

AI and Human Rights: A Critical Ethico-Legal Overview

Vikrant Sopan Yadav *

Abstract: The capabilities of artificial intelligence (AI) in ensuring human rights are tremendous. However, it may also have some denting effect on human rights. The use of AI-based surveillance, face detection, etc. has proved racially discriminatory, resulting in grave human rights violations. AI experts have also admitted to the possibility of developers' bias resulting in biased AI inventions. This research article is an attempt to analyze the possible adverse impact of AI technology on the protection of human rights. The author has done an analytical overview of practical instances of AI-related human rights violations in the recent past. An empirical analysis comprising of an observation tool was employed to observe and analyze the expert opinion expressed during a conference. Based on doctrinal and empirical analysis, the author has made some recommendations, such as including technology-related human rights in national and international human rights statutes, to strike a balance between human rights and AI innovation, with the ultimate goal of protecting human rights.

Keywords: artificial intelligence (AI), human rights (HR), issues, statute

HYPOTHESIS

AI technology is a double-edged sword, as along with protecting human rights, it can also cause a severe breach of human rights.

RESEARCH METHODOLOGY

This work is a qualitative investigation of artificial intelligence (AI) and its possible potential for a breach of human rights (HR). The research methodology used in this research is both doctrinal and empirical. It also uses a doctrinal methodology to understand the concept of AI and the possible cases of breach of HR resulting from the use of AI technology. The empirical methodology is limited to the application of observation tools, wherein the author has attended an educational event containing expert deliberations on 'AI and HR.'

* Vikrant Sopan Yadav (✉)
NICMAR University, Hyderabad, India
e-mail: vyadav@nicmar.ac.in

Observations about the opinion expressed by experts are covered in current research work in the form of two tables.

ARTIFICIAL INTELLIGENCE (AI)

Marvin Minsky, the well-known AI scholar, defines AI as “the science of making machines do things that would require intelligence if done by men,” (as quoted UNESCO 2014, 17). Stanford University (2016, 4) report defines AI as “a science and a set of computational technologies that are inspired by-but typically operate quite differently from-the ways people use their nervous systems and bodies to sense, learn, reason, and take action.” In the words of UN High Commissioner for Human Rights (2018-2022), Michelle Bachelet (quoted in Dziedzic 2021), “Artificial intelligence can be a force for good, helping societies overcome some of the great challenges of our times. However, AI technologies can have negative, even catastrophic, effects if they are used without sufficient regard to how they affect people’s human rights.”

AI AND HUMAN RIGHTS VIOLATIONS

Various instances of using AI technology resulting in breaches of human rights have been experienced in the recent past. Some of the prominent human rights (HR) issues associated with the use of AI technology are as follow:

AI, Bias, and Discrimination

Article 2 of the Universal Declaration of Human Rights (UDHR) and Article 2 of the International Covenant on Civil and Political Rights (ICCPR) provide freedom without discrimination. However, many instances of AI committing racial discrimination have been reported resulting in violation of HR granted under the abovementioned provisions. The potential of AI to reproduce, strengthen, or exaggerate unfavourable preconceptions is a serious concern. Such biases can multiply depending on the method used to collect the data.

Bias in AI can be intentional or unintentional. However, intentional bias may be a rare case scenario, whereas the possibility of unintentional bias can be found frequently. Such an unintentional bias may result from input level bias due to the use of biased/partial historical data, incomplete, incorrect, or unverified data, data suffering from an inadequate representation of population, etc.

In 2015, *Google Photos*, which is considered an advanced recognition software, categorized a photo of two Black people as a picture of

gorillas. When keywords such as ‘Black girls’ were inputted into the Google search bar, the algorithm showed sexually explicit material in response. Researchers have also found that an algorithm that identifies which patients need additional medical care undervalued the medical needs of Black patients. (Baweja and Singh 2020)

The United States criminal justice system uses recidivism risk-scoring algorithms in its decision-making about detention at various levels, from setting bail to imposing crime sentences (Angwin 2016). “The software has led to more black defendants falsely labelled as high risk and given higher bail conditions, kept in pre-trial detention, and sentenced to longer prison terms. Additionally, because risk-scoring systems are not prescribed by law and use inputs that may be arbitrary, detention decisions informed by these systems may be unlawful or arbitrary” (Andersen 2018, 19). U.S. Attorney General Eric Holder gave a warning in 2014 that the risk scores might induce bias into the legal system. He demanded that the U.S. Sentencing Commission should look into their application. “Although these measures were crafted with the best of intentions, I am concerned that they inadvertently undermine our efforts to ensure individualized and equal justice,” he said, adding, “they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and our society” (Angwin 2016).

