Himachal Pradesh National Law University, Shimla (India)
HPNLU Law Journal

Volume IV *ISSN: 2582-8533* *April 2023-March 2024*

Volume IV (2023)

PROMOTING 'AI' INCLUSIVITY IN INDIA: A Progressive Legal Model to Mitigate Bias

Titiksha Narkhede

DOI: https://doi.org/10.70556/hpnlu-lj-v4-2023-11

This article can be downloaded from: https://www.hpnlu.ac.in/journal-level-3.aspx?ref-id=45

Recommended Citation:
Titiksha Narkhede, *Promoting 'AI' Inclusivity in India: A Progressive Legal Model to Mitigate Bias* IV HPNLU. L. J. 193 (2023).
https://doi.org/10.70556/hpnlu-lj-v4-2023-11

This Article is published and brought to you for free and open access by Himachal Pradesh National Law University, Shimla. For more information, please contact lawjournal.editor@hpnlu.ac.in

Contents

Volume IV | *ISSN: 2582-8533* | *April 2023-March 2024*

PROMOTING 'AI' INCLUSIVITY IN INDIA: A Progressive Legal Model to Mitigate Bias

Titiksha Narkhede*

[Abstract: *Artificial Intelligence (AI) is dualistic, where it has a significant impact on emerging technological inventions and the danger of deepening pre-existing inequalities. India is a combination of diverse landscapes and an exponential growth rate of implementation of AI systems in various fields. This raises critical concerns about inclusivity, transparency, and accountability of, by, and towards stakeholders involved in the deployment and regulation of AI-operated systems. This article identifies the limitations of India's current AI regulation in addressing bias, compares it with global principles, particularly from the European Union (EU) legislature on Artificial Intelligence (EU AI Act), and incorporates elements in a progressive legal model to prevent the perpetuation or amplification of existing societal, cultural, or economic inequalities into the AI systems.*]

I

Introduction

AI is a multidisciplinary field focused on creating systems that are capable of simulating human-like intelligence. It is often in general parlance described as a machine or software that mimics human intelligence, a key characteristic of AI is its ability to replicate behaviour humans perceive as intelligent.[1] It is defined through its methodologies and objectives as the science and engineering of making intelligent machines.[2] In more modern frameworks, it is broadly categorised into narrow and

* The author is a student at Maharashtra National Law University, Mumbai. Email: 2024m041@mnlumumbai.edu.in

1 Joost Kok, *et al.*, *Artificial Intelligence: Definition, Trends, Techniques, and Cases*, ENCYCLOPAEDIA OF LIFE SUPPORT SYSTEMS, *available at*: https://www.eolss.net/sample-chapters/c15/e6-44.pdf. (last visited on Dec. 10, 2024).

2 Dalvinder Singh Grewal, *A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering* 16 IOSR JOURNAL OF COMPUTER ENGINEERING (IOSR-JCE) 09 (2014) *available at*: https://professionalismandvalue.org/wp-content/uploads/2021/02/A_Critical_Conceptual_Analysis_of_Definitions_of_Artificial_Intelligence_as_Applicable_to_Computer_Engineering.pdf (last visited on Dec. 10, 2024).

general AI.[3] These definitions often highlight underlying processes like learning, reasoning, and perception, intrinsic to the technique, such as neural networks and deep learning algorithms. The EU AI Act[4] defines AI as software developed with machine learning, logic-based, or statistical approaches, capable of generating outputs such as predictions or decisions affecting the environment they interact with.[5] AI's diverse definitions underscore its dynamic nature, evolving with advancements in technology and societal expectations.

To understand the nature of AI, it is pertinent to delve into the evolution of AI since its genesis. The emergence of AI as a field of study was only after it marked its quest to replicate and copy cognitive functions associated with human intelligence. The theoretical idea emerged with the development of theory on cybernetics, logic, and machines simulating human thoughts through models, analogies, and abstractions. Thereafter, the invention of the digital computer provided the infrastructure to realise these theories and marked the beginning of this ambitious project with the development of domains of AI driven by mathematics, computation, and philosophical breakthroughs. Further, milestones such as the Dartmouth Conference[6] led to the formalisation of AI concepts by leading researchers like John McCarthy, Herbert Simon, and Allen Newell.[7] These technologies back then relied on heuristic methods to tackle complex problems, hence a need to acquire their reasoning by analogies and prior experiences was seen. Nonetheless, these principles laid down the groundwork for early AI systems like theorem solvers, game-playing algorithms, and the interpretation of simple natural language. During the initial encounter of AI systems with humans, it was often argued that this assumption of AI replicating human intelligence is flawed since the human brain is deeply connected to sensation and interaction with the world. Which a machine is incapable of having, and was criticised for trying to break down intelligence into rules and formulas, by stating that this approach ignored the rich,

---

3 Pei Wang, 'On Defining Artificial Intelligence' (2019) Journal of Artificial General Intelligence 1 <10.2478/jagi-2019-0002> accessed 10 December 2024

4 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 Artificial Intelligence Act, 2020 *available at*: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (last visited on Dec. 10, 2024).

5 Vladislav V. Arkhipov, *Definition of Artificial Intelligence in the Context of the Russian Legal System: A Critical Approach* GOSUDARSTVO I PRAVO (2022) 168 <10.31857/S102694520018288-7> (last visited on Dec. 10, 2024).

6 Nivash Jeevanandam, *Exploring the Significance of the Dartmouth Workshop*, INDIAAI (28 Apr., 2022) *available at*: https://indiaai.gov.in/article/exploring-the-significance-of-the-dartmouth-workshop (last visited on Dec. 10, 2024).

7 Marvin L Minsky, *Artificial Intelligence* 215(3) SCIENTIFIC AMERICAN 246 (1966) *available at*: https://www.jstor.org/stable/24931058.

flexible way humans think and understand things.[8] These arguments, however, today have become redundant since AI has made great progress, especially in tasks like the recognition of faces, translating language, and generating texts. Nonetheless, a lack of emotions or deep common sense is to date a thin line of distinction between an AI and a human being.

Every evolving technology comes with its challenges, back then in the 90s a few issues in AI were identified, but with a boom in AI technology today the scope of its malfunction has widened to include critical areas like (i) data privacy and protection, (ii) accountability and liability, (iii) transparency and explainability, (iv) unethical use of AI, (v) employment and economic impact, (vi) security and cyber threat, (vii) intellectual property and copyright infringement, (viii) human-AI interaction and autonomy, (ix) cross border and international regulations, and (x) bias and discrimination.

This article dives into one of the key issues out of the ones mentioned above, which have been prevailing and threatening the systems, i.e., bias. Back in 1983, Reading[9] examined how AI could be used to analyse and interpret social issues, and their peripheral concerns, particularly around bias. Despite improvements in AI, both personal and social biases continue to be a part of AI technologies, therefore, influencing how these systems process and interpret social data, resulting in the difficulty of achieving completely neutral AI. However, it was only in 1990 that, for the first time, philosophical, psychological, and ethical aspects surrounding AI were explored by Henley,[10] and the issues encountered while developing AI with a focus on the influence of a human being's inherent bias were addressed.

While analysing if AI systems adopt biases, it is necessary to first understand what the word bias implies. There are around 27 identified classifications of biases, out of which the ones that AI systems commonly catch are as follows.

Firstly, cognitive bias emerges from different ways in which the mind processes information, thereby leading to skewed judgments and decisions. They can be reinforced in systems by the use of algorithms reflecting faulty cognitive tendencies. For instance, a type of cognitive bias called confirmation bias.[11] It is commonly recorded to exist in applications and websites that use AI models designed to analyse user history, data, behaviour, etc., and predict content aligned with their interests. Cognitive bias can

---

[8] Hubert L Dreyfus, *Artificial Intelligence* 412 Sage Publications & American Academy of Political and Social Science 21 (1974) *available at*: https://www.jstor.org/stable/1040396.

[9] Hugo F Reading, *Artificial Intelligence* 18 ROYAL ANTHROPOLOGICAL INSTITUTE OF GREAT BRITAIN AND IRELAND 183 (1983) *available at*: https://www.jstor.org/stable/2801770.

[10] Tracy B Henley, *Natural Problems and Artificial Intelligence* 18 CAMBRIDGE CENTER FOR BEHAVIORAL STUDIES (CCBS) 43 (1990) *available at*: https://www.jstor.org/stable/27759223.

