ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS: AN OVERVIEW

Milind Khande*

THE GROWING INFLUENCE OF AI

Artificial Intelligence (AI) is reshaping societies across the globe, permeating every facet of human life. The classic definition of AI dates back to 1955 when John McCarthy and his fellow researchers characterized artificial intelligence as "making a machine behave in ways that would be called intelligent if a human were so behaving".[1] AI encompasses numerous subdisciplines including natural language processing, machine inference, statistical machine learning, and robotics.[2] [3] [4] AI technologies are becoming integral to modern convenience and efficiency, from autonomous vehicles to virtual assistants like Siri and Alexa. Within AI, machine learning (ML) serves as a key subfield in which systems learn from the present data rather than being explicitly programmed involves less human involvement and complex AI decision-making. Some observers believe this will eventually lead to artificial general intelligence or superintelligence that either achieves or surpasses human intelligence.[5]

ML algorithms, allow computers to adapt to new data and improve performance over time (decision tree algorithm, support vector machine algorithm). In health care, AI-enabled diagnostic tools and predictive analytics are revolutionizing patient care, while algorithms in finance result in optimum investment and fraud detection. There is much government interest in the effectiveness of AI in bringing about smart governance disaster management and infrastructure planning, thereby proving its worth for tackling large-scale problems. Yet we

*BA LLB, SECOND YEAR, NATIONAL UNIVERSITY OF STUDY AND RESEARCH IN LAW, RANCHI.

[1] John McCarthy and others, 'Proposal for the Dartmouth Summer Research Project on Artificial Intelligence' (2006) 27(4) AI Magazine 12 https://aaai.org/ojs/index.php/aimagazine/issue/view/165

[2] Rodney Brooks, 'The Origins of Artificial Intelligence' (FoR & AI, 27 April 2018) https://rodneybrooks.com/forai-the-origins-of-artificial-intelligence/

[3] John Mallery, 'Intelligent Computation in National Defense Applications: Artificial Intelligence or Magic?' (2018 Roundtable on Military Cyber Stability, Washington, DC, 17 July 2018).

[4] Rodney Brooks, 'The Seven Deadly Sins of AI Predictions' (MIT Technology Review, 6 October 2017).

[5] Max Tegmark, Life 3.0: Being Human in the Age of Artificial Intelligence (Knopf 2017).

cannot ignore the inherent risks that AI brings with it and careful assessment is required to prevent imbalances that can be caused by growing AI.

SCOPE OF AI

Indeed the possibility that more powerful AI could lead to discoveries in science, as well as enable game-changing progress in some of humanity's greatest challenges and opportunities, has long been a key motivation for many at the frontier of AI research to build more capable systems.[6] Although AI has been in use for many years, there has been concurrent advancement with large sets of accessible data, increased computational power, and newly developed machine-learning algorithms, and this confluence has accelerated the dissemination of machine learning such that it is broadly prevalent, even among non-technical users.[7]

The use of AI can be employed in a variety of fields one the leading areas are military and education for national security purposes and easement of conducting concepts and imparting clarity among students respectively. In recent years, technological developments have enabled the autonomous collection of massive amounts of data, including measurements, satellite and infrared imagery, and electronic signals, through ISR systems, such as unmanned aerial vehicles (UAV) and satellites.[8]

The amount of data is so large that its effective exploitation cannot rely on traditional human analysis for increased situational awareness. However, the systemic and unsupervised use of AI to process and analyze ever larger sets of data could potentially raise concerns of cognitive bias imported into the analysis – underlining the necessity to keep a human being in the analytical process.[9] The widespread adoption of AI could have a net effect on international stability in other ways. AI systems could change strategy in war, including by substituting machines for human decision-making in some mission areas, and therefore removing certain aspects of human psychology from parts of war.[10]

---

[6] James Manyika, 'Getting AI Right: Introductory Notes on AI & Society' (2022) 151(2) Daedalus 5 https://www.jstor.org/stable/48662023 accessed 28 December 2024.
[7] John D Winkler and others, 'The Future of Technology' in Reflections on the Future of Warfare and Implications for Personnel Policies of the US Department of Defense (RAND Corporation 2019) 7 http://www.jstor.org/stable/resrep20004.4 accessed 29 December 2024.
[8] Nadia Marsan and Steven Hill, 'International Law and Military Applications of Artificial Intelligence' in Andrea Gilli (ed), The Brain and the Processor: Unpacking the Challenges of Human-Machine Interaction (NATO Defense College 2019) 55 http://www.jstor.org/stable/resrep19966.12 accessed 29 December 2024.
[9] Ibid
[10] Kenneth Payne, Strategy, Evolution, and War: From Apes to Artificial Intelligence (Georgetown University Press 2018).