Although facial recognition technology is yet to achieve perfection, many states have already started its use for surveillance. One such worrying example of governmental manipulation of AI-based facial recognition can be found in China.

The (Chinese) authorities are also using a vast, secret system of advanced facial recognition technology to track and control the Uighurs, a largely Muslim minority. Experts said it is the first known example of a government intentionally using artificial intelligence for racial profiling. The facial recognition technology, which is integrated into China’s rapidly expanding networks of surveillance cameras, looks exclusively for Uighurs based on their appearance and keeps records of their comings and goings for search and review. (Mozur 2019)

Deep Fakes

Deep Fakes are another challenge posed by AI technology. It helps the generation of fake (but sounding realistic) audio-visual recordings of reputed personalities and political leaders. Such deep fakes have a greater potential to spread hatred, thereby threatening public

tranquillity. Such tools can be used for spreading targeted communal propaganda through social media platforms. “In 2017, a software developer nicknamed *deepfakes* on Reddit online platform posted his creations that he swapped the Hollywood celebrities’ faces onto the faces of porn artists. After the creations spread rapidly, deepfakes became a new trend.” (Çolak 2021)

Similar is the challenge of shallowfakes: “videos that are either presented out of context or are doctored with simple editing tools. They are crude but undoubtedly impactful. A shallowfake video that slowed down Nancy Pelosi’s speech and made the speaker of the U.S. House of Representatives sound slurred reached millions of people on social media.” (Sample 2020)

According to WIPO (2020, 6), “The more profound issues of personal identity, the right to privacy, the right of publicity, and the ability to control the use of one’s image for any purpose appear more appropriately to be human rights issues, rather than purely or even primarily copyright issues.”

Article 17 of GDPR provides for the “right to erasure,” i.e., the right to be forgotten is a perfect human right weapon that may provide remorse in such cases. However, international human rights statutes such as UDHR and ICCPR are yet to recognize such rights.

AI and Unemployment

The right to work/ free choice of employment has been recognized as a human right under all international HR statutes such as Article 23 of UDHR, Article 6 of The International Covenant on Economic, Social and Cultural Rights (ICESCR), and Article 1(2) of the ILO. AI is estimated to have a tremendous effect on employment opportunities. Technology has been criticized for the loss of jobs ever since the evolution of computers. Experts, however, are skeptical about the impact of AI technology on employment across all sectors.

In 2017, *Changing Precision Technology*, a Chinese factory producing mobile phones, replaced 90% of its human workforce with machines, which led to a 250% increase in its productivity and a substantial 8% drop in defects. Similarly, Adidas has moved towards ‘robot-only’ factories to improve efficiency. (Baweja and Singh 2020)

AI and Restrictions on Freedom of Movement

Workplaces worldwide have been using technological devices to track the movement of employees during working hours. In the recent COVID pandemic, facial recognition was used to track COVID

patients and identify the virus-affected individuals. Such instances of mandated use of facial recognition are likely to raise serious privacy concerns and may also be considered restrictive of freedom of movement.

Commenting on the risk involved in the use of face recognition for biometrics, a scholar commented, “An increasingly go-to solution for States, international organizations and technology companies are biometric technologies... These technologies, which include facial recognition, are increasingly used to identify people in real-time and from a distance, potentially allowing unlimited tracking of individuals.” (Dziedzic 2021)

Amazon has recently been criticized for actively offering ‘Rekognition’, a facial recognition software, to law enforcement agencies for use with police body cameras, allowing officers to identify people in real-time. Police departments in Orlando, Florida, and Washington County, Oregon, tested the product. (Wong 2018)

AI and Privacy

In recent times, privacy has received recognition as a basic human right by almost all democratic nations across the world. In a landmark decision in *Puttuswamy v. Union of India* on August 24, 2017, the Supreme Court of India proclaimed the right to privacy to be a basic right protected under Article 21 of the Constitution of India. (Justice K. S. Puttaswamy (Retd.) and Anr. v. Union of India and Ors. 2017)

However, in today’s technology-driven era, instances of interference with the privacy of individuals are reported almost every day. AI-enabled data gathering poses several privacy concerns, such as easy grant of informed consent, restricting data collection, the ability to opt out, and even removing data on demand.