[11] Raymond S Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises* 2 REVIEW OF GENERAL PSYCHOLOGY 175 (1998) *available at*: https://pages.ucsd.edu/~mckenzie/nickersonConfirmationBias.pdf.

distort rational judgments and result in unfair outcomes due to a lack of exposure to diversified training datasets during their development.[12]

Secondly, social bias is introduced in the systems from data fed to them during their training, which contains societal stereotypes, like, for instance, facial recognition often performs poorly for women and minority groups. A theory as put forth by Glassner[13] on how fear-driven social dynamics contribute to these biases, on which the models then pick is a valid explanation as such practices reinforce inequality on multidimensional levels, in sensitive areas such as healthcare, education, recruitment, law enforcement, and finance and thereby unfair impact on certain groups.[14]

Thirdly, implicit bias unconsciously sneaks into the systems because of data and algorithms, reflecting stereotypes in society like those surrounding gender, race, etc. Since they are extremely difficult to identify and have significant real-world consequences, like unfair hiring, loan approvals, etc., it is crucial to address them.[15]

Fourthly, statistical bias leads to the exacerbation of existing inequalities in minority groups due to selection bias. Where the sample data does not fully represent the broader population, skewed predictions/ results are generated. Hindsight bias in statistical models can distort historical data interpretations, affecting the AI's future predictions.[16] Therefore, careful data selection is a key to overcoming the issue and delivering more accurate and equitable outcomes.[17]

Fifthly, media bias occurs when algorithms are designed with political, commercial, or sensational biases, distorting public perception and spreading misleading narratives. Media outlet's[18] motives shape content, which is amplified using AI, creating a feedback loop. This reinforces user beliefs, fuelling confirmation bias[19] and limiting exposure to diverse viewpoints.

Sixthly, cultural bias arises when systems are trained by data reflecting a specific cultural perspective, often ignoring others. The choice of words influenced by cultural norms

---

12 Amos Tversky and Daniel Kahneman, *Judgment under Uncertainty: Heuristics and Biases* 185 AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE 1124 (1974) *available at*: https://www2.psych.ubc.ca/~schaller/Psyc590Readings/TverskyKahneman1974.pdf.

13 Barry Glassner, THE CULTURE OF FEAR – WHY AMERICANS ARE AFRAID OF THE WRONG THINGS (1999).

14 Dan Ariely, *Predictably Irrational available at*: https://radio.shabanali.com/predictable.pdf.

15 Elizabeth F Loftus, 'Creating False Memories' (1997) 277 Scientific American, a division of Nature America, Inc. 70. https://www.jstor.org/stable/24995913 accessed 12 December 2024

16 Neal J Roese and Kathleen D Vohs, 'Hindsight Bias' VII Assn. Psychol. Sci. 5 (2012) <https://doi.org/10.1177/1745691612454303> accessed 13 December 2024

17 Richard H Thaler, MISBEHAVING - THE MAKING OF BEHAVIORAL ECONOMICS (2015)

18 Ben H Bagdikian, THE NEW MEDIA MONOPOLY (2004).

19 *Supra* at 11.

affects decision-making.[20] For example, if an AI is developed in Western culture and is used in non-Western culture, it ought to fail. They unknowingly lead us to systemic inequalities. Therefore, to enable a global usage of AI systems, the diversification of datasets is the solution.[21]

This in-depth analysis of the concept of AI, and the core issue surrounding it, it can be concluded that the bias issue AI might face is essentially enabled by the data fed, and the algorithm it thereby forms, hence to overcome it, the algorithms used, are to be regulated, and thereby build trust in the technology. However, this can be enabled only with the inclusion of standards of inclusivity, transparency, and accountability. Further, to address these challenges, a legal regulatory model for AI is essential. The next part of the article delves into the jurisprudence around AI regulation, focusing on finding the best approach for India. Since, in India, if AI development is not carefully managed, the existing biases in society could be carried over into the systems and could create a much greater risk. Since biases, once embedded in digital systems, become harder to identify and prove, and thereby ultimately raise questions of fairness and reliability on AI technology.

II

Theoretical Nomenclature of Legal Models for Law & Regulation of AI

The issue of bias that AI faces can be addressed once the approaches for creating a law that can best tackle the problem are known; therefore, understanding the models of law and regulation of AI is a must, only then can the best-fit model be identified and implemented. The classification of these models can be based on different angles they aim to focus on while regulating AI, that is, based on (A) Approach towards Regulating AI, (B) Mechanism and Scope of Regulating AI, and (C) Integrated Governance of AI.

The approach towards regulating AI firstly contains a risk-based approach. It is a unique approach to bifurcate AI, based on its degree of risk in case of malfunction, in society. This is a blanket approach to categorize types of AI systems into (i) low-risk, (ii) medium-risk, and (iii) high-risk, wherein regulation of those falling in a low-risk zone will have minimal regulatory compliance and penalty on the breach, whereas those falling in a high-risk zone will have strict compliance and hefty penalty for breach. This approach is used in the EU and Singapore. Secondly, the ethics-based approach promotes infusing AI development and deployment with ethical principles. The focus here is on the values conveyed into the systems by their designers. Irrespective of being

[20] *Supra* at 12.
[21] *Supra* at 14.

non-binding guidelines, they are highly effective as they align with global principles and guidelines like the OECD[22] and the IEEE, respectively.[23] Thirdly, a human-centric approach promotes the alignment of systems with fundamental principles and human rights. A model created keeping in mind this approach shall prevent violation of equality, privacy, and liberty principles like UNESCO's AI Ethics Recommendations,[24] and the EU Charter of Fundamental Rights.[25] Fourthly, a market-driven approach leaves the ethical development and use of AI to the market forces, with minimal government intervention. The companies are left by themselves to determine ethical guidelines for AI. In the US, the focus of the government is mainly on the innovation part of AI, and these guidelines are left for the companies to decide. Public opinion and consumer preferences then determine how good the guidelines of the business are, thus directly affecting the reputation of the business.[26] Fifthly, a principle-based approach emphasises the development of AI guided by overarching ethical principles like transparency and accountability. It provides flexibility to the organisations to adapt these principles to their contexts. Canada's directive on automated decision-making[27] and Singapore's model of AI governance[28] follow a principle-based approach. Sixthly, a prescriptive approach is a rule-based framework where detailed and binding standards govern the systems. It minimises ambiguity and provides for an explicit compliance scheme ensuring utmost transparency and accountability. For instance, the EU's audits, penalties, and certification requirements for high-risk AI systems employ a hybrid approach, a combination of multiple approaches, while developing one

---

[22] OECD, *OECD AI Principles Overview* (May 2024) *available at*: https://oecd.ai/en/ai-principles.

[23] IEE SA, *Autonomous and Intelligent Systems (AIS) Standards*, *available at*: https://standards.ieee.org/initiatives/autonomous-intelligence-systems/standards/ (last visited Dec. 10, 2024).

[24] UNESCO, RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE (Nov. 23, 2021) *available at*: https://unesdoc.unesco.org/ark:/48223/pf0000381137.

[25] Charter of Fundamental Rights of the European Union (2000) OFFICIAL JOURNAL OF THE EUROPEAN COMMUNITIES *available at*: https://www.europarl.europa.eu/charter/pdf/text_en.pdf.

[26] Raul Alan and Mushka Alexandra, *The U.S. Plans to "Lead the Way" on Global AI Policy*, LAWFARE (Feb. 13, 2024) *available at*: https://www.lawfaremedia.org/article/the-u.s.-plans-to-lead-the-way-on-global-ai-policy.

[27] Treasury Board of Canada, DIRECTIVE ON AUTOMATED DECISION-MAKING (2021) *available at*: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592 (last visited Dec. 10, 2024).

[28] AI Verify Foundation (AIVF) and Infocomm Media Development Authority (IMDA), *Singapore Proposes Framework to Foster Trusted Generative AI Development*, *available at*: https://www.sgpc.gov.sg/api/file/getfile/Press%20Release%20Model%20AI%20Gov%20Framework%20GenAI.pdf?path=/sgpcmedia/media_releases/imda/press_release/P-20240116-1/attachment/Press%20Release%20Model%20AI%20Gov%20Framework%20GenAI.pdf (last visited Dec. 10, 2024).

comprehensive regulatory framework, which is an efficient way to include multiple things in one law. For instance, the EU AI Act is a law based on a hybrid approach to regulating AI.[29]

Thereafter, the mechanism and scope of regulating AI consists firstly of a self-regulation is an influential mechanism wherein assessing how effective and good corporate governance of a business can be done. In this method, the companies themselves establish and enforce certain guidelines, principles, and systems without any government mandate. Microsoft's Responsible AI Guidelines[30] and Google's AI Principles[31] are prevailing examples of this approach. Secondly, co-regulation is a collaborative attempt by the government and the business stakeholders to jointly develop regulations. It merges self-regulation with public authority. It is an effective method since ongoing communication between the lawmakers and the relevant stakeholders can ensure the maximum utility of the law. The Australian ethics principles are drafted with inputs from various stakeholders, and the EU's GDPR is another example where constant involvement from both the public and private sectors is seen. Thirdly, a command-and-control regulation is a model where direct, strict rules and regulations on business are enforced by the government. It is a top-down approach with rules that must be followed and penalties for non-compliance. Fourthly, soft regulations are those regulations that are contrary to the command-and-control approach, wherein by enabling flexibility with the guidelines, businesses follow the ethical and operational standards voluntarily. Fifthly, an incentive-based regulation is an approach that motivates the business to follow certain standards for obtaining certain benefits, be it monetary, goodwill, or certification. Perks given on successful compliance encourage the stakeholders to abide by the compliance, rather than a penalty-based approach, it provides a friendly framework, and has a decentralised regulation has a structure similar to delegated legislation, wherein the responsibility of regulation is distributed across multiple stakeholders, local government, civil society, industry leaders, etc. For instance, blockchain and smart contracts use technology that enables automatic enforcement of AI regulation through decentralised applications, as well as local government initiatives.

Further, the integrated governance of AI contains, firstly, sector-specific regulations, which are industry-specific rules and regulations that are intrinsic to a particular field with regard to their influence or effect on society. Based on the severity of the impact caused these standards are set and followed. For instance, the healthcare system will

---

[29] Nicolas Petit and Jerome De Cooman, *Models of Law and Regulation for AI*, EUROPEAN UNIVERSITY INSTITUTE, *available at*: https://cadmus.eui.eu/handle/1814/68536 (last visited Dec. 10, 2024).

[30] Microsoft, *Microsoft Responsible AI Principles and Approach*, *available at*: https://www.microsoft.com/en-us/ai/principles-and-approach (last visited Dec. 10, 2024).

[31] Google, *AI Principles: Our approach to building beneficial AI*, *available at*: https://ai.google/responsibility/principles/ (last visited Dec. 10, 2024).

require standards and rules which will be different from those of finance, since the risks and challenges in these two fields are different. The EU Medical Device Regulations,[32] which govern those devices powered by AI, are a sector-specific attempt at regulation by the EU. Similarly, the Financial Stability Board has promulgated principles on AI.[33] Secondly, sector-agnostic regulation is a unified set of principles that are applied across various sectors. They focus on common transparency, accountability, etc., principles that remain common throughout various sectors. The OECD principles of AI[34] are one example of sector-agnostic regulation. Thirdly, a prohibition-based regulation prohibits certain AI systems that are considered too risky or unethical. Those beyond the scope of the fundamental principles of a country. For instance, China's approach to facial recognition and surveillance, and the EU's model wherein such AI, having the capacity of social scoring, are banned from being developed. Fourthly, with a global collaborative model promotes international cooperation for governing AI, since AI is a type of technology and technology after a limit does not differ from area to area, and is universally similar, the principles and laws governing it should also be of such nature, that the globe aligns with them, that is possible by global collaborative efforts, like OECD AI Principles, G20 AI Principles, UNESCO AI Ethics Recommendations, etc.

Thus, the first classification, based on approach, focuses on principles for governing AI and addresses the 'how' of regulation. The second one, based on mechanism and scope, focuses on implementation frameworks and regulatory responsibility of regulating AI and addresses the 'who' of regulation, and lastly, the third one, containing an integrated governance model, focuses on holistic governance based on both the approach, the mechanism, and the scope. This classification addresses 'what', 'why', 'how', and 'who', all of them together.

III

Global Regulations on AI: Approaches Towards AI Law and Regulation

The above approaches and methods, in various permutations and combinations, are used by different countries to address issues surrounding ethics, fairness, transparency, accountability, and societal impacts. The realisation of theoretical ideas of AI regulation

---

[32] Regulation (EU) 2017/745 of the European Parliament and of the Council on Medical Devices (Apr. 5, 2017) L 117/1 OFFICE JOURNAL OF THE EUROPEAN UNION (May 2017) *available at*: https://www.medical-device-regulation.eu/wp-content/uploads/2019/05/CELEX_32017R0745_EN_TXT.pdf.

[33] Financial Stability Board, THE FINANCIAL STABILITY IMPLICATIONS OF ARTIFICIAL INTELLIGENCE (14 Nov. 2024) *available at*: https://www.fsb.org/uploads/P14112024.pdf.

[34] *Supra* at 22.

for the first time is brought into the practical world by the EU. The EU AI Act is a pioneering law that sets standards for the globe to follow. It aims to manage the risk of AI, while not discouraging innovation. The AI technologies under the EU AI Act are classified based on their risk level in the regulatory framework, which incorporates the human-centric approach of the theoretical model of governance, by realising its diverse implications.[35] The Act is a meticulously drafted piece, which aligns itself with the global regulatory principles in place, for instance 'precautionary principle.'[36] It seems to have a dual nature where one aspect of it aims to regulate the innovation of the technology with minimal human intervention, while also laying down strict oversight by humans, in those applications developed and falling under the high-risk band, this creates an environment of accountability of the technology creators towards its stakeholders. While having a dualistic nature, it further goes on to provide for exemption from these provisions for military and research, indicating double standards, and the threat of a 'black-box'[37] and lack of transparency.[38]

An interesting requirement that the law draws is to label the deepfake content, hence, ensuring transparency in the generative models of AI, since developing of trust of stakeholders in the systems is important, however, it somehow lacks clarity on the enforcement, like how effective could regulations be without that much of oversight is a question.[39] The Act promotes innovation, but its strict measures somehow surpass that goal. AI systems require huge amounts of datasets, but the requirement of the law to keep data to a minimum might come in the way of developing a to-the-mark system. This has been a topic of continuous discussion. However, this idea can be suggestively used in India to cancel out irrelevant information fed to the systems. Experts support

---

[35] Benjamin Herd, *et al.*, *The European Artificial Intelligence Act: Overview and Recommendations for Compliance*, FRAUNHOFER INSTITUTE OF COGNITIVE SYSTEMS (May 2024) *available at*: https://publica.fraunhofer.de/entities/publication/f87f2c06-4abc-4b6c-987b-3e3d5413a923.

[36] Precautionary principle – In a legal context this principle lays down preventive actions to avoid uncertain potential harm that might occur

[37] Black Box – In context of AI, these refer to ethical concerns arising from AI systems, due to its opaque nature, thereby making it difficult to understand, and hold it accountable for produced results

[38] Dmitryi Kuteynikov and Osman Izhaev, *Analysing Risk-Based Approach in the Draft EU Artificial Intelligence Act* 4 LEGAL ISSUES IN THE DIGITAL AGE 97 (2023) *available at*: https://lida.hse.ru/article/view/18252; Asress Adimi Gikay, *Risks, Innovation, and Adaptability in the UK's Incrementalism versus the European Union's Comprehensive Artificial Intelligence Regulation* 32 INT'L J. LAW AND INFORMATION TECHNOLOGY (2024) htps://doi.org/10.1093/ijlit/eaae013.

[39] Philipp Hacker, *AI Regulation in Europe: From the AI Act to Future Regulatory Challenges* in OXFORD HANDBOOK OF ALGORITHMIC GOVERNANCE AND THE LAW (Ifeoma Ajunwa & Jeremias Adams-Prassl (eds.) OUP 2024) *available at*: https://arxiv.org/pdf/2310.04072.

the enactment of such laws that allow for responsible and risk-minimising AI.[40] The EU AI Act is a blanket legislation the in-depth questions are yet to be addressed. For instance, AI systems under the Act are not legally recognised personalities, so when the question of accountability arises, such as in the case of AI malfunction, it is unclear who should be held responsible, i.e., the creator, the user, or the AI itself. Up until the AI performs functions that fall under narrow AI and general AI, this framework will work, but when super AI systems are invented, the question will require a circle back. The Act has attempted to set an influential precedent, yet its blanket applicability provisions will require considerable thought from the regulators and stakeholders. Its success will be determined by its adaptability to the changes.[41]

Despite its step towards becoming a global leader in AI governance, the regulation cannot be adopted in India, essentially because of India's socio-economic context. The applications in India are focused on addressing local issues of administration, healthcare, education, recruitment, law enforcement, and finance. The requirements of the EU AI Act come with high-cost compliance.[42] This can be quite off-putting for startups and small and medium businesses that are the backbones for innovation in most of India's industries. The data minimalism model of law does not go with the principles of a populous, diversified nation. India's federal structure, with a variety of issues of diverse kinds, cannot be uniformly regulated with the same strict principles.[43] Nonetheless, India can adapt and build upon the principles given by the EU. The present governance of AI in India is fragmented, relying on the already existing laws, which are incapable of addressing this complexity. The *Information Technology Act, 2000* (IT Act) is a primary legislation for cybersecurity and electronic commerce, but it does not address issues that might arise in AI. It is vehemently argued that this inability reduces user trust in AI.[44] There are other laws, like the Consumer Protection Act, 2019, which could extend its applicability to AI consumers; however, how' of it is yet to be

---

[40] Johanna Chamberlain, *The Risk-Based Approach of the European Union's Proposed Artificial Intelligence Regulation: Some Comments from a Tort Law Perspective* 14 EUROPEAN JOURNAL OF RISK REGULATION 1 (2023).

[41] Tobias Mahler, *Between Risk Management and Proportionality: The Risk-Based Approach in the EU's Artificial Intelligence Act Proposal*, NORDIC YEARBOOK OF LAW AND INFORMATICS 245 (2021) *available at*: https://ssrn.com/abstract=4001444.

[42] European Commission, *European Artificial Intelligence Act Comes into Force* (Aug. 01, 2024) *available at*: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4123.

[43] Satish Chandra, *Regulating AI Is a Challenge*, THE HINDU BUSINESS LINE (Aug. 28, 2024) *available at*: https://www.thehindubusinessline.com/opinion/regulating-ai-is-a-challenge/article68577700.ece.

[44] Nithesh Naik, *et al.*, *Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility?* 9 FRONTIERS IN SURGERY 862322 (2022) *available at*: https://www.frontiersin.org/journals/surgery/articles/10.3389/fsurg.2022.862322/full.

determined. Further, since this Act is aimed at tackling consumer disputes,[45] it will not apply to the companies developing AI systems during their development stage, compliance and will come into play only when consumers find the systems to malfunction. The Digital Personal Data Protection Act, 2023, aims to regulate and secure individual privacy in a digital setting, but it fails to mandate provisions against biased decision-making AI systems.[46] To date, India has established NITI Aayog's Guidelines and an AI Standardisation Committee, which is aimed solely at AI and its regulation. The NITI Aayog is the forefront think tank of the government; nonetheless, it lacks an enforcement mechanism. NITI Aayog does not have defined inclusivity metrics to determine the nature of AI, which limits its ability to ensure a transparent and accountable AI system. On the other hand, the AI Standardisation Committee operating under the Bureau of Indian Standards is focused on creating technical standards for AI systems. It attempts to levy such standards which are India-specific while also aligning with international standards such as ISO and IEC. It, however, focuses on technical standards and neglects bias mitigation, creating such standards which align with a reactive approach while disregarding the already existing condition of the country, and a lack of mechanisms for public accountability and transparency.

Thus, the European Union stands out with the EU AI Act, which explicitly mandates bias testing for high-risk AI applications and bans harmful practices like social scoring.[47] The EU ensures transparent compliance through audits, penalties, and certifications. It follows a command-and-control model, combining hybrid, risk-based, prescriptive, and human-centric approaches. India, meanwhile, through documents like NITI Aayog's guidelines,[48] AI Standardisation Committee, [49] and the Digital Personal Data Protection Act, 2023,[50] promotes inclusivity but lacks strong enforcement measures. Its model is characterised by soft regulations and a principle-based approach. Further, across

---

[45] Sanjay Pareek, *et al.*, *An Empirical Study on the Factors Influencing the Use of Artificial Intelligence in Indian Financial Services* 29(S1) ACADEMY OF MARKETING STUDIES JOURNAL 1 (2025) *available at*: https://www.abacademies.org/articles/an-empirical-study-on-the-factors-influencing-the-use-of-artificial-intelligence-in-indian-financial-services-17215.html.

[46] Paarth Naithani, *Regulating Artificial Intelligence under Data Protection Law: Challenges and Solutions for India* 14(2) INDIAN J. L. & JUSTICE 436 (2023) *available at*: https://ir.nbu.ac.in/server/api/core/bitstreams/de9a9562-409f-444d-8977-feb8ceecc422/content.

[47] *Key Issue 3: Risk-Based Approach - EU AI Act*, *available at*: https://www.euaiact.com/key-issue/3 (last visited Dec. 17, 2024).

[48] Anna Roy, *National Strategy for Artificial Intelligence*, NITI AAYOG (2018) *available at*: https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf.

[49] Ministry of Electronics and Information Technology, Report of Committee: Artificial Intelligence Committees Reports (2019) *available at*: https://www.meity.gov.in/artificial-intelligence-committees-reports (last visited Dec. 17, 2024).

[50] Digital Personal Data Protection Act, 2023.

jurisdictions, countries have employed varied approaches to mitigate algorithmic bias, with significant differences in their legal and regulatory frameworks. Australia does not have specific guidelines for algorithmic bias, but it promotes fairness through its overarching ethical frameworks. Relevant legislations include the Privacy Act, 1988,[51] the Competition and Consumer Act, 2010,[52] and the Australian Human Rights Commission Act, 1986.[53] It follows a co-regulation mechanism with a principle-based and ethics-based approach. Similarly, Brazil adopts a co-regulatory and risk-based approach, where the draft AI Bill[54] classifies AI applications based on rights and risks, indirectly addressing bias. The country aligns its human rights commitments with international frameworks. Canada implements a directive-based model wherein the Directive on Automated Decision-Making[55] aims to reduce bias in public-sector AI systems, although regulation in the private sector remains weak. Canada employs an incentive-based mechanism with a principle-based approach, stressing transparency and accountability. China, while not directly targeting algorithmic bias, embeds risk control and accountability within a broader ideological framework. The Personal Information Protection Law[56] is central, and the country adopts a command-and-control mechanism, with a prescriptive and prohibition-based approach. In Japan, AI Utilisation Guidelines stress fairness and human dignity, and though indirect, they help mitigate bias. Alignment with OECD standards indicates an ethics-based, self-regulatory framework. Singapore takes a pragmatic route with its Model AI Governance Framework, which includes fairness assessment tools and promotes voluntary compliance. It uses an incentive-based principle and risk-based approach supported by public-private partnerships. The United Kingdom emphasises fairness in its National AI Strategy, though it lacks a centralised system for bias mitigation. Under the Data Protection Act, 2018, the country follows a decentralised regulation mechanism, with a market-driven and human-centric approach. Lastly, the United States maintains sector-specific guidelines that vary in enforcement. Agencies like the

---

[51] Office of the Australian Information Commission, *The Privacy Act*, *available at*: https://www.oaic.gov.au/privacy/privacy-legislation/the-privacy-act (last visited Dec. 17, 2024).

[52] Australian Competition and Consumer Commission, *Legislation we enforce*, *available at*: https://www.accc.gov.au/about-us/accc-role-and-structure/legislation-we-enforce (last visited Dec. 17, 2024).

[53] Australian Human Rights Commission Act, 1986.

[54] Rafael A.F. Zanatta & Mariana Rielli, *The Artificial Intelligence Legislation in Brazil: Technical Analysis of the Text to Be Voted on in the Federal Senate Plenary*, DATA PRIVACY BR RESEARCH (Dec. 10, 2024) *available at*: https://www.dataprivacybr.org/en/the-artificial-intelligence-legislation-in-brazil-technical-analysis-of-the-text-to-be-voted-on-in-the-federal-senate-plenary/.

[55] *Supra* at 27.

[56] Personal Information Protection Law, *available at*: https://personalinformationprotectionlaw.com/ (last visited Dec. 18, 2024).

FTC and NIST support fairness testing. The U.S. adopts a decentralised regulatory model, following a market-driven and principle-based approach.

With this,[57] it can be concluded that most countries are on their way to enacting AI legislation. The next part explores the possibilities of adopting or adapting certain features of transparency and accountability from the EU and other jurisdictions to eliminate the manufacturing of biased AI systems in India.

IV

India-Centric Approach to Overcome Bias: Regulatory Framework for AI

In India, varied inequalities have shaped minds to think a certain way, and various laws have been enacted to identify and curb the biases arising out of social stratification in society. While the existing laws deal with the existing bias and inequalities, the ones which arise systematically through AI systems and technologies are not yet addressed and governed under any binding legislation. In Maharashtra, a chatbot known as '*Aaple Sarkar Bot*', which is an AI-powered chatbot, was launched in 2019.[58] It aimed to provide users with access to information about public services managed by the state government as a part of the Right to Service Act, 2015, helping the users with queries relating to healthcare, education, development, tracking of water connection, driving licenses, etc. As visionary and convenient as the usage of this chatbot sounds, it is an example of our inherent thinking leading to linguistic non-inclusivity, as the system from its initial development and deployment was not trained with data from the official language[59] of the state in which it was supposed to be used thus infringing upon the right to the accessibility of the citizens under the *Right to Information Act, 2005* (RTI Act).[60]

---

[57] Amlan Mohanty & Shatakratu Sahu, *India's Advance on AI Regulation*, CARNEGIE ENDOWMENT FOR INTERNATIONAL PEACE (Nov. 21, 2024) *available at*: https://carnegieendowment.org/research/2024/11/indias-advance-on-ai-regulation?lang=en.

[58] DC Correspondent, *AI Chatbot Provides Information on 1,400 Public Services in Maharashtra*, DECCAN HERALD (Mar. 08, 2019) *available at*: https://www.deccanchronicle.com/technology/in-other-news/080319/this-ai-chatbot-provides-information-on-1400-public-services-in-mahar.html.

[59] Maharashtra Official Languages Act, 1964, S. 2 declares Marathi as the official language of Maharashtra, mandating its use in state-level communications and public service delivery

[60] Right to Information Act, 2005, S. 4(4) requires the public authorities to provide information in the local language to ensure accessibility for all citizens

Further, the lack of practical applicability of the updated Information Technology Rules, 2021, failed to prevent the 2024 election from witnessing digital deception through deepfakes and AI-generated content, and videos featuring deceased political figures being circulated without proper labels, leading to voter manipulation.[61] Circulating deceptive AI-generated content to manipulate voter perception is a form of undue influence,[62] as it interferes with voter's ability to make informed choices. Misinformation during election campaigns is violative of Section 123 (4) of the Representation of the People Act, 1951 as it prohibits the publication of false statements about a candidate's character or conduct to influence voting. Fake videos of deceased political figures could amount to false statements intended to mislead voters and tarnish reputations, constituting a corrupt practice. Further, in Delhi, an unregulated facial recognition technology was used to identify and arrest suspects, leading to the amplification of systemic biases and criminalising minority communities[63] since the data used was faulty. India's socio-economic diversity challenges conventional assumptions of algorithmic fairness in governance systems.[64] Peter and Carman[65] conducted research that significantly found AI research to be culturally biased, favouring the Western population. Another research study compared ethical considerations in India, the UK, and the USA, and it obtained different perspectives to address challenges that are *sui generis* to India.[66] Research[67] conducted on large language models that are used in the

---

[61] Sahana Venugopal and Saumya Kalia, *From IT Bots to AI Deepfakes: The Evolution of Election-Related Misinformation in India*, THE HINDU (May 24, 2024) *available at*: https://www.thehindu.com/elections/lok-sabha/from-it-bots-to-ai-deepfakes-the-evolution-of-election-related-misinformation-in-india/article68015342.ece.

[62] Representation of the People Act, 1951, § 123 (2) defines the term 'Under Influence' as any direct or indirect interference with the free exercise of electoral rights

[63] 'Explained | Delhi Police's Use of Facial Recognition Technology' *The Hindu* (21 August 2022) https://www.thehindu.com/sci-tech/technology/explained-delhi-polices-use-of-facial-recognition-technology/article65793897.ece#:~:text=facial%20recognition%20technology%3F-,The%20Delhi%20Police%20first%20obtained%20FRT%20for%20the%20purpose%20of,Haldar%20vs%20NCT%20of%20Delhi accessed 19 December 2024

[64] Nithya Sambasivan, *et al*., Re-Imagining Algorithmic Fairness in India and Beyond, PROCEEDINGS OF THE 2021 CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (2021) *available at*: https://arxiv.org/pdf/2101.09995.

[65] Uwe Peters and Mary Carman, *Cultural Bias in Explainable AI Research: A Systematic Analysis* 79 JOURNAL OF ARTIFICIAL INTELLIGENCE RESEARCH 971 (2024) *available at*: https://jair.org/index.php/jair/article/view/14888.

[66] Amina Catherine Ijiga, *et al*., *Ethical Considerations in Implementing Generative AI for Healthcare Supply Chain Optimization: A Cross-Country Analysis across India, the United Kingdom, and the United States of America* 7 IJBPAS, Archive 048 (2024) https://doi.org/10.53771/ijbpsa.2024.7.1.0015.

[67] Siddharth Garg, *Women May Pay a "MOM PENALTY" When AI Is Used in Hiring, New Research Suggests* NYU (Dec. 12, 2023) *available at*:

Contd…

hiring process concluded bias in resume segregation based on its content and negatively affecting those women who have maternity-related gaps in their resumes.[68] Algorithmic bias leads to the marginalisation of underprivileged populations.[69] Several AI systems developed during the COVID-19 period showed social and ethical implications of systemic bias.[70] AI's potential mistakes are due to the unbalanced datasets in them. [71] The learning systems used in education are mostly American-centric learning systems[72] , which would lead to algorithmic bias if used in India, thereby advocating for such systems to support and train with data from the education system, as follows in India. The AI tools developed for use in the Indian education system have traces of biased datasets.[73]

Below is a *Hybrid Integrated Governance Model* (HIG Model) for Promoting AI Inclusivity for Mitigating Bias at the Grassroots in India which is divided into two types (i) Sector-Specific Amendments; for illustration few sectors are taken like healthcare, education, recruitment, law enforcement, and finance and a legal framework for them is proposed, and (ii) Sector Agnostic AI Legislation.

Sector-Specific Models

Firstly, in Healthcare, for sector-specific amendments with grassroots-level provisions in order to enable inclusivity, the laws must be changed to address the issue. The legislatures should ensure that diversified datasets are included in the training of algorithms in this sector, and the predictive solutions for diseases such as diabetes or hypertension can include samples from different regional, linguistic groups, and socio-economic classes. Amendment to the IT Act[74] should be made to represent demographic diversity in data collection, enabling algorithms to work on behalf of all sections of society. Local data inclusion will play an important role in addressing regional

---

https://engineering.nyu.edu/news/women-may-pay-mom-penalty-when-ai-used-hiring-new-research-suggests.

[68] Akshaj Kumar Veldanda, *et al.*, *Investigating Hiring Bias in Large Language Models*, *available at*: https://openreview.net/pdf?id=erl90pLIH0 (last visited Dec. 19, 2024).

[69] Sarthak Bhatia, Anuj Kumar and Stuti Tandon, *Uncovering the Challenges from Algorithmic Bias Affecting the Marginalized Patient Groups in Healthcare*, INTERNATIONAL CONFERENCE ON INNOVATIVE COMPUTING AND COMMUNICATION (2024) *available at:* https://ssrn.com/abstract=4848690.

[70] Janet Delgado and others, *Bias in Algorithms of AI Systems Developed for COVID-19: A Scoping Review* 19 JOURNAL OF BIOETHICAL INQUIRY 407 (2022).

[71] Atharva Prakash Parate, *et al.*, *Review of Data Bias in Healthcare Applications* 20 INTERNATIONAL JOURNAL OF ONLINE AND BIOMEDICAL ENGINEERING 124 (2024).

[72] Ryan S Baker and Aaron Hawn, *Algorithmic Bias in Education* 32 INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE IN EDUCATION 1052 (2022).

[73] Gurumurthy Kasinathan and Yogesh K S, *Exploring AI in Indian School Education* (2019) *available at*: https://itforchange.net (last visited Dec. 20, 2024).

[74] Information Technology Act, 2000, S. 43A.

healthcare disparities. For instance, the system diagnosing diseases[75] like kala-azar or dengue should be trained on affected region-specific data, and integration of the data could be through accounts like Ayushman Bharat Health Account,[76] which will be a cost-effective solution for genuine data collection. Grassroots-level applications, including feedback from disadvantaged communities,[77] should guide the design of patient-centric algorithms to meet the diverse needs of India's population. Further, to promote transparency in the process, mandatory statutory periodic audits of healthcare algorithms should be there to report their performance across India's diverse population. The healthcare fraternity can be made to attend timely workshops to identify errors in AI diagnostics. Cost-effective online modules in regional languages[78] should ensure that healthcare professionals all over the country are well-trained. To increase transparency and public trust, the RTI Act[79] should mandate public disclosures of audit results for AI algorithms used in healthcare. Such measures should foster accountability and ensure that only bias-tested algorithms are employed in both public and private healthcare systems. In order to ensure accountability, a strict liability approach is needed to hold developers and deployers of biased algorithms to hold them accountable for harm caused to the patients. Healthcare-specific amendments modelled after the Consumer Protection Act, 2019,[80] can help patients seek redress for adverse outcomes caused by flawed AI tools. To include negligence in deploying flawed algorithms, the *Bharatiya Nyaya Sanhita*, 2023, can be amended. To make algorithms patient-centric and support grassroots intervention public-private partnership by the Ministry of Health and Family Welfare[81] would be an effective step. For instance, to cater specifically to the needs of disadvantaged populations, AI-based maternal health tools should be piloted in rural districts under Janani Suraksha Yojana schemes.[82] NITI Aayog[83] can further enhance oversight by developing regulatory bodies that facilitate third-party certification of AI algorithms, auditing, and grievance redressal mechanisms. The enforcement mechanisms will need to levy fines or even disqualify

---

75 Epidemic Diseases Act, 1897, S. 2A.

76 Ministry of Health and Family Welfare, NATIONAL HEALTH POLICY, 2017 *available at*: https://mohfw.gov.in/sites/default/files/9147562941489753121.pdf.

77 Constitution of India, Art. 46 and 243G.

78 *Id*., Art. 343.

79 *Supra* at 60, S. 4 & S. 6

80 Consumer Protection Act, 2019, S. 2(42).

81 National Health Mission, *National Health Mission Guidelines*, *available at*: https://nhm.gov.in/index1.php?lang=1&level=1&sublinkid=197&lid=136 (last visited Dec. 19, 2024).

82 Ministry of Health and Family Welfare, *Janani Suraksha Yojana Guidelines*, *available at*: https://nhm.gov.in/WriteReadData/l892s/97827133331523438951.pdf.

83 *Supra* at 48.

the use of biased algorithms[84] by healthcare organisations, and thus comply with ethical AI standards.

Secondly, in education, inclusivity can be ensured in AI should be made fair through a bias mitigation certification framework. Admission tools adopted by universities must be certified as algorithms to account for diversity.[85] Cost-effective certification under the Start-Up India Scheme[86] must encourage small businesses and start-ups to innovate responsibly. Diversified data standards must be established. Thus, the grading systems applied by the central boards[87] should include the datasets portraying the rural and under-represented communities, rather than urban bias. Open-source AI tools translating textbooks[88] into tribal languages like Gondi or Santali[89] can help overcome the language barrier, and in remote areas, data-sharing networks would enable smaller universities to access higher-quality datasets. Feedback from the stakeholders, such as rural teachers, should inform the development and deployment of AI tools,[90] and students should be able to lodge appeals through digital engagement portals against grading decisions to enhance inclusiveness and accountability. Further, policymakers should make AI-driven educational systems more transparent by mandating public disclosure of the algorithmic processes used in decisions, such as college admissions. Institutions should publish simplified reports explaining how factors such as academic performance and reservation quotas are weighed, supported by amendments to the RTI Act. AI education tools[91] must have ethics boards with representation, and ensure that they reach both the rural and the urban equitably. Routine audits are necessary to see whether an AI system does not disadvantage vernacular[92] over English medium students for purposes like grading. The findings of these audits should be made public through non-technical summaries so that the stakeholders can trust them and understand and address the biases in AI-driven education tools. Accountability should also be guaranteed in an AI-driven education system through performance-based incentives, independent monitoring, and community engagement. Provisions can be placed for rewards in the form of scholarships or public awards to those schools and

---

84 *Supra* at 74, S. 72 A.

85 University Grants Commission Act, 1956, S. 12A.

86 StartupIndia, *Startup India Action Plan, 2016*, *available at*: https://www.startupindia.gov.in/content/dam/invest-india/Templates/public/Action__Plan.pdf.

87 Central Board of Secondary Education Bye-Laws, 1988.

88 Ministry of Human Resource Development, NATIONAL EDUCATION POLICY, 2020 *available at*: https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf.

89 *Supra* at 77, Art. 350A.

90 *Supra* at 77, Art. 243G.

91 *Supra* at 88, Ch. 23.

92 *Supra* at 77, Art. 14 & Art. 21.

institutions of learning that employ ethical AI.[93] Grievances and suggestions by review and complaint committees under UGC[94] or AICTE[95] regarding biased assessment scores or unfair admission policy consultation with teachers, as well as parents, can help highlight and rectify cultural disparities. Amendment[96] support can ensure fairness, impartiality, and compliance with wide-ranging stakeholder expectations effectively.

Thirdly, in employment and recruitment, inclusivity reforms can be achieved by a tiered certification framework for technology-driven systems. Exams like UPSC, related to public health or recruitment, must be evaluated at a more stringent scale based on dataset diversity, bias mitigation, and real-world impact[97] to ensure equal opportunity in public employment. Scalable evaluations of the systems would be possible only by classifying them as high, medium, and low-impact. For example, AI tools for AIIMS recruitment would achieve stringent certification standards;[98] in contrast, initial resume filtering tools for such startups would conform to streamlined metrics. Businesses that score high on responsible innovation practice should be rewarded under incentive-based grading systems. Businesses receiving top ratings in an Index of Fair Employment can promote themselves as inclusive, like Zomato, with regard to regional diversity. All recruitment systems using AI must also acquire bias mitigation certifications, such as those used on Naukri.com and LinkedIn, which, before deployment, are made to run through simulations for extended periods in controlled conditions, a 'Bias Elimination Sandbox'.[99] Shared AI infrastructure by the government should democratize access to advanced technologies so that smaller organisations can compete ethically. For instance, open-source AI tools that translate job postings into multiple regional languages[100] should bridge the gap for non-English-speaking candidates. Public-private partnerships should fund AI-based skill-matching tools at state employment exchanges in regions that can enhance job accessibility.[101] Further, transparency practices in recruitment technologies should encourage accountability and trust. Laws such as the Companies Act, 2013,[102] should be amended to require disclosure of processes using AI and fairness measures at periodic intervals. Public Sector Undertakings should publish metrics of reservation policies, gender representation, and socio-economic diversity;[103] hiring systems should align with constitutional mandates. There should be a centralised

---

[93] *Samagra Shiksha Scheme Guidelines*, Ministry of Education, Government of India 2018 *available at*: https://samagra.education.gov.in/docs/samagra_shiksha.pdf.

[94] UGC (Grievance Redressal) Regulations, 2018, Reg. 4.

[95] All India Council for Technical Education Act, 1987, S. 10.

[96] *Supra* at 91.

[97] *Supra* at 77, Art. 16.

[98] All India Institute of Medical Sciences Act, 1956.

[99] *Supra* at 74.

[100] *Supra* at 77, Art. 343.

[101] *Supra* at 77, Art. 41.

[102] Companies Act, 2013, S. 134(3)(m).

[103] *Supra* at 77, Art. 16(4).

Digital Employment Portal accessible in regional languages, allowing job applicants to simulate resume evaluations with step-by-step explanations of the hiring criteria. It should be designed to include grievance redressal mechanisms[104] in the process that will make the candidates report their views and grievances concerning the seeming bias in improving fairness for various demographics. Periodic audits of recruitment systems should ensure compliance with inclusivity benchmarks. A National Algorithm Audit Bureau should standardise audit practices, supported by legislative amendments to the IT Act.[105] Rolling audits must have the capability to be used incrementally in the datasets and algorithms without halting operations. Public summaries must highlight issues such as a scoring mechanism that penalises employment gaps because of a caregiving responsibility, to correct the issue.[106] Further, accountability in recruitment technologies must be strengthened with sector-specific ethics committees and mandatory audits. These committees should review AI systems to ensure compliance with ethical norms and diversity standards, such that tools used by firms like Infosys or TCS take into account socio-economic inclusion without compromising merit-based selection.[107] The Labour Codes[108] should also establish such committees with members drawn from labour unions, NGOs, and community leaders to provide varied outlooks for the evaluation. Smaller organisations will require that the government develop data-sharing networks under the National Skill Development Mission,[109] whereby rural employment agencies have easy access to large-scale datasets without necessarily spending exorbitant amounts. Public-private partnerships[110] are necessary in setting up AI Incubation Hubs in universities to ensure innovation and affordability in recruitment tools. Ethics board membership should rotate regularly to bring fresh insights, and annual reports capturing the overall evaluations should display biases corrected in the recruitment systems, for example, the urbanised over the ruralised candidates.[111]

Fourthly, in Law Enforcement and Surveillance, inclusivity amendments to the Indian Police Act, 1861,[112] can mandate the constitution of Community Oversight Committees in every district to ensure local representation to review surveillance policies and specific concerns. Caste-based violence investigations should be monitored by independent panels.[113] Community Monitoring Centres must be established in urban

---

[104] *Supra* at 80.
[105] *Supra* at 74, S. 72A.
[106] Maternity Benefit Act, 1961.
[107] *Supra* at 77, Art. 15(4).
[108] Code on Wages, 2019, S. 3.
[109] Ministry of Skill Development and Entrepreneurship, NATIONAL POLICY ON SKILL DEVELOPMENT AND ENTREPRENEURSHIP, 2015 *available at*: https://www.nitiforstates.gov.in/policy-viewer?id=PNC1160P000010.
[110] *Supra* at 86.
[111] *Supra* at 77, Art. 46.
[112] Indian Police Act, 1861, S. 23.
[113] Scheduled Castes and Schedules Tribes (Prevention of Atrocities) Act, 1989, S. 4.

slums and rural areas[114] residents should be allowed to inspect surveillance practices and make them transparent. Digital literacy campaigns must be initiated[115] so that citizens can be empowered with their rights towards privacy and redressal of grievances. Participatory budgeting[116] should provide communities with a say in where the surveillance money is spent. This means voting on priorities to ensure that spending is aligned with the local needs, of more CCTV cameras or better police response infrastructure. To ensure transparency, public scrutiny and proper guidelines for surveillance technologies, along with the amendments to the RTI Act,[117] must make it a rule for law enforcement agencies to submit yearly transparency reports stating the usage of surveillance tools, budgetary provisions, and the safeguards for privacy. The National Automated Facial Recognition System[118] has to be applied with cautious pre-deployment impact assessments, and privacy and civil liberties[119] should be safeguarded. The application of drones in surveillance should be governed by warrant-based requirements;[120] therefore, judicial sanction has to be provided prior to their application in sensitive tasks. Data-sharing agreements should satisfy transparent legal processes involving private parties and law enforcement agencies. An amendment to the Digital Personal Data Protection Act, 2023,[121] can be used to prohibit companies from sharing user data without an explicit requirement. A digital employment portal in regional languages will assist citizens in knowing how surveillance technologies impact them and provide grievance redressal mechanisms. [122] To ensure accountability of law and order, enforcement authorities must be made more responsible by independent oversight bodies, and articulated policies and procedures for the use of data. There should be a mechanism to ensure allegations of excessive force, abuse of surveillance tools, or violation of privacy are independently investigated.[123] IT rules[124] amendment must include immediate suspension and legal actions against officers who access personal information without authorisation. In order to curb misuse, the data retention policy must limit how long surveillance data is stored. Public-private partnerships supporting AI Incubation Centres[125] can provide affordable and innovative surveillance technologies for law enforcement. Pre-deployment impact assessments can ensure the

---

[114] *Smart Cities Mission Guidelines, 2015*, Ministry of Urban Development, Government of India (June 2015) *available at*: https://smartcities.gov.in/themes/habikon/files/SmartCityGuidelines.pdf.

[115] Digital India Initiative, 2015.

[116] *Supra* at 77, Art. 243G.

[117] *Supra* at 79.

[118] *Supra* at 74.

[119] *K. S. Puttaswamy and Anr.* v. *Union of India and Ors*, (2017) 10 SCC 1.

[120] Drone Rules, 2021, Rule 21.

[121] *Supra* at 50, Ch. IV.

[122] *Supra* at 80.

[123] *Supra* at 114.

[124] *Supra* at 74, S. 72A.

[125] *Supra* at 91.

evaluation of all systems for compliance with privacy protections before implementation.[126] Drone surveillance should require judicial approval under the updated Unmanned Aircraft Systems Rules, 2021,[127] especially for monitoring protests or large public gatherings. To ensure transparency, data-sharing agreements should require explicit user consent or judicial authorisation[128] to safeguard citizen's rights while maintaining security standards.

Fifthly, in Financial Services, the Banking Regulation Act, 1949,[129] could make it mandatory for banks and financial institutions to maintain minimum levels of financial inclusion, especially in terms of defined allocations for women, small-scale entrepreneurs, and self-help groups.[130] Amendments to the Reserve Bank of India Act, 1934,[131] could ease KYC procedures by allowing acceptance of alternative identification methods, such as Aadhaar-linked mobile numbers,[132] locally issued IDs, or certifications from Panchayats. Enhanced Priority Sector Lending Guidelines could motivate banks to open rural branches and provide focused banking services to low-income groups. Open financial architectures under the Digital India Mission[133] could eliminate duplication and reduce the costs of deploying financial products. For example, a tool like Unified Payments Interface[134] can be localised for regional languages and further simplified for end-user access. Shared financial infrastructure through data-sharing networks under the National E-Governance Plan[135] could facilitate cooperation among financial institutions. Public-private partnerships for cost-sharing models can be explored for pooling resources for such large initiatives. A certification framework before the deployment of financial algorithms[136] can prevent discrimination. Community engagement programs, in collaboration with NGOs and Panchayats,[137] can educate citizens on financial services. Digital feedback portals under the Digital India Programme can help underprivileged sections report grievances or monitor the effectiveness of local schemes. District-level ethics committees could be in place to provide support for efforts toward inclusivity, with annual reporting on progress.

---

[126] *Supra* at 74.
[127] *Supra* at 80.
[128] *Supra* at 114.
[129] Banking Regulation Act, 1949, S. 35A.
[130] Microfinance Institutions (Development and Regulation) Bill, 2022, S. 6.
[131] Reserve Bank of India Act, 1934, S. 45JA.
[132] Aadhar Act, 2016, § 7
[133] Digital India, *Digital India Programme, 2015*, *available at*: https://www.digitalindia.gov.in/about-us/ (last visited 24 December 2024).
[134] Payment and Settlement Systems Act, 2007, S. 10A.
[135] Ministry of Electronics and Information Technology, *NeGP Guidelines, 2006*, *available at*: https://www.meity.gov.in/static/uploads/2024/02/Guidlines_Operational_Model_V42_231210.pdf.
[136] *Supra* at 74.
[137] *Supra* at 77, Art. 243G.

Incentives such as tax breaks or grants under the Income Tax Act, 1961,[138] could be offered to institutions that excel in inclusivity to encourage the adoption of inclusive practices. Further, transparency can be promoted by the publication of annual disclosure reports[139] that provide user-friendly formats of demographic loan data, nonperforming assets, and service impacts. Rural cooperative banks and microfinance institutions can also go fully digital with transactions for clear record-keeping and better oversight by regulatory authorities.[140] Clear and effective guidelines under the Microfinance Institutions (Development and Regulation) Bill[141] can help prevent predatory practices and raise borrower awareness by streamlining loan interest rate disclosures. Transparency centres could be established at the village level that show and explain simple financial reports to the local people.[142] Public summaries of independent audit findings could be circulated in rural areas so that the community is aware of whether the financial services are fair and effective. To increase accountability, the grievances in respect of discrepancies in service, algorithmic mistakes, or misreporting are to be dealt with through district-level financial grievance redress cells that would give expeditious redressal.[143] The penal provisions can be brought into the Reserve Bank of India Act, 1934, towards the malpractitioners and misreporters. Data deletion policies under *Bharatiya Sakshya Adhiniyam*, 2023, would be effective to ensure that all sensitive financial information would be deleted within a time period unless it is actively needed for the process of investigation. Participatory budgeting mechanisms would further empower the local communities to monitor and guide financial allocations and ensure accountability according to community priorities. Bias audits of algorithmic decisions would ensure that low-income or rural applicants are not prejudiced against. Grading the financial institutions for accountability can also be facilitated through tax incentives offered to top performers under the Income Tax Act, 1961.[144]

Sector Agnostic Model

This should reflect upon the spirit of the Competition Act, 2002, to set objectives of its own. Similar to the Competition Commission of India, the *AI Regulatory Authority of India* (AIRAI) should operate dualistically as an investigation authority and as a quasi-judicial authority. The AIRAI must function as the SEBI in the system as a watchdog to make AI ethical and fair.[145] Further *Regional AI Compliance Boards* (RACB) should be patterned after State Pollution Control Boards.[146] RACBs need to provide local-level

[138] Income Tax Act, 1961, S. 80G.
[139] *Supra* at 102, S. 134.
[140] *Supra* at 134, S. 25.
[141] *Supra* at 130, S. 13.
[142] *Supra* at 77, Art. 243G.
[143] *Supra* at 80, S. 17.
[144] *Supra* at 138, S. 35 AC.
[145] Securities and Exchange Board of India Act 1992, S. 11.
[146] Water (Prevention and Control of Pollution) Act, 1974, S. 17.

regulatory oversight and take care of region-specific AI issues. For example, a RACB for Punjab shall regulate AI applications in the agricultural sector, and the crop disease prediction system tool needs to be customised to regional requirements. *AI Dispute Resolution Tribunal* (AIDRT) should be modelled on the National Company Law Tribunal[147] as a specialised forum for AI-related disputes. For instance, the AIDRT can hear complaints against ride-sharing algorithms accused of surge pricing based on socioeconomic status. Drawing from the Consumer Protection Act, 2019, establishing user-friendly grievance systems. Similar to the Industrial Disputes Act, 1947, the Act should implement a tiered dispute resolution process, beginning with internal mechanisms and escalating to RACBs or the AIDRT. Out-of-the-court settlement, aligning with the ADR laws of India, will be a great initiative for speedy settlement. The penalties can be based on the severity of violations, from moderate to severe.

Firstly, fairness principles can be drawn from provisions similar to the Equal Remuneration Act, 1976,[148] wherein the Act can mandate AI systems to provide unbiased decisions in any area of usage, aligning with the principles of equality. This can be achieved by means such as fair training data and periodic auditing to discover and remove biases, as well as providing transparency through clear accountability structures. Remedial access through the courts must also exist to challenge unfair AI decision-making, as with provisions of the Act to correct imbalanced pay situations. With these principles in AI regulation, we can more effectively execute society's values of fairness and equality so that such technologies benefit all users without recreating systemic biases.

Secondly, transparency motivated by the RTI Act,[149] the AI developers can be compelled to disclose key decision-making processes, such as a credit-scoring AI should document its methodology so that people understand decisions such as loan denials. Analogous to the Census Act, 1948,[150] the provisions in this Act should require AI systems to be trained on diverse datasets. For instance, a facial recognition AI used in airports should be trained on datasets representing all skin tones to avoid racial or ethnic biases. Analogously, from the Insolvency and Bankruptcy Code, 2016,[151] the Act can compel public disclosures of non-compliance.

Thirdly, accountability modelled after the Environment Protection Act, 1986,[152] accountability can be established by holding developers and all other intermediaries involved in the AI lifecycle for harmful outcomes. For example, such healthcare systems prioritising discrimination based on the wealth of patients should be discouraged and penalised under the Act, ensuring rectification of biases. Furthermore, the audits

[147] *Supra* at 102, S. 408.
[148] Equal Remuneration Act, 1976, S. 4.
[149] *Supra* at 60, S. 4.
[150] Census Act, 1948, S. 4-5.
[151] Insolvency and Bankruptcy Code, S. 8.
[152] Environment Protection Act, 1986, S. 15.

requiring regular third-party audits of AI systems, as modelled after the Companies Act,[153] are required for developing unbiased systems.

Fourthly, inclusivity can be enabled by borrowing principles from the Scheduled Castes and Scheduled Tribes (Prevention of Atrocities) Act, 1989.[154] To ensure the representation of marginalised groups, like the training datasets for language translation, which can include tribal languages and regional dialects, to prevent digital exclusion. For purposes of further streamlining fairness in applicability, this Act should also be aligned with sector-specific laws, like healthcare, law enforcement, etc., as discussed above.

Fifthly, in order to promote AI innovation, the act should create rules and regulations. The Startup India initiative is an on-point inspiration for this. Provisions for monetary benefits to startups complying with ethical AI checkboxes. Funding mechanisms in the form of grants, venture capital support, and subsidised loans. For innovation and risks to be encouraged, sandbox zones, similar to the Special Economic Zones,[155] can be established as focused areas to test new AI technologies that are emerging. These zones would offer a regulated environment with relaxed regulations that would allow startups and researchers to test the most advanced AI solutions while adhering to safety and ethical standards. Through collaboration between the public and private sectors, these zones would represent hubs for responsible AI development. The Act should use the Skill India Mission to avail of specialised training programs in collaboration with universities and industries that focus on ethical AI practices, data privacy, and algorithmic transparency. Scholarships and outreach programs to underrepresented groups may ensure diversity and inclusivity within the AI workforce. This, at the same time, may also be adopted under the Act, similar to the National Rural Employment Guarantee Act, 2005,[156] whereby it will fund AI-driven solutions for regional problems. Moreover, the Act should institute an Ethical AI Certification Program, which would certify the AI systems and organisations to meet predefined ethical standards. This would not only increase consumer confidence but would also give competitive advantages to compliant startups. The Act could mandate the establishment of an institution of a National Ethical AI Council responsible for overseeing policy implementation, redressing grievances, and making recommendations to keep pace with techno-progress.

Thus, the above HIG Model can align AI in India with global principles of inclusivity, transparency, and accountability and contain provisions for certification, public notices/public engagement, timely audits of the data, formation of ethics committees, and incentive-based grading systems.

---

[153] *Supra* at 102, S. 143.

[154] *Supra* at 113, S. 3.

[155] Special Economic Zones Act, 2005, S. 15.

[156] National Rural Employment Guarantee Act, 2005, S. 3.

V

Conclusion

In this article, the author has called for an urgent requirement for a binding legal framework for regulating emerging AI technologies, while also promoting innovation. The starlight features, such as regulating and prohibiting existing biases, can prevent existing inequality in the real world from getting amplified and perpetuated in AI technology. The aim of creating an equality-based virtual world would become a little easier with proper legal backing. Every technology takes time to get assimilated by the consumers, and given the fact that AI has the potential to do as much as it has to do good, it is difficult for users to rely on it. Therefore, AI systems should be of such quality and integrity that they can be used and relied upon effectively, which in the long run will be profitable to the developers, the users, and India through increased investments in the Indian AI market. The abovementioned HIG Model can be a way to promote AI inclusivity and mitigate bias at its very inception stage, thereby ensuring fairness and equity in AI systems. By drawing from the available model of regulations and taking lessons from the present legal system of the nation, we can pave our way toward humanity-sensitive AI. It also means that with the help of ethics and transparency, AI will help to close accessibility gaps so that everyone gets equal opportunities.