentability, similarly CRIs were also considered the same once upon a time, during 1999 when the second amendment bill was passed for the Indian Patents Act. But in the case of *Telefonaktiebolaget Lm Ericsson v. Intex*[13], it was held that if CRIs conclusively possesses a technical effect then such inventions can be within the scope of patentability. A similar ratio was followed in the *HTC v. Apple*[14] case and *Alice Corp. v. CLS Bank International*[15], that a patent cannot be given out for an abstract thought but so as long the invention induces technical functioning then such invention falls well within the scope of patentability. Hence, it can be understood that in southern approach regarding AI, this question of 'technical functioning' is going to be the determining factor regarding the patentability of AI inventions. The whole premise of 'technical functioning'[16] is based upon the resolution of a technical issue, which in turn could be interpreted as an outcome, the value of which can be quantified in terms of economics. This approach is recurring common law approach to IP which further sheds light upon the need for regulation of AI from the southern perspective, in which the premium is imposed upon monetary resources. The southern approach to AI which associates its value in terms of quantified economics is also necessary for its regulation from a standpoint where effective steps can be taken to stop piracy done by abuse of AI, recent cases of which has come up.

[8] *People v. Wakefield,* 175 A.D. 3d. 158 (New York Appellate Division 2019)
[9] KEARNS and ROTH, "The Ethical Algorithm: The Science of Socially Aware Algorithm (2019)"
[10] SINGER and METZ, "Many Facial-Recognition Systems Are Biased, Says U.S. Study (2019)", New York Times, 19 Dec. 2019
[11] COGLIANESE, "AI in Adjudication and Administration (2021)", Penn Carey Law School, Penn Carey Scholarship Repository

The European Union has decided to deal with this moral question by creating a model work for a responsible AI, in which it has further talked about integration of Ethics and AI to minimize problems of the nature of AI bias and etc. A critical analysis of the praxis reveals that the EU has attempted to constitute the conception of a Trustworthy AI from an amalgamation of Operationalism on top of Fundamental Rights approach. In its attempt the EU has tried to draft the European Strategy for Artificial Intelligence, through the creation of a High-Level Expert Group on Artificial Intelligence (AI HLEG), in 2018. In their quest for responsible AI or as the Council of Europe has named it, Trustworthy AI[17], one of the main things that this initiative is aimed at deals with the role that is played by law and morality or ethics, regarding the global governance related to AI landscape. The notion of Trustworthy AI is based on three central elements which are – (i) lawfulness, i.e., compliance with the applicable laws, (ii) ethicalness, i.e., AI abiding by the principles and values of human ethics practiced in the human society,
(iii) robustness, which just revolves around the concept AI being technically robust from a social perspective.
The ethical question addressed in (ii) has further been broken down into seven key points, which state that AI should be mindful of the following – (i) human agency (fundamental rights approach), (ii) technical robustness, (iii) compliance to privacy regimes and global data governance landscape, (iv) traceability, explainability, and communication collectively named as transparency (which is also a direct measure against the Black Box Paradox related to AI) (v) fairness (which includes non-discrimination, unfair bias, and universal design), (vi) sustainability and environmentalist qualities and a compliance to associated laws, (vii) accountability (which extends to auditability and etc.). Apart from the above-mentioned key points there also exists a special section on examples of beneficial opportunities related to the operation of AI and AI applications that could pose a threat to humankind which includes surveillance, operation of lethal weapons and other covert AI systems that are within the veil of confidentiality.

Within the key points itself, a limitation of 'time-boundness' can be observed, e.g., the question AI's compliance with the existing global data governance landscape and data related laws is time bound, because even if the AI's compliance with the existing laws can be ensured time and again the compliance mechanism has to be checked and reshaped to ensure compliance with the laws of those times or automation has to be used in this regard.