Modern AI applications can assist in adopting these educational technologies in the following aspects:

1. Providing modern AI-based learning analytics and adaptive learning[11]. For example, AI-based agents can collect personal information and predict learners' preferences or learning paths.[12] [13] [14] [15]

2. Facilitating modern AI-based interaction in VR/AR learning environments.[16] For example, AI-based games in VR/AR can better foster learners' immersion and interaction compared to games without AI. [17] [18]

3. effective computing/robotics with highly accurate modern AI models.[19] For example, some deep neural networks can be adopted for analyzing bio-feedback signals such as EEG or brainwaves, which are collected from effective computing devices.[20][21]

[11] Haoran Xie and others, 'Editorial Note: From Conventional AI to Modern AI in Education: Re-Examining AI and Analytic Techniques for Teaching and Learning' (2021) 24(3) Educational Technology & Society 85 https://www.jstor.org/stable/27032857 accessed 29 December 2024.
[12] Haoran Xie and others, 'Discover Learning Path for Group Users: A Profile-Based Approach' (2017) 254 *Neurocomputing* 59.
[13] K Almohammadi, H Hagras, D Alghazzawi, and G Aldabbagh, 'Users-Centric Adaptive Learning System Based on Interval Type-2 Fuzzy Logic for Massively Crowded E-Learning Platforms' (2016) 6(2) *Journal of Artificial* Intelligence and Soft Computing Research 81.
[14] D Zou and others, 'A Comparative Study on Linguistic Theories for Modelling EFL Learners: Facilitating Personalized Vocabulary Learning via Task Recommendations' (2021) 29(2) *Interactive Learning Environments* 270.
[15] J Wang and others, 'Top-N Personalized Recommendation with Graph Neural Networks in MOOCs' (2021) 2 Computers and Education: Artificial Intelligence 100010.
[16] Xie and others (n 13) 2
[17] E Rahimi and A Ahmadi, 'An AI-Based Tennis Game by Application of Virtual Reality Components' in *2017 Iranian Conference on Electrical Engineering (ICEE)* (2017) 2165, doi:10.1109/IranianCEE.2017.7985421.
[18] S Hammedi, F Essalmi, M Jemni, and AA Qaffas, 'An Investigation of AI in Games: Educational Intelligent Games vs Non-Educational Games' in *2020 International Multi-Conference on Organization of Knowledge and Advanced Technologies (OCTA)* (IEEE 2020) 1-4.
[19] Xie and others (n 13) 2
[20] SK Goh and others, 'Automatic EEG Artifact Removal Techniques by Detecting Influential Independent Components' (2017) 1(4) *IEEE Transactions on Emerging Topics in Computational Intelligence* 270.
[21] X Chen, X Tao, FL Wang and H Xie, 'Global Research on Artificial Intelligence-Enhanced Human Electroencephalogram Analysis' (2021) *Neural Computing and Applications*, doi:10.1007/s00521-020-05588-x.

4. Developing innovative learning applications with modern AI techniques.[22] For example, some recent AI techniques such as generative adversarial networks (GAN) can create new images, videos, or styles,[23] which can be employed in drawing learning.[24] [25]

AI AND AADHAAR

Aadhaar is the world's largest biometric-based identification system, developed by the Unique Identification Authority of India (UIDAI). It provides a unique 12-digit identity number to Indian residents, which their demographic and biometric data link. It captures and stores biometrics like fingerprints, iris scans, and facial in a centralized and highly secure database operated by the (UIDAI). Although Aadhar serves a variety of functions; leveraging its advanced technical infrastructure, service delivery, and digital governance there are many loopholes in the functioning and administration. We find that the law and regulations are vague and have failed to notify several important guidelines and processes on issues ranging from enrolment to security standards, which has resulted in various parts of the Aadhaar scheme operating in a legal vacuum.[26]

There exist no adequate performance accountability mechanisms. UIDAI is both a data controller (i.e., it controls the procedure/process of data usage) and a data protection authority (i.e., it is in charge of the grievance redressal process), The UIDAI thus wields extraordinary power over the lives of Aadhaar number holders, without adequate accountability mechanisms.[27] There have been several instances of breaches of the personal data of lakhs of Indian citizens, some through mishandling of data done by the government. In 2018 around 200 official government websites accidentally made personal Aadhaar data public; the problem was exacerbated to such a level, that one could access thousands of government databases with confidential information simply by Googling it.[28] On October 9th, a threat actor going by the alias 'pwn0001' posted a thread on Breach Forums brokering access to 815 million "Indian

[22] Xie and others (n 13) 2
[23] X Mao and others, 'On the Effectiveness of Least Squares Generative Adversarial Networks' (2019) 41(12) *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2947.
[24] Y Jin and others, 'GAN-Based Pencil Drawing Learning System for Art Education on Large-Scale Image Datasets with Learning Analytics' (2019) *Interactive Learning Environments*, doi:10.1080/10494820.2019.1636827.
[25] V Sorin, Y Barash, E Konen, and E Klang, 'Creating Artificial Images for Radiology Applications Using Generative Adversarial Networks (GANs)–A Systematic Review' (2020) 27(8) *Academic Radiology* 1175.
[26] Vrinda Bhandari and Renuka Sane, 'A Critique of the Aadhaar Legal Framework' (2019) 31(1) National Law School of India Review 72 https://www.jstor.org/stable/26918423 accessed 30 December 2024.
[27] Ibid
[28] Aadhaar Security Breaches: Here Are the Major Untoward Incidents That Have Happened with Aadhaar and What Was Affected' (Tech2)

