JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

From Data to Discrimination: Addressing Gender Inequality in AI Systems

Theme of Submission: Human Rights, Technology, and the Digital Society Sub theme of Submission: AI and Automation

Name of the Author: Anushree Kumar

Name of College: Indian Law Society's Law College, Pune

Email Id: anushree.kumar123@ilslaw.in

Abstract

Artificial intelligence systems increasingly shape decisions in critical areas such as hiring, healthcare, and law enforcement. However, these systems are often perpetuated and encoded with certain biases that reinforce discriminatory practices in the digital society. It manifests in several forms, including facial recognition systems leading to wrongful arrests of black women; recruitment algorithms favoring male candidates; and voice assistants perpetuating gender stereotypes through submissive, female-coded responses, alongside biases in online ad targeting.

Left unaddressed, such biases threaten to exacerbate the existing inequalities, undermining the fundamental right to equality in an increasingly technology-driven world. Mitigating gender bias in AI is not just a technical challenge but a societal imperative and a policy decision that can be enforced through law. This paper argues that mitigating gender bias in AI requires a combined effort embedding ethical frameworks in AI design, bias audits, and regulatory oversight to ensure accountability.

While scholars have highlighted the dangers of algorithmic biases, the discourse often focuses on technical fixes without addressing the systemic societal factors that produce and amplify these biases. Additionally, a gap exists in connecting AI governance to broader human rights frameworks, which could offer a more cohesive solution.

This paper bridges that gap by applying a human rights lens to AI governance, proposing an integrated framework that prioritizes inclusivity in data collection and transparency in AI development. By situating the discussion within the broader context of technology's impact on human rights, this paper underscores the necessity of treating AI bias not merely as a technical flaw but as a moral and legal failing that demands urgent

attention. Its intellectual contribution lies in reshaping the conversation around AI from technological advancements and innovations to accountability and equality, aiming to influence the policy and practice around it.

Introduction

The increasing ubiquity of artificial intelligence (AI) in decision-making processes has critically shaped certain sectors. The use of AI systems in hiring, healthcare, law enforcement, and many more is exponentially on the rise.[1] AI systems, built on vast datasets and trained to identify patterns, are widely regarded as tools of efficiency, objectivity, and precision. However, AI models aren't free from bias and their output can be affected by the data they are trained on. This bias, particularly in relation to gender, has emerged as a challenge in the digital society. AI already has proven records of gender discrimination that have been far-reaching and stark. Bias in AI algorithms can create a disadvantage for female recruiters - claiming they have fewer or no opportunities in an industry sector than male recruiters. Many instances of misguided apprehensions followed by excessive surveillance due to AI grappling with women of color are examples of AI's failure in the judiciary sector. [2]Most AI-powered voice assistants are coded with submissive, female-sounding personas which perpetuate harmful stereotypes. These factors reveal that AI Models are susceptible to many biases.[3]

The issue of gender bias in AI cannot be dismissed as a mere technical flaw or an unintended consequence of innovation. This is rather a breach of fundamental provisions of law such as the right to equality as articulated in Article 14 of the Indian Constitution.[4] Article 15 of the Consitution[5] further prohibits discrimination based on sex. These highlight how this issue is a constitutional concern. Beyond the domestic Indian law, the international legal framework continually reinforces this principle. Article 7 of the Universal Declaration of Human Rights (UDHR)[6] states that all human beings are equal before the law and are entitled without any discrimination to equal protection of the law, and the Convention of the Elimination of All Forms of Discrimination Against Women (CEDAW)[7] obliges state parties to combat gender discrimination and stereotypes in all spheres of society.