Cambridge Analytica, the data firm owned by Robert Mercer, where one of the board members was the former Trump aide Stephen K. Bannon, allegedly used Facebook data obtained improperly to build voter profiles. “...contractors and employees of Cambridge Analytica, eager to sell psychological profiles of American voters to political campaigns, acquired the private Facebook data of tens of millions of users — the largest known leak in Facebook history.” (Confessore 2018)

“Given the rapid and continuous growth of AI, filling the immense accountability gap in how data is collected, stored, shared, and used is one of the most urgent human rights questions we face” (Dziedzic 2021).

EMPIRICAL ANALYSIS

With the object of collecting expert inputs on the problem statement, the author observed the opinion expressed by scholars during the International Symposium on Law and Peace organized in one of the premier educational institutes in India. One of the tracks for deliberation was on “AI and Human Rights.” The findings of the observation tool are summarized below.

Findings of Observations

Experts voiced their concerns about the potential of AI to violate human rights in the following words:

Speaker 1 [Vice Chancellor]

- If HR and human values are neglected in the process of development of AI, instead of helping, it may create problems;
- Considering the current invasion of smartphones in human life, humans are likely to be slaves of AI or similar scientific inventions, which may result in violation of HR;
- Invention of technology is for the benefit and welfare of society. A lack of awareness about using technology for ethical purposes results in violating HR.

Speaker 2 [Research Scholar]

- AI may be used as a sham tool for the commission of cybercrimes, e.g., online harassment of vulnerable sections, including HR defenders, through hacking;
- AI inventions may create a monopoly of the developed over the underdeveloped.

Speaker 3 [Judge]

- Machine thinking and acting like a human may suffer from human-induced bias during the development process;
- In some domains, AI may not be convenient, and human involvement/ intervention may continue to require it.

Speaker 4 [Research Scholar]

- International HR statutes do not provide for the case of possible HR violations resulting from AI technology. At present, domestic judicial interpretations are the only source of law covering such cases;
- There is no binding HR obligation in Corporate Sector. Only non-binding sources such as UN Guiding Principles on Business and HR are available. It often results in a breach of privacy;
- There is a complete lack of third-party audits of AI and HR in the business world.

From the above observations about the opinion expressed by experts, one can conclude that the use of AI can potentially violate HR. It may raise some serious issues affecting HR, such as privacy of data, the possibility of biasness, etc. In order to avoid potential human rights violations, the legal and regulatory systems should ensure ultimate human control over the use of AI technology to ensure fault liability. International and national HR statutes do not provide answers to human rights violations by or through AI technology, resulting in a lack of adequate legal control.

Experts' Suggestions

To ensure a balance between the use of AI and the protection of HR, these expert speakers proposed several suggestions, which are summarized as follows:

Speaker 1 [Vice Chancellor]

- Human rights values and their importance should be advocated amongst AI scientists;
- Law and regulations should be developed to regulate the dealing of information/data;
- A specialized regulatory mechanism should be set up to oversee the prevention of human rights violations through technological advancements.

Speaker 2 [Research Scholar]

- AI invention should not be approved/permitted unless adequate safeguards for HR protection are ensured;
- To avoid robots causing harm to human beings, there is a need for '3 laws of robotics' given by Isaac Asimov;
- Humans should always judge discrimination due to its complexities. AI should not take that away from us.

Speaker 3 [Judge]

- New sets of technology-related HR should be recognized, such as, right of choice and informed consent to determine the use of virtual data;
- AI should be used as a tool for the protection of HR; for e.g., forensic medicine data is used to predict victims and offenders in crime which can be used in crime detection, prevention;
- Use of AI may result in loss of job, but in cases involving larger public interest, it should be used e.g., in judiciary and medical wherein there is a lack of skilled workforce.

Speaker 4 [Research Scholar]

· There is a need for a statutory mandate for corporate policies to prevent HR violations in the business world.

SUGGESTIONS/OUTCOMES OF RESEARCH

Based on the result of doctrinal and empirical analysis, and in addition to the suggestions mentioned by experts in the existing literature, the author proposes a few suggestions to ensure the prevention of human right violation by using AI and for recognition of technology related rights.