Citizen Aadhaar & Passport" records. There have been similar leaks of the Cowin database leaked by a threat actor, exposing the personal information of individuals registered on the Cowin website for the COVID-19 vaccination. The leaked data included details such as AADHAAR numbers, PAN card information, mobile numbers, and home addresses.[29] This highlights the vulnerability of data protection and imparting AI algorithms would make it even out of the hands of officials to control breaches.

As Aadhaar involves the recording of biometric data, and taking into account that there are several instances of data leaks a policy can be adopted in line with GDRP which uses guidelines such as DPIA (data protection impact assessment). This is mandatory when using new technologies and the data processing is "likely to result in a high risk to the rights and freedoms of natural persons." Facial-recognition technology is likely to fit this description, DPIA is also required if there is "sensitive" data processing, including biometric data for identification purposes, on a large scale.[3031] DPIA, data controllers must assess the "necessity and proportionality" of the data processing and the risks to the rights and freedoms of the individuals concerned. Further, they must set "safeguards, security measures, and mechanisms" to mitigate these risks.[32] A similar regulatory framework should be developed and adopted in the Indian context as a collection of biometric data can be hijacked and misused which brings the fundamental rights conferred by the constitution at stake.

CONSTITUTIONAL RIGHTS AFFECTED BY AI

1. Right to Privacy (Article 21): As it is said that data is the new oil of the 21st century so there have to be channels protecting data and providing safe passage for transferring of data. Digital privacy is about the ability to shape one's own online identity and decide when, how, and where to share parts of that identity with people, companies, or other selected entities.[33] It ensures individuals can safeguard their identity, data, and personal choices, fostering freedom of thought, speech, and action at its core, privacy is essential for preserving human dignity,

[29] 'PII Belonging to Indian Citizens, Including Their Aadhaar IDs, Offered for Sale on the Dark Web' (Re-security) https://www.resecurity.com/blog/article/pii-belonging-to-indian-citizens-including-their-aadhaar-ids-offered-for-sale-on-the-dark-web accessed 28 December 2024.
[30] Els J Kindt, Transparency and Accountability Mechanisms for Facial Recognition (German Marshall Fund of the United States 2021) http://www.jstor.org/stable/resrep28527 accessed 2 January 2025.
[31] European Commission, General Data Protection Regulation (2016) art 35(3)(c).
[32] Els J Kindt, Transparency and Accountability Mechanisms for Facial Recognition (German Marshall Fund of the United States 2021) http://www.jstor.org/stable/resrep28527 accessed 2 January 2025.
[33] Nuala O'Connor and others, 'Privacy in the Digital Age' (2015) Great Decisions 17 http://www.jstor.org/stable/44214790 accessed 7 January 2025.

individuality, and trust in personal and societal relationships. The journey has been particularly long and roller-coaster in the Indian context from M.P Sharma (1954) To the Dpdpa Act (2023).

2. In M.P. Sharma[34] The court considered the scope of searches and seizures under the state's authority to conduct investigations without recognizing privacy as a protected constitutional right and reasoned that the Constitution did not explicitly guarantee a right to privacy.

3. Kharak Singh[35] Struck down intrusive domiciliary visits as unconstitutional majority held that privacy was not explicitly a fundamental right under the Constitution However, Justice Subba Rao's dissent laid the groundwork for the concept of personal liberty under Article 21.

4. The case of Govind v State of Madhya Pradesh (1975)[36] While the Court did not unequivocally declare privacy as a fundamental right, it acknowledged that certain aspects of privacy could be protected under Article 21 as part of the right to personal liberty. Stating that the right to privacy is not absolute and could be restricted stating the importance of proportionality a balance that needs to be maintained between individual good and larger public interest.

5. R. Rajagopal[37] SC addressed that the right to privacy extends to protecting an individual from both state and private intrusion, linking it to Article 19(1)(a) (freedom of speech and expression).

6. `In PUCL[38] The court directly tried to link privacy with Article 21, the Court held that telephonic tapping and surveillances constitute a serious intrusion into a person's private space. Emphasizing privacy SC explicitly said that infringement must adhere to the principles of legality, necessity, and proportionality.