The demonstration of individual rights in the legal context of gender bias in AI is not the only challenge. This bias has the potential effects of aggravating sociocultural inequalities, and

[1] Priyadi, Unggul & Agus Arwani, Digital Transformation: Artificial Intelligence Shaping the Future of Public Sector, 7 New Applied Studies in Management, Economics & Accounting 54 (2024).
[2] Jonny Bairstow, AI's Influence on Shaping Ethical Practices and Driving Technological Innovation Across Sectors (2024).
[3] Ramya Srinivasan & Ajay Chander, Biases in AI Systems, 64 COMM. OF THE ACM 44, 44-49 (2021).
[4] Article 14 of the Indian Constitution, Constitution of India, art. 14.
[5] Article 15 of the Indian Constitution, Constitution of India, art. 15.
[6] Article 7 of the Universal Declaration of Human Rights (UDHR), G.A. Res. 217 A (III), U.N. Doc. A/810 at 71 (1948).
[7] Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW), G.A. Res. 34/180, U.N. Doc. A/34/46 (1979).

hence, disintegrating the rule of law in a world that is becoming increasingly driven by technology.[8] Combatting this problem is therefore not only a concern for the developers seeking to enhance AI technologies, it is in the interest of society at large. This paper argues that the mitigation of gender bias in AI is a societal imperative that must be approached with the lens of legality and morality beyond just technical obstacles. It requires a strategy that is more inclusive of ethical, compliance, and international human rights frameworks. It also maintains that inclusion and active participation in all engagements with and of AI should be the global rule around bias mitigation.

Research Questions

1. What is the presence of manifestations and impacts of gender bias in AI systems within the Indian context?
2. How do current legal and regulatory frameworks in India address AI bias?
3. What lessons can India learn from the experience of other countries that created a human rights-based legal framework for AI governance?
4. What policy model can be proposed to combat gender bias in AI?

Objectives

1. To identify legal gaps and weaknesses in India's approach to AI governance.
2. To develop a rights-based, regulatory framework for AI design that focuses on transparency, accountability, and inclusivity.
3. To integrate constitutional principles and international human rights standards into AI policy.

Methodology

This research is doctrinal and comparative in nature. It relies on primary sources including constitutional provisions, domestic statutes such as the Information Technology Act, 2000, international human rights instruments, and judicial precedents. Secondary sources would include scholarly articles, policy reports, and case studies. The analysis puts Indian domestic

[8] Gichoya, Judy Wawira, et al. "AI Pitfalls and What Not to Do: Mitigating Bias in AI." The British Journal of Radiology 96.1150 (2023): 20230023.

legal frameworks side-by-side with international best practices to propose a hybrid model for India.

Literature Review

I. Perception of gender bias in AI systems

Gender bias in artificial intelligence (AI) refers to systematic discrimination embodied in algorithms that perpetuate inequities against women and marginalized gender groups. This bias manifests for reasons such as historical inequalities, inaccurate data, and design errors in AI systems. Different scholars from various disciplines have made researches on how such systems reflect and amplify the prevalent biases, thus throwing into limelight the social, legal, and technological faces of the issue.

1. Bias in AI Systems Definition

Friedman and Nissenbaum (1996)[9] provide a foundational understanding of bias in computer systems, categorizing it into three types:

1. **Existing biases;** these originate in society and other institution.
2. **Technical bias,** which arises from the limitations of technology, including data processing algorithms;
3. **Emergent bias,** which develops during the system's application in dynamic real-world contexts.

These categories help outline how gender bias enters AI, since algorithms reproduce and amplify the inequalities embedded in society.[10] Barocas et al. (2016) posit that machine learning models, often seemingly neutral, quickly inherit prevailing biases found within training data, which themselves derive from historical inequities; for example, gender imbalances in the labor force are simply encoded into hiring tools, perpetuating inequity.[11]

2. Manifestations of Gender Bias

Society carries this gender bias in many ways on the applications of AI.

[9] Friedman, Batya, and Helen Nissenbaum. "Bias in Computer Systems." ACM Transactions on Information Systems (TOIS) 14, no. 3 (1996): 330-47.
[10] Supra note 9
[11] Barocas, Solon, and Andrew D. Selbst. "Big Data's Disparate Impact." California Law Review 104 (2016): 671.

1. **Facial Recognition Technology**

   One of the most researched areas is that of bias in facial recognition systems. Buolamwini and Gebru's (2018) study on Gender Shades exposed stark inaccuracies in facial recognition technology. They proved that systems like those from IBM, Microsoft, and Face++ had error rates as high as 34.7% for darker-skinned women compared to near-perfect accuracy for white men.[12] These discrepancies are not just technical bugs but have real implications with serious injustices like false arrests. For instance, the reports from the United States show how biased facial recognition algorithms led to the misidentification of Black women and furthered racial and gender inequality.[13]

2. **Recruitment Algorithms**

   Recruitment tools powered by AI are also biased, such as Amazon's AI-based hiring system, which showed favoritism towards male applicants than their female counterparts with equivalent qualifications.[14] The bias arose from training on historical hiring data where men were more dominant in the industry.[15] Such systems therefore run counter to core principles of fairness and equal opportunity guaranteed in employment law.

3. Voice Assistants and Stereotyping

Voice assistants like Alexa, Siri, and Google Assistant have been critiqued for reinforcing gender stereotypes.[16] West et al. (2019) argue that the submissive, female-coded responses of these assistants perpetuate harmful gender norms, reinforcing stereotypes of women as passive and accommodating. The design

[12] Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (2018): 77-91.
[13] Meissner, Christian A., and John C. Brigham. "Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review." Psychology, Public Policy, and Law 7, no. 1 (2001): 3-35.
[14] Rietdijk, Sofie. "The Relationship Between AI Recruitment and Gender Bias: A Literature Review." (2024).
[15] Obermeyer, Ziad, et al. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." Science 366, no. 6464 (2019): 447-453.
[16] Builtin, AI Voice Assistant Bias: Gender Stereotypes and the Consequences They Create (Oct. 30, 2023), https://builtin.com/artificial-intelligence/ai-voice-assistant-bias#:~:text=The%20Gender%20Bias%20Problem%20in,gender%20stereotypes%20this%20decision%20perpetuates.

decisions underpinning such systems fail to account for their broader societal impact, particularly in shaping cultural perceptions of gender roles.[17]

3. Bias in data collection

The basis of AI systems is the data used to train them. D'Ignazio and Klein (2020) in their influential work Data Feminism, emphasize that lack of diversity in datasets is one of the primary drivers of bias. For example, women and other marginalized groups are often underrepresented in datasets, which in turn leads to poor performance algorithms for those populations. They argue that inclusivity in data collection is not just a technical requirement but a political and ethical obligation.[18]

Besides representation, race, gender, and socio-economic status intersect in such a way that further complicates the bias in AI. Scholars have underlined the necessity of the intersectional lens to understand how compounded forms of discrimination[19] shape outcomes in AI applications. Most AI systems, however, treat gender as a standalone category and fail to take into account how other factors can exacerbate gender bias.

II. Legal and Regulatory Responses to AI Bias

The legal and regulatory landscape around AI bias remains relatively nascent with few frameworks that tackle the problem head-on. Most of the responses remain piecemeal and tend to center around specific domains such as employment or privacy rather than broader systemic issues of bias and discrimination.

1. Domestic Legal Frameworks India:

Articles 14 and 15 provide the constitutional guarantees of equality and non-discrimination. The provisions have not been implemented successfully on AI systems. The Information Technology Act, 2000[20], focuses more on cybercrime and data protection but does not touch on algorithmic accountability and bias in AI systems (Bhatia, 2021). Scholars opine that the

[17] Sarah Myers West, Meredith Whittaker & Kate Crawford, Discriminating Systems, AI Now (2019), https://ainowinstitute.org/discriminating-systems.
[18] Catherine D'Ignazio and Lauren F. Klein, Data Feminism (MIT Press 2023).
[19] Simone Cusack, "Discrimination against Women: Combating its Compounded and Systemic Forms," 34 Alternative L.J. 86 (2009).
[20] Information Technology Act, No. 21 of 2000, Acts of Parliament, 2000 (India).

Indian approach to AI regulation has been reactive rather than proactive, addressing the issues as they arose, instead of mitigating systemic risks beforehand.[21]

For example, India has shown incredible advancements in AI deployment, especially in the sectors of healthcare and governance. There has been little effort towards auditing these systems for biases. Without laws to oblige fairness audits or algorithmic transparency, such biased systems are free to operate unrestrained.

Global Contexts:

In the U.S. and the UK, anti-discrimination laws form a basic foundation from which biased AI systems can be challenged. For instance, Title VII of the Civil Rights Act in the U.S.[22] and the Equality Act, 2010[23], in the UK outlaw discriminatory practices in the hiring and other spheres of life. However, Wachter et al. (2020) indicate that such laws are rarely sufficient to counter the issues presented by AI.[24] For example, opacity in machine learning models - a problem known commonly as the "black box" problem[25] - makes it impossible to prove intentional discrimination, an element that must be shown in most anti-discrimination statutes.[26]

2. International standards and frameworks CEDAW and human rights law

The Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW)[27] provides a robust framework for addressing gender bias. Articles 2 and 5 obligate states to eliminate systemic discrimination and stereotypes, including those perpetuated through technology. Ahmed (2022) argues that applying CEDAW to AI governance could provide a much-needed rights-based framework for addressing bias.[28]

[21] Bhatia, Swati, Veenu Arora & Shweta Batra, The Role of Artificial Intelligence in Developing an Effective Training Assessment Tool, in AI-Oriented Competency Framework for Talent Management in the Digital Economy 244-55 (CRC Press 2023).
[22] Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e et seq. (2020).
[23] Equality Act 2010, c. 15 (UK).
[24] Wachter, Sandra, Brent Mittelstadt & Chris Russell, Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI, 41 Comput. L. & Sec. Rev. 105567 (2021).
[25] Castelvecchi, Davide. "Can We Open the Black Box of AI?" Nature News, 538.7623 (2016): 20.
[26] Von Eschenbach, Warren J. "Transparency and the black box problem: Why we do not trust AI." Philosophy & Technology 34.4 (2021): 1607-1622.
[27] CEDAW, art. 1.
[28] Ahmed, Imran, Gwanggil Jeon, and Francesco Piccialli. "From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where." IEEE Transactions on Industrial Informatics 18.8 (2022): 5031-5042.

European Union AI Act

The EU's AI Act of 2021 takes one step forward in the regulation of AI systems. There is a risk-based classification system, which will strictly monitor the high-risk applications like hiring and facial recognition.

Yet, critics like Veale & Borgesius (2021) argue that there are poor mechanisms of enforcement under the Act, especially concerning ensuring compliance outside the EU. The Act does not have explicit provisions for intersectional biases; rather, it is focused on the more general categories of harm.

3. Gaps in Regulation

Despite these strides, significant gaps still persist within the regulatory landscape:

1. **Lack of Algorithmic Audits:** Most jurisdictions do not require regular audits for AI systems, which means there is no check on biases in algorithms.[29]
2. **Mechanisms for accountability:** Developers or deployers of AI systems often lack accountability under existing frameworks for unfair outcomes.[30]
3. **Intersectionality:** Only a few legal frameworks account for the compounding effect that intersectional forms of discrimination like gender and race have on AI outputs.[31]

III. Ethical and Social Implications of AI Bias

Beyond the legal and regulatory, ethical dimensions of AI bias are no less important. These perspectives cast a spotlight on the broader implications of biased AI systems, calling for systemic change.

1. Systemic source of bias

The important point to consider in her book Algorithms of Oppression, Noble (2018) argues that AI bias is more than a technical flaw-it is a reflection of the systemic inequalities. She decries the underrepresentation of women and other minorities in the tech industry as directly

[29] Santoni de Sio, Filippo, and Giulio Mecacci. "Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them." Philosophy & Technology 34.4 (2021): 1057-84.

[30] Hu, Shiming, and Yifan Li. "Policy Interventions and Regulations on Generative Artificial Intelligence: Key Gaps and Core Challenges." Proceedings of the 25th Annual International Conference on Digital Government Research. 2024.

[31] Hartmann, David, et al. "Addressing the regulatory gap: moving towards an EU AI audit ecosystem beyond the AI Act by including civil society." AI and Ethics (2024): 1-22.

influencing how datasets are framed and interpreted. Male, white viewpoints dominate AI development, with a deepening of gender hierarchies embedded into technological systems.[32]

2. Technical Fixes vs. Transformation

Many researchers (Barocas et al., 2017) pointedly question the overemphasis put on technical interventions, of bias mitigation algorithms and measures with fairness metrics. All instruments do is reduce rather than remove the effects that show bias, but such doesn't address the heart issue of disparity in education and working or job opportunities for women and would instead require all-around systemic overhauls, including incorporating all sorts of diversity at working teams in AI-related field and participatory methodology incorporated into design.[33]

3. Human Rights and Ethical AI

a. Embedding Human Rights Principles:

Scholars like Kirkpatrick, 2020, urge that the governance of AI must adopt human rights principles, including inclusive data collection and obligation for transparency in the algorithms behind decision-making.[34]

b. Corporate Responsibility:

The United Nations Guiding Principles on Business and Human Rights[35] emphasize the duty of enterprises to mitigate adverse impacts based on discrimination. According to Deva (2021), these principles should include AI developers, thereby holding them liable for the social implications of their technological creations.

IV. Case Studies

1. Global Case Analyses

[32] Noble, Safiya Umoja. "Algorithms of Oppression: How Search Engines Reinforce Racism." Algorithms of Oppression. New York University Press, 2018.

[33] Barocas, Daniel A., et al. "Association Between Radiation Therapy, Surgery, or Observation for Localized Prostate Cancer and Patient-Reported Outcomes After 3 Years." JAMA 317.11 (2017): 1126-1140.

[34] Nawaz, Fahad, Wisal Ahmad, and Muhammad Khushnood. "Kirkpatrick Model and Training Effectiveness: A Meta-Analysis 1982 to 2021." Business & Economic Review 14.2 (2022): 35-56.

[35] United Nations Guiding Principles on Business and Human Rights, U.N. Doc. HR/PUB/11/04 (2011).

Amazon Case

In 2018, Amazon had to cancel its AI-driven recruitment system because of its in-built gender bias. The algorithm displayed a stark preference for male applicants for technical roles; it had been trained on historical recruitment data that were predominantly comprised of male candidates. The tool discriminated against résumés that featured language linked to female-oriented activities, such as "women's chess club," or those that came from institutions for women.[36]

Analysis

Although Amazon's tool never was implemented, the potential violation of Title VII of the U.S. Civil Rights Act, which prohibits employment discrimination on the basis of sex, shows legal risks involved in the use of AI systems lacking safeguards against bias. It is one of those situations that indicate the need to incorporate fairness evaluation and diverse training data sets into AI systems. Moreover, it emphasizes the necessity for tailored regulatory frameworks within the United States to guarantee that artificial intelligence adheres to established anti-discrimination legislation.

R (on the application of Bridges) v. South Wales Police

The United Kingdom gives an excellent example of the use of judicial activism to address R (on the application of Bridges) v. South Wales Police. Here, Edward Bridges, a civil liberties campaigner, challenged South Wales Police's use of automated facial recognition (AFR) technology on grounds that he said it was violating his right to private life protected under Article 8 of the European Convention on Human Rights (ECHR). The AFR system faced considerable criticism due to its elevated error rates, especially in the misidentification of women and people of color. In 2020, the Court of Appeal ruled in favor of Bridges, highlighting that the implementation of AFR technology was deficient in legal protections and disproportionately violated rights related to privacy and equality. [37]

Analysis

[36] Reuters, Insight: Amazon scraps secret AI recruiting tool that showed bias against women, Reuters (Oct. 10, 2018), https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against- women-idUSKCN1MK0AG.

[37] R (on the application of Bridges) v. South Wales Police [2020] EWCA Civ 1058.

This ruling established an important precedent, illustrating the judiciary's function in limiting the unregulated use of biased artificial intelligence systems. The decision further emphasized the significance of proactive legal structures, such as algorithmic impact evaluations and judicial supervision, to guarantee adherence to human rights and anti-discrimination standards.

2. National Case Analyses (India)

A landmark precedent of legal scrutiny over artificial intelligence systems within the Indian framework is the case in **Justice K.S. Puttaswamy v. Union of India (2017)[38].** The Supreme Court had declared the right to privacy as a fundamental right within Article 21 of the Indian Constitution and further introduced the concept of proportionality tests when an infringement of this right occurred. Although the case is mainly about the Aadhaar biometric system, its impact expands into the artificial intelligence governance arena, particularly in systems collecting and managing personal data. Principles of necessity, legitimacy, and proportionality outlined in this case can be used to challenge artificial intelligence systems that focus disproportionately or negatively on disadvantaged groups. For example, a predictive policing tool, which differentially tracks a particular community based on biased data, can be challenged on the grounds of constitutional guarantees of equality and privacy.

Similarly, the judgment in Navtej Singh Johar v. Union of India (2018)[39] expanded the scope of anti-discrimination principles as enshrined in the Indian Constitution. The Supreme Court reaffirmed the constitutional commitment to inclusivity and equality by decriminalizing consensual same-sex relationships and emphasizing dignity and autonomy. This judgment has far-reaching implications for AI governance, particularly in the area of bias reduction against LGBTQ+ individuals. The artificial intelligence-based systems of recruitment algorithms and credit assessment mechanisms with bias against the nonheteronormative identities shall be called for judicial review in accordance with the principles of anti-discrimination in this judgment.

In toto, the two judgments- Puttaswamy and Navtej Singh Johar-have comprehensively established the constitutional basis of oversight over AI systems for adherence to the principles of fairness, inclusivity, and equality.

[38] K.S. Puttaswamy v. Union of India, (2017) 10 SCC 1 (India).
[39] Navtej Singh Johar v. Union of India, (2018) 10 SCC 1 (India).

3. Cross-Jurisdictional Learnings

The EU's preventive approach to the regulation of AI is of immense interest for India. The EU draft AI Act divides AI applications based on the level of risks associated with them and seeks stricter controls for applications found to be high-risk, such as recruitment and policing. By ensuring bias audits, diversity in data, and algorithm transparency, the EU regulatory regime hopes to prevent biased outcomes and provide accountability. In addition, the General Data Protection Regulation (GDPR)[40] offers broad protections for personal rights that include the right to explanation as well as the right to contest automated decisions. This helps individuals challenge biased outcomes resulting from artificial intelligence and contributes to greater accountability from developers.

Though India may draw lessons from the European Union's systematic framework of regulations, it has to address its unique issues. A prime issue is that artificial intelligence training data does not contain sufficient diversity. Such datasets very often fail to mirror the vast range of castes, classes, and linguistic differences existing in the country. In this manner, the system of inequality and social hierarchy gets accentuated in AI results. In this regard, Indian law needs to mandate proportionate representation of all social classes in the data sets used for training artificial intelligence. Additionally, considering the very low levels of digital literacy and technological access among India's underprivileged, regulatory mechanisms must include public education programs and accessible avenues for redressal of grievances. This would allow the victim to file objections to discriminatory AI systems and to seek appropriate relief.

Systemic inequalities like caste and gender disparities are also a challenge to the fair governance of AI in India. The EU is on the opposite side, as regulatory efforts are aimed at preventing bias. Instead, India must take an intersectional approach that makes explicit socio-economic factors shape AI outcomes. For example, predictive policing tools may target Dalit and Muslim communities disproportionately because of the historical biases embedded in the data of law enforcement. A comprehensive regulatory regime for AI in India should involve not just bias testing, but the strategies to address those inherent imbalances proactively through design and data representations.[41]

[40] General Data Protection Regulation (GDPR), Regulation (EU) 2016/679.
[41] Bircan, Tuba, and Mustafa F. Özbilgin. "Unmasking inequalities of the code: Disentangling the nexus of AI and inequality." Technological Forecasting and Social Change 211 (2025): 123925.

Ultimately, the Indian judiciary will be able to significantly influence the governance of AI by relying on precedents established both within and outside the country's borders. The principles enunciated in Puttaswamy and Navtej Singh Johar judgments provide a constitutional footing for challenging discriminatory AI regimes that infringe fundamental rights. Moreover, judicial reasoning in cases like Bridges v. South Wales Police can provide Indian courts with much-needed guidance in evaluating AI surveillance technologies. India may succeed in creating a good model of governance that gives reasonable, transparent, and responsible AI systems by enrooting these legal precepts through a seamless regulation policy. Such an approach may eventually eliminate gender bias among other biases in artificial intelligence. The same would thereby endorse the principles of the constitutional order in the emerging cyber world as well.

Framework Suggested Towards Counter Measures against AI systems and Their Gender Biases

The persistence of gender bias in artificial intelligence, despite the rising awareness and slow but growing regulatory efforts, is evidence of the inadequacy of the current framework to address the root causes and systemic impacts. Solutions offered tend to be technical or piecemeal regulations that do not bring social, legal, and ethical components into the equation. This paper proposes a holistic model, based on legal analysis and a human rights approach, to address gender bias in AI systems. The proposed model underscores three fundamental pillars: mechanisms for legal reform and accountability, frameworks for algorithmic governance, and interventions at the systemic societal level. These pillars are interrelated and jointly seek to reduce gender bias while promoting inclusivity, transparency, and accountability.

Pillar 1: Legal Reform and Accountability Mechanisms

Legal frameworks lay a strong foundation in guaranteeing the accountability of developers and deployers of AI. However, most jurisdictions have not developed cohesive or enforceable laws that take into account algorithmic bias and its intersectional dimensions. One of the ways to solve the problem of opacity of algorithms-the "black box problem"-is to require developers to be transparent about methodologies, training datasets, and decision-making processes of AI systems. Such transparency can be enforced with the help of regulatory bodies similar to that of the financial auditing standards, holding public and policymakers' attention. Independent AI system auditing should be compulsorily done by independent auditors who are certified in their domains and are not employees of the AI companies. Such an independent audit can be brought about mandatorily. For example, an oversight body like that of the CCI would regulate such auditing, concentrating on gendered impact. Laws on nondiscrimination, such as Title VII of the U.S. Civil Rights Act or equality guarantees under Articles 14 and 15 of the Indian Constitution, should cover discriminatory outcomes produced by AI systems. Victims of algorithmic bias should be allowed to seek legal remedies. The law must also take into consideration compounded biases that can emerge when combining gender and race biases. These should be mandated with appropriate protection for marginalized groups in relation to the intersectionality theory by Kimberlé Crenshaw. Strict liability should be imposed on developers and deployers of AI systems in areas like healthcare, hiring, and law enforcement, with or without intent, just like the principle of absolute liability in Indian environmental law in MC Mehta v. Union of India (1987). Liability should also be imposed on corporations deploying AI systems so that responsibility is shared. Enhancement of oversight of the judiciary will call for special judicial education as well as creation of algorithm bias dispute resolution tribunals comprising technical as well as legal expertise.

Pillar 2: Algorithmic Governance Frameworks

Whereas technological answers would necessarily form part of correcting structural issues with AI systems, there also has to be appropriate governance frameworks that provide ethical foundations as well as the capability for the law. Ethics-by-design principles should be enacted into law and thus be complied with by developers at design and deployment time in standard protocols of detecting and mitigating bias with the help of guidelines, like the EU's High-Level Expert Group on AI Ethics Guidelines. The human element of oversight should be introduced to all high-stakes decisions AI may take, such as the algorithm determining whether a candidate is fit for the job. Data inclusivity standards must be established, mandating proportional representation of genders and marginalized groups in training datasets, enforced through data protection laws similar to GDPR. To protect marginalized groups' privacy, data collection should prioritize anonymization and informed consent, aligning with the UN Guiding Principles on Business and Human Rights. Algorithmic Impact Assessments (AIAs), modeled after environmental impact assessments, should be mandatory for all AI systems that are deployed in sensitive domains, evaluating potential biases in training data and algorithms and societal and legal impacts with a focus on gender equity. These should be done pre-deployment and at regular intervals to deal with emergent biases.

Pillar 3: Systemic Societal Interventions

Legal and technical solutions alone cannot eliminate gender bias from AI, making systemic societal interventions critical for addressing structural inequities. The lack of diversity in AI development teams is a significant driver of bias, necessitating legislative mandates in the workplace for diversity, such as quotas in tech companies focusing on AI roles, modeled after Norway's gender quota laws for corporate boards. Educational reforms introducing AI ethics and gender studies into STEM curricula can cultivate developers attuned to bias and inclusivity. Public awareness campaigns are essential to educate the public on the risks and consequences of AI bias while empowering individuals to identify and report discriminatory outcomes. Governments and civil society organizations should collaborate to strengthen oversight by providing grants to organizations conducting independent research on AI bias and establishing platforms for individuals to report instances of bias, enabling civil society to monitor systemic trends and advocate for change.

The rationale for this model is rooted in legal principles and established precedents. It aligns with constitutional guarantees of equality and international human rights obligations, such as CEDAW and ICCPR, which require the elimination of systemic discrimination. The model puts preventive justice at the forefront by mandating audits and impact assessments, thus ensuring that discriminatory systems are identified and mitigated before causing harm. The liability framework draws on principles of strict and absolute liability, ensuring accountability for biased systems. Additionally, the model fosters participatory governance by involving civil society and promoting public awareness, ensuring marginalized voices are included in AI policymaking. This comprehensive approach aims to address gender bias in AI systems holistically, recognizing the multifaceted nature of the problem and integrating legal, technical, and societal solutions.

Potential Challenges and Limitations of the Proposed Model

The proposed model faces challenges that may impede its adoption. Technical complexity in AI systems, especially deep learning models, often renders their decision-making opaque

even when transparency is mandated. Furthermore, balancing algorithmic transparency with intellectual property protections might be resisted by corporations, who may fear that the revelation of trade secrets will undermine their interests.

On the legal front, extending anti-discrimination laws to AI and implementing strict liability frameworks may face significant legislative barriers. Overburdened judicial systems and the need for specialized AI tribunals present further barriers, especially in resource-constrained jurisdictions. Similarly, mandatory algorithmic impact assessments (AIAs), though essential, may prove resource-intensive, especially for smaller companies, potentially consolidating power in the hands of larger corporations.

Systemic obstacles also plague societal interventions. Long-term educational reforms are necessary to achieve diversity in AI development, while diversity quotas usually evoke resistance or risk tokenism. Public awareness campaigns are unlikely to reach marginalized communities where digital access or literacy is poor. This would ultimately add another layer of complexity in terms of enforcing local regulations over AI development globally. Indeed, there is a global lack of agreement over who should govern AI development and whether the development should be internationalized.

Conclusion

This paper has discussed the critical issue of gender bias in AI systems and proposed a comprehensive model to mitigate such biases with a multifaceted approach. The increasing presence of AI technologies in all sectors from hiring and law enforcement to health care and education calls for a serious reevaluation of their design and deployment so that they do not perpetuate or even worsen existing gender inequalities.

The entire research revealed a great void in the current governance of AI, especially in legal frameworks that are not responsive to algorithmic discrimination, lack of transparency in AI systems, and the underrepresentation of marginalized groups in AI development. All these were filled by proposing a model integrating legal reforms, algorithmic governance frameworks, and societal interventions. This model is designed to ensure that AI technologies are developed and deployed in a manner that is fair, inclusive, and accountable to all societal groups, particularly to women and marginalized communities.

It has researched the paper's very questions: how legal frameworks can reduce gender bias in AI, what kind of technical solutions can be created to mitigate bias, and the way societal interventions can fuel systemic change