---

[12] Guidelines for Examination of Computer Related Inventions, Government of India
[13] 2023:DHC:2243-DB
[14] [2021] EWHC 1789 (PAT)
[15] 134 S. Ct. 2347 (2014)

[16] JEDRUSIK, "Patent protection for software-implemented inventions (2017)", The WIPO Megazine
[17] SMUHA, "The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence (2019)", Computer Law Review International, 20(4), pp. 97-106

The usage of automation would again lead to the Black Box Paradox and a periodic up gradation of the AI would directly conflict with the concept of robustness as mentioned in the praxis for the Trustworthy AI. Also, in the draft submitted by AI HLEG, there is a notion of trade-offs regarding 'accountability' of AI, which says if a dilemma ever rises regarding the simultaneous and perhaps mutually exclusive application of multiple ethical guidelines a trade-off has to take place. The main problem with this notion of trade-offs lies within the fact that the certain fundamental right and correlated principles are absolute and they cannot be traded off (e.g., the right to dignified life, principle of equality, the inherent right of humans to be adjudicated by humans and etc.). This further exposes that AI, which is essentially a neural link, running the limited computational power of some drives, might have to pick and choose from a bundle of fundamental rights, through a thought process which is not absolutely reliable and will always remain within the veil of the Black Box Paradox. This very notion of trade-offs is ethically unsound and unacceptable.

In conclusion from a comparative study of both the approaches to AI it can be stated that the southern perspective, focused on the economic value of AI also reveals the fact that the global south envisions AI as tool for economic activities, whereas the construct of the trustworthy AI or the conception of ethics and AI envisions, perhaps a legislative capacity for AI.

V. CONCLUSION

In conclusion it can be stated that the usage of AI should be immediately stopped in the fields of Adjudication, Administration and it definitely should not make up to a point where AI is used in the field of legislation. The usage of AI in the field of legislation will be violative of the social contract theory just as the usage of AI in Adjudication is violative of the principles of free and fair trials. There are concerns regarding the Black Box Paradox which are violative of the principles of natural justice, Furthermore the concerns related to Data Privacy regimes and AI Bias elucidate why the usage of AI in adjudication and legislation is harmful towards humanity. Hence, it can be conclusively stated that the hypothesis conceived in the paper, which says that AI will never be able to impart impartial justice true.

REFERENCE

[1] GORSKI, "Why a human judge? (2023)", EUCRIM, Issue 1/2023, pp. 83 - 88
[2] CHESTERMAN, "Artificial Intelligence and the limits of Legal Personality (2020)", International and Comparative Law Quarterly, Cambridge University Press, 69(4), pp. 819-844
[3] RAWASHDEH, "AI's mysterious Black Box problem explained (2023)", University of Michigan Dearborn, < https://rb.gy/r82u9> accessed on 31 Aug 2023
[4] MANYIKA et al, "What Do We Do About the Biases in AI? (2019)", Harvard Business Review,
[5] < https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai> accessed on 31 Aug 2023
[6] ULENAERS. "The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge? (2020)", Asian Journal of Law and Economics, Issue 11, No. 2
[7] DMITRUK, "Towards Explainable Artificial Intelligence (2019)", UH Bio-computation Group, University of Hertfordshire Higher Education Corporation (2016), < http://biocomputation.herts.ac.uk/2019/11/06/towards-explainable-artificial-
[8] intelligence.html> accessed on 31 Aug 2023
[9] Brady v. Maryland, 373 U.S. 83 (1963), Supra note 60 at 30
[10] People v. Wakefield, 175 A.D. 3d. 158 (New York Appellate Division 2019)
[11] KEARNS and ROTH, "The Ethical Algorithm: The Science of Socially Aware Algorithm (2019)"
[12] SINGER and METZ, "Many Facial-Recognition Systems Are Biased, Says U.S. Study (2019)", New

York Times, 19 Dec. 2019.
[13] COGLIANESE, “AI in Adjudication and Administration (2021)”, Penn Carey Law School, Penn Carey Scholarship Repository
[14] Guidelines for Examination of Computer Related Inventions, Government of India 2023:DHC:2243-DB
[15] [2021] EWHC 1789 (PAT)
[16] 134 S. Ct. 2347 (2014)
[17] JEDRUSIK, “Patent protection for software-implemented inventions (2017)”, The WIPO Megazine