International human rights instruments such as UDHR, ICCPR, and ICESCR should be amended and expanded to include futuristic human rights applicable in the digital age, which are as follows:

- a. Right to algorithmic transparency: An algorithm is a set of steps a computer program follows to make a decision. Human rights instruments such as UDHR, ICCPR, etc. should recognize the right to open and accessible information/disclosure about how algorithmic tools support the decision and deliver information. The right should encompass informed decision making enabling the humans to check whether the AI model is thoroughly tested or not and the reasoning behind a particular decision making by such model.
- b. Right to opt out: It is a right of human beings to exclude themselves or withdraw consent for processing their personal data. It confers greater control over human beings' personal data and privacy.
- c. Right to exemplary remedy: In cases of HR violations resulting from use of AI, exemplary compensation and withdrawal of AI tool/model should be recognized as a remedial human right. This will achieve a twofold purpose, on the one hand remedying the wrong and, on the other hand, ensuring accountability and prevention of repetition of HR violation.
- d. Right to access: AI-based innovations create monopolistic intellectual property rights. Individuals whose data has been used for developing the algorithm must be provided with a right to free access to the AI model developed based on such data.
- e. Ethical Guidelines: A model code of conduct/ethical guidelines must be prepared to ensure the developer's moral and legal obligation to prevent AI-aided human rights violations.

Intellectual property regulators must have a periodic review of the actual use of AI technology for the betterment of human life and to ensure the same is not resulting in a breach of HR. All intellectual property rights must be subject to observation of HR.

Legal systems should hold humans (such as inventors/developers, producers, people yielding monetary benefits from using AI technology, etc.) liable for compensating the victims. Such liability should extend even in the case of self-reliant AI technology causing harm to HR.

CONCLUSION

AI technology has already become a part of the day-to-day lifestyle of a vast section of today's society. Due to its potential for serving humanity in living a more leisurely and safe life, an AI-centric future is an unavoidable phenomenon. However, as observed in this research work, along with its potential for a better humanitarian society, its possible danger to human rights cannot be neglected. The legal and regulatory systems must carry out necessary reforms to provide for human liability for acts of even a self-reliant/autonomous AI technology to ensure responsibility for every wrong resulting from using such AI technology. Determination of liability on individuals benefiting from such invention of technology (like developers, owners/operators, etc.) is required to ensure accountability for compensating the victims of HR violation. Recognition of new age of AI-centric rights, as proposed by the author hereinabove, is also the need of the hour to ensure a futuristic HR system.

REFERENCES:

- Andersen, L. 2018. "Human Rights in the Age of Artificial Intelligence." <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf> [accessed: 09.02.2022].
- Angwin, J. L. 2016 (May 23). "Machine Bias." *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [accessed: 23.02.2022].
- Baweja, S., and S. Singh. 2020 (July 16). "Beginning of Artificial Intelligence, End of Human Rights." <https://blogs.lse.ac.uk/humanrights/2020/07/16/beginning-of-artificial-intelligence-end-of-human-rights/> [accessed: 03.03.2022].
- Çolak, B. 2021 (January 19). "Legal Issues of Deepfakes." Institute for Internet and Just Society. <https://www.internetjustsociety.org/legal-issues-of-deepfakes> [accessed: 22.02.2022].
- Confessore, N. 2018 (April 4). "Cambridge Analytica and Facebook: The Scandal and the Fallout So Far." *The New York Times*.

- <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html> [accessed: 27.02.2022].
- Dziedzic, M. 2021 (September 15). “Urgent action needed over artificial intelligence risks to human rights.” United Nations. <https://news.un.org/en/story/2021/09/1099972> [accessed: 03.03.2022].
- Justice K. S. Puttaswamy (Retd.) and Anr. v. Union of India and Ors., 4161. Supreme Court, August 24, 2017.
- Mozur, P. 2019 (April 14). “One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority.” *The New York Times*. <https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html> [accessed: 25.01.2022].
- Sample, I. 2020 (January 13). “What are deepfakes – and how can you spot them?” *The Guardian*. <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them> [accessed: 03.02.2022].
- Stanford University. 2016. “Artificial Intelligence and Life in 2030.” https://ai10020201023.sites.stanford.edu/sites/g/files/sbiybj18871/files/media/file/ai100report10032016fnl_singles.pdf [accessed: 29.01.2022].
- UNESCO. 2014 (September 17). “Report of Comest on Robotics Ethics.” <https://unesco.blob.core.windows.net/pdf/UploadCKEditor/REPORT%20OF%20COMEST%20ON%20ROBOTICS%20ETHICS%2014.09.17.pdf> [accessed: 23.02.2022].
- Wong, J. 2018. (May 23). “Recipe for authoritarianism!: Amazon under fire for selling face-recognition software to police.” *The Guardian*. <https://www.theguardian.com/technology/2018/may/22/amazon-erkennung-facial-recognition-police> [accessed: 10.02.2022].
- World Intellectual Property Organisation (WIPO). 2020. “WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence.” Washington: Association of American Publishers. https://www.wipo.int/export/sites/www/about-ip/en/artificial_intelligence/call_for_comments/pdf/org_aap.pdf [accessed: 13.01.2022].