7. SC made a transformative move in the case of K.S. Puttaswamy (Retd.) v. Union of India[39]. By unequivocally recognizing the Right to Privacy as a fundamental right under Article 21. Justice Chandrachud while writing the dissent noted that although technology has become a universal language that straddles culture and language, it has also reshaped dialogue between citizens and the state, and has the potential to confront the future of freedom and power itself,

[34] M.P. Sharma and others v Satish Chandra [1954] 1 SCR 1077.
[35] Kharak Singh v The State of UP & Others AIR 1963 SC 1295.
[36] Govind v State of Madhya Pradesh & Others AIR 1975 SC 1378.
[37] R Rajagopal & Others v State of Tamil Nadu & Others AIR 1995 SC 264.
[38] People's Union for Civil Liberties v Union of India & Others AIR 1997 SC 568.
[39] Justice K.S. Puttaswamy (Retd.) & Another v Union of India & Others AIR 2017 SC 4161.

Privacy concerns have also been fuelled by the almost unchecked practice, and capability, of both state and private actors to collect and process vast troves of metadata of individuals.[40]

It is a landmark judgment because it emphasizes privacy as an individual right. The right to privacy can be clubbed into a part right i) the Right to bodily and mental integrity ii) the Right to decisional autonomy iii) the Right to control personal information.[41]

The Digital Personal Data Protection Act, 2023 (DPDPA) represents India's most significant step towards establishing a comprehensive data privacy framework Act intends to control the processing of personal data by public and private entities and ensure accountability and transparency in the data-handling practices. Building on the European Union's GDPR, the DPDPA permits cross-border data transfers to jurisdictions outside of India other than those jurisdictions specifically identified by the Indian government on its list of countries to which data transfers are restricted.[42]

The DPDPA regulates the processing of digital personal data, i.e., personal data collected in digital form, or collected in non-digital form and subsequently digitized. Whilst the DPDPA's data definition is similar to that provided under the GDPR, it excludes from its scope personal data made publicly available by the data principal or by any other person under a legal obligation to make that data publicly available.[43] This differentiation appears to be made to refine the purview of the DPDPA, focusing the regulatory glare only on sensitive personal data and keeping away publicly available information from all regulatory compliance requirements. The DPDPA provides that data fiduciaries may lawfully process personal data only with the consent of the data principals or for certain specified "legitimate uses". Such legitimate uses include: the processing of personal data voluntarily shared by the data principal for a specified purpose (provided that the data principal does not object).[44]

[40] Vrinda Bhandari, 'Privacy Concerns in the Age of Social Media' (2018) 45(3/4) India International Centre Quarterly 66 http://www.jstor.org/stable/45129854 accessed 8 January 2025.
[41] Rohith S B and Sethupriya N, 'A Study on Impact of Artificial Intelligence on Right to Privacy in India' (2024) 4(3) Indian Journal of Legal Review (IJLR) 170, APIS–3920–0001, ISSN 2583-2344.
[42] Susan Ariel Aaronson, 'How AI Sovereignty Efforts May Distort Trade in AI' in The Age of AI Nationalism and Its Effects (Centre for International Governance Innovation 2024) 7 http://www.jstor.org/stable/resrep63493.9 accessed 9 January 2025.
[43] 'India's Digital Personal Data Protection Act 2023 vs the GDPR: A Comparison' (<www.lw.com>)
[44] Ibid

AI-DRIVEN BIASES IN HIRING AND LAW ENFORCEMENT

A set of instructions or commands used to carry out a particular operation is known as an algorithm, the algorithm analyses massive data patterns through data mining, searching, and using ways to predict, like our point of view encoded in the code. Algorithms frequently contain these biases due to the lengthy history of racial and gender prejudices, both intentional and unconscious.[45] The primary reason for these biases can be the data being analyzed by the algorithm is racially motivated or contains preconceived notions about a particular race or group of people. When biases exist in algorithmic data, AI may replicate these prejudices in its decision-making, a mistake known as algorithmic bias.[46]

One issue arises when datasets are skewed towards accessible and more "mainstream" groups due to the ease of data collection. If the collected data inadequately represent a particular race or gender, the resulting system will inevitably overlook or mistreat them in its performance.[47]

A research team at Princeton University discovered that algorithms lack access to the absolute truth. The machine corpus contains biases that closely resemble the implicit biases observed in the human brain.[48] The naïve analysis performed on the entire data set can easily produce skewed results depending on which mixture of populations is predominant in the data at the time. [49]After recollecting data, it is the functioning of the algorithm to work upon the invalid or biased data that has been filed this directly hinders a person's right to Equality under Article 14 as it snatches from the very beginning the chance which they have or to which they were legally entitled.

The AI algorithm-based discrimination is also based on the biases that are present in the developer because AI works on the instructions that are being provided by the developer. From setting goals for machine learning to selecting the appropriate model and determining data characteristics such as labels[50], all the commanding part is done by the developer. An engineer is responsible for developing the algorithmic model; If they hold certain beliefs and

[45] Z Chen, 'Ethics and Discrimination in Artificial Intelligence-Enabled Recruitment Practices' (2023) 10 Humanities and Social Sciences Communications 567 https://doi.org/10.1057/s41599-023-02079-x.
[46] MC Jackson, 'Artificial Intelligence & Algorithmic Bias: The Issues with Technology Reflecting History & Humans' (2021) 16 Journal of Business & Technology Law 299.
[47]Chen (n 46) 6
[48]Chen (n 46) 6
[49] DA McFarland and HR McFarland, 'Big Data and the Danger of Being Precisely Inaccurate' (2015) 2(2) Big Data & Society https://doi.org/10.1177/2053951715602495.
[50] supra note 36

preconceptions, those personal biases can be transmitted to the machine.[51] The Amazon hiring case illustrates this, where engineers considered education, occupation, and gender when assigning labels to the algorithm. When gender is considered the crucial criterion, it influences how the algorithm responds to the data.

Since biases are prevalent in any human language, language models are vulnerable to the same biases. Unfairness emanates from skewed behaviour that wrongly uses biases to create a certain outcome that discriminates against a certain group. When dealing with words describing gender, e.g., men and women, certain attributes can be ascribed to each category, significantly reinforcing stereotypes.[52] E.g. associating strong characteristics with men like resilience and toughness while attributing weak traits with women like softness and tenderness. In such cases, it is evident that the language model is not to blame for the bias, but rather the training.[53]

Garrido-Muñoz et al. describe a systemic approach to eliminate biases which are righty mentioned by Albaroudi et al. in five gist points.

1. Defining stereotype knowledge by identifying the protected properties and the related stereotyped aspects, enabling them to populate their stereotyped knowledge to identify potential biases that may harm the system.

2. Need for software engineers to evaluate the model to establish how it behaves with stereotyped and protected expressions.

3. Need for developers to analyze the results of the evaluation; This is meant to pinpoint the expressions or categories resulting in higher bias.

4. Software engineers must reevaluate the model and loop the last steps until they receive an acceptable response.

5. The procedure results should be reported by attaching model cards to attain transparent model reporting.

The Case of State vs. Loomis 2016[54] Stands out as a decision that marks a new chapter at the challenging crossroads of technology and law, especially in matters touching on the use of

[51] Sheilla Njoto, 'Gendered Bots? Bias in the Use of Artificial Intelligence in Recruitment
[52] E Albaroudi, T Mansouri and A Alameer, 'A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring' (2024) 5(1) AI 383 https://doi.org/10.3390/ai5010019.
[53] Ismael Garrido-Muñoz and others, 'A Survey on Bias in Deep NLP' (2021) 11(7) Applied Sciences 3184.
[54] Tate v Loomis, Docket No 42007 (Idaho Ct App, 22 December 2014).

artificial intelligence in criminal justice. The Wisconsin Supreme Court held that a trial court's use of an algorithmic risk assessment in sentencing did not violate the defendant's due process rights even though the methodology used to produce the assessment was disclosed neither to the court nor to the defendant.[55] At Loomis's sentencing hearing, the trial court referred to the COMPAS assessment in its sentencing determination and, based in part on this assessment, sentenced Loomis to six years of imprisonment and five years of extended supervision.[56]

Loomis argued against the decision and contented that COMPAS violates his rights and decisions given through the algorithm can be biased. Because COMPAS reports provide data relevant only to particular groups and because the methodology used to make the reports is a trade secret, Loomis asserted that the court's use of the COMPAS assessment infringed on both his right to an individualized sentence and his right to be sentenced on accurate information.[57]

Near-Repeat Hypothesis, a small number of individuals or households experience a disproportionate amount of crime and by focusing prevention resources on these repeat victims, the impact on crime will be greater than if entire communities were targeted.[58] Proponents of predictive policing, then, argue that such patterns are useful to focus police efforts.[59]

1. Officers would be able to focus attention on crime-prone areas;[60]

2. predictions would render police work more efficient and provide relief for overworked police officers – all based on statistical knowledge[61]

This development is welcome among some communities because it bears the promise of enhancing the precision of prediction and providing speedy, actionable results, also are the main driver for the use of algorithms for predictive policing. These developments are also met with criticism regarding several aspects; technological failure and cost inefficiencies as digital

[55] State v Loomis, Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing (https://harvardlawreview.org)
[56] Ibid
[57] Ibid
[58] Michael Townsley, Ross Homel and Janet Chaseling, 'Infectious Burglaries: A Test of the Near Repeat Hypothesis' (2003) 43(3) The British Journal of Criminology 615 https://doi.org/10.1093/bjc/43.3.615.
[59] M Kaufmann, 'Chapter 22: AI in Policing and Law Enforcement' in Handbook on Public Policy and Artificial Intelligence (Edward Elgar Publishing 2024) https://doi.org/10.4337/9781803922171.00031 accessed 17 January 2025.
[60] LW Kennedy, JM Caplan and E Piza, 'Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modelling as an Algorithm for Police Resource Allocation Strategies' (2011) 27(3) Journal of Quantitative Criminology 339.
[61] C Beck and C McCue, 'Predictive Policing: What Can We Learn from Wal-Mart and Amazon About Fighting Crime in a Recession?' (2009) 76(11) Police Chief 18.

illiteracy among officers regarding the usage of technology.[62] A more fundamental concern relates to the role of the state and democratic accountability, Mark Andrejevic highlights the role of predictive policing as a catalyst in the "collection of data without limits"[63], which leads to a general increase in surveillance.[64]

Right of equality (Article 14): Within this pre-crime approach in public policy, large databases and algorithmic calculation engender a governance by patterns and categorisations that are the basis for predictions The decision-making processes that inform these patterns are a part of complex technological developments and remain largely invisible. This underlines the relevance of studying the social life of prediction algorithms, such as "the nature of that work, how it is organised day-to-day, what tacit understandings are built into this organisation, its situatedness within a network of other organisational arrangements"[65] What lies in a name comes with an expectation or association, in the worst case a prejudice.[66] For example, an algorithm monitoring historical data tends to categorise an area as 'high risk', which creates an expectation that crime would most possibly occur and influences the behaviour of the law-enforcement personnel, even though predictions might lack whole casuistry.

Similarly, individuals identified by such systems may face enhanced scrutiny purely based on a predictive label, irrespective of their actual behaviour. This framework can reinforce stereotypes, especially when the datasets reflect systematic bias, for example: when certain communities like marginalized communities receive greater attention from the police.

Marda and Narayan (2020) highlight that records in a criminal database (not a conviction database) lend themselves more readily to some areas of the city and sections of society.[67] Account of employees suggests that people from posh areas are "hardly called".[68] This reinforces the cyclic pattern of crime among the people of slums or low social strata leading to more heightened scrutiny and discrimination.

---

[62] M Kaufmann, 'The Co-Construction of Crime Predictions: Dynamics Between Digital Data, Software and Human Beings' in HO Gundhus, KV Rønn and NR Fyfe (eds), Moral Issues in Intelligence-Led Policing (Routledge 2018) 143.
[63] M Andrejevic, Automated Media (1st edn, Routledge 2019) https://doi.org/10.4324/9780429242595.
[64] Townsley, Homel and Chaseling (n 59) 8
[65] Ibid
[66] Mareile Kaufmann, 'Chapter 22: AI in Policing and Law Enforcement' in Handbook on Public Policy and Artificial Intelligence (Edward Elgar Publishing 2024) https://doi.org/10.4337/9781803922171.00031.
[67] Vidushi Marda and Shivangi Narayan, 'Data in New Delhi's Predictive Policing System' in FAT '20: Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, 27–30 January 2020, Barcelona, Spain (ACM 2020) 1, 8 https://doi.org/10.1145/3351095.3372865.
[68] Ibid

Crimes are more likely to be recorded when they come from organized colonies and have a higher likelihood of getting their grievances documented; whereas crimes from shanty settlements are plotted at the same spot due to a lack of accurate information, leading to an imbalance in what is classified as a "hotspot" of crime, which in turn leads to over-policing areas inhabited by individuals from vulnerable groups, and also creates a cycle of confirmation bias within an institution that is already embedded with societal, cultural, gender and caste biases.[69]

Marda and Narayan's (2020) research highlights the need for a comprehensive system, and the importance of focusing these assessments on institutional formalities and standard operating procedures before analyzing the sociotechnical system itself proposed that the system of predictive policing should be studied through the lens of institutional culture and limitations within which it will function,[70] The reason is institutional culture—encompassing attitudes toward technology, accountability, and law enforcement priorities—can significantly influence the effectiveness and fairness of these systems if observed independently of each other, also proposed approach focusing on basic aspects surrounding these systems;

1. Acquiring knowledge of the particular public sector institution regarding how the institution operates, how information is documented and communicated within the institution, brights lines for discretionary action (algorithm can be designed to automatically flag or process decisions based on predefined rules, minimizing subjective judgment) also redressal processes and systems in place for addressing complaints and concerns from stakeholders (e.g., employees, citizens). In which the algorithm and operation will operate.[71]

2. The procurement process should be transparent and should strictly comply with the specifications required from the system the process must include provisions for regular independent audits; and performance reviews.[72]

3. Governance limitations at institutional and bureaucratic levels should be taken into account to filter and rectify any biases before employing any AI algorithm.[73]

[69] Vidushi Marda and Shivangi Narayan, 'Data in New Delhi's Predictive Policing System' in FAT '20: Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, 27–30 January 2020, Barcelona, Spain (ACM 2020) 1, 8 https://doi.org/10.1145/3351095.3372865.
[70] Ibid
[71] Ibid
[72] Ibid
[73] Ibid

4. Algorithmic Impact Assessments should be institutionalized as a prerequisite for deployment so that respective developers can formulate documents entailing any unforeseen risks or consequences to the possibly inflicted harm and timelines on which these harms may be mitigated.

FREEDOM OF SPEECH AND EXPRESSION (ARTICLE 19)

It is important to note that the application of artificial intelligence in the online media environment can have both positive and negative implications for individuals' right to freedom of expression. Despite the advantages that AI moderation has, in terms of limiting hate speech, and harmful and false information, there are equally issues with over-censorship as well as bias in algorithms.

Blocking while using automated responses is used for the analysis of images and blocking texts can be done through the selection of "toxic" and harmful content and labelling them in a particular fashion this can be understood in the context of NLP (natural language processing) which plays a critical role in automated text correction by enabling machines to understand, analyze, and modify human language with precision. A similar outline can be drawn for the automated image generation. Tools can also be designed to classify whether an image contains a feature such as nudity; One approach to detecting nudity in an image has been to analyze the proportion of pixels in an image that fall into a specific colour range that has been pre-identified as representing skin colour; This kind of tool is vulnerable to misclassification of underrepresented skin tones and of objects or scenes with the same colour palette as the training data.[74]

Image generation also notably includes the area of "deepfakes," composite videos and images created based on real footage that portray fictional statements and actions.[75] Powered by Generative Adversarial Networks (GANs) Different from normal video editing methods, deepfake utilizes encoder-decoder architectures that enable facial movements to be mapped with much higher precision including voice patterns, thus making it even harder to detect, enabling hyper-realistic manipulation of audio, video, and images by synthetically altering or generating human likenesses. Deepfakes can threaten rights to privacy and dignity.[76] Many

[74] Ibid
[75] Ibid
[76] 'Deepfake Nude AI App "DeepNude" Shut Down After Backlash' The Verge (27 June 2019) https://www.theverge.com/2019/6/27/18760896/deepfake-nude-ai-app-women-deepnude-non-consensual-pornography

technology researchers believe that deepfakes realistic-looking content developed using machine learning algorithms will herald a new era of information warfare. In 2019, a deepfake video of U.S. Speaker Nancy Pelosi surfaced online, portraying her as if she were intoxicated while delivering a speech. Several governments are now proposing to have platform recommendation algorithms accommodate public interest considerations and legal requirements, and platforms are taking comparable measures on their initiative.[77] Without adequate safeguards, government regulation of content recommendation could impede the freedom of expression of social media users. Prescribing what should be downranked risks becoming a form of censorship, and what must be prioritized is a form of propaganda.[78]

Any framework governing algorithmic content prioritization should avoid broad, vague mandates that could be exploited to suppress dissent or manipulate public discourse A balance between addressing misinformation and ensuring freedom of expression is key to regulatory approaches that aim for autonomy along with accountability without restricting the diversity of public discourse.

RECOMMENDATIONS

Bias Mitigation and Intersectional Representation: To mitigate algorithmic discrimination based on race, gender, and intersecting identities, steps need to be taken against systemic biases that are embedded in training data. Today most datasets are likely to underrepresent marginalized groups or conflate identities into grossly reductive categories (e.g., binary gender classifications or overly limited skin-tone ranges), all of which will seep into biases held by machine learning models. An intersectional evaluation further requires a dataset representing the defined genders with a range of phenotypes that enable subgroup accuracy analysis.[79]

Images poorly exposed due to sensor calibrations can pose difficulties for automated facial recognition systems. By labeling faces with skin type, we can increase our understanding of performance on this important phenotypic attribute.[80] Protocols collected data have to always give precedence to inclusive sampling through working alongside demographically diverse

[77] Emma Llansó and others, 'Artificial Intelligence, Content Moderation, and Freedom of Expression' (2020) Google Scholar Reference 1.
[78] Ibid
[79] Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' in Proceedings of the Conference on Fairness, Accountability and Transparency, PMLR, 2018.
[80] Ibid

communities and employing tools such as the Monk Skin Tone.[81] (an MST) dataset to codify phenotypic variation. Continuous monitoring of model performance through real-time fairness dashboards is critical to ensuring algorithmic systems remain equitable and accountable post-deployment These Fairness Indicators, enable developers to track fairness metrics, integrating dashboards into ML pipelines allows for proactive interventions, such as retraining models with updated data or adjusting decision thresholds to mitigate harm.

Human-Rights/Ethical Considerations: As artificial intelligence (AI) processes are now being better incorporated into the decision-making process across domains such as law enforcement, finance, healthcare, and employment, the concern for its grave potential impact on fundamental human rights has also grown. While AI has the potential to enhance efficiency and innovation, its governance must be rooted in human rights principles, ensuring transparency, fairness, and accountability. There exist several legal frameworks, including the EU AI Act the OECD AI Principles, Asilomar guidelines of 2017 which attempt to regulate some of the impacts of AI; yet, there remain hurdles that need to be overcome in the areas of eliminating algorithmic discrimination, ensuring explainability, and providing avenues for legal redress for harm inflicted by AI.

Human rights considerations must make themselves felt in the design, application, and evaluation of AI products, as well as in all government procurement and deployment of AI.[82] Given AI's borderless nature collaboration between global and local governments to ensure the ethical development, transparent deployment, and unbiased regulation of artificial intelligence systems. International cooperation is crucial for setting standardized regulations, such as those outlined in the EU AI Act, OECD AI Principles, and UN AI frameworks. To fulfill their obligation to protect citizens, public authorities also need to step up their vigilance regarding what private companies are doing with AI.[83]

Regulatory Framework: An all-encompassing AI regulation framework established by the Government should include legally binding oversight mechanisms, mandates for algorithmic transparency, and strict bias mitigation procedures to guarantee ethical and accountable AI implementation. Because the regulatory part of the algorithm can’t be made up to private players entirely because, the shortcomings of self-regulatory approaches however, are well-

[81] Monk, Ellis. "The monk skin tone scale." (2019).
[82] Eileen Donahoe and Megan MacDuffee Metzger, ‘Artificial Intelligence and Human Rights’ (2019) 30(2) Journal of Democracy 115 https://dx.doi.org/10.1353/jod.2019.0029
[83] Ibid

established and are best summarized by "the motivation problem": even when industry actors have superior information and expertise advantages to devise regulatory solutions, they do not necessarily have the incentives to do so – to self-regulate in ways consistent with public regulatory goals rather than with their own private individual interests.[84] [85] [86]

The imposition of Algorithmic Impact Assessments (AIA) should be made mandatory before the deployment of AI into crucial sectors such as law enforcement, hiring, health care, and financial services and be subject to third-party assessments that weigh in on AI bias, accuracy, and fairness. The development of transparency requirements would further compel AI developers to disclose decision-making processes to ensure explainability in such high-stakes use of AI technologies as automated hiring or predictive policing. Another thing that governments need to do is put in place HITL, which stands for human-in-the-loop systems for high-risk applications of AI (Human-in-the-loop aims to train an accurate prediction model with minimum cost by integrating human knowledge and experience. Humans can provide training data for machine learning applications and directly accomplish tasks that are hard for computers in the pipeline with the help of machine-based approaches).[87]

This is to place a human review and contest of the AI-enabled decisions. Ascribing to AI ombudsman bodies and legally embedding the "Right to Explanation" will allow people affected by algorithmic decisions to challenge these decisions most notably concerning employment, credit opportunities, and law enforcement interventions. AI-powered tools can assist ombuds institutions in processing and analyzing large volumes of data, identifying patterns, and detecting potential systemic issues enhancing their efficiency and effectiveness.[88] Prioritizing fairness, transparency, and human rights while preventing the unchecked deployment of biased or opaque algorithms.

[84] Robert Baldwin, Martin Cave and Martin Lodge, Understanding Regulation: Theory, Strategy, and Practice (Oxford University Press 2012).
[85] Martin Lodge and Kai Wegrich, Managing Regulation: Regulatory Analysis, Politics and Policy (Palgrave Macmillan 2012).
[86] Madalina Busuioc, 'AI Algorithmic Oversight: New Frontiers in Regulation' in Handbook of Regulatory Authorities (Edward Elgar Publishing 2022) 470–486.
[87] Wu, Xingjiao, et al. "A survey of human-in-the-loop for machine learning." Future Generation Computer Systems, vol. 135, 2022, pp. 364-381.
[88] Madise, Ülle, and Kertti Pilvik. "The role of Ombuds institutions in a rush of digitalization." La Ley del Ararteko. Construyendo el futuro: una reflexión sobre las defensorías del pueblo: 2023ko ekainaren 12an eta 13an Vitoria-Gasteizen egindako mintegia = Seminario celebrado en Vitoria-Gasteiz los días 12 y 13 de junio de 2023. Eusko Legebiltzarra = Parlamento Vasco, 2024.

CONCLUSION

It has become an indispensable part of present-day governance, business, and people. Its applications include law enforcement, finance, healthcare, and so on, such as Aadhaar, a digital identity system. Although artificial intelligence propels users toward efficiency, automation, and data-based decision-making, it also offers glaring chances for the risk of bias, discrimination, and privacy breaches, some kinds of unchecked or unregulated application of AI systems are bound to lead to algorithmic biases in hiring and policing, mass surveillance, and, thus, violations of fundamental rights, including privacy (Article 21), equality (Article 14), and freedom of expression (Article 19).

A much-needed regulatory framework, among others, should emphasize algorithmic transparency, independent audits, human intervention, and fairness metrics. Algorithmic impact assessments (AIAs) should all be made compulsory before use in critical areas such as law enforcement and recruitment, to ascertain the dangers posed by AI decision-making. Further, legal provisions, like the Digital Personal Data Protection Act (DPDPA) in India, should be reinforced for data security and user rights. Since AI is a global phenomenon, national regulatory agencies and international cooperation must evolve routes to accountability standards. Lastly, intersectional data representation, ongoing fairness monitoring, and human-in-the-loop (HITL) oversight mechanisms will ensure that there are sufficient approaches for program optimization toward addressing the challenges of bias. Without adequate safeguards, ruling AI will only exacerbate societal inequalities; however, with responsible governance, it can become a very strong tool for good transformation upholding human rights and democratic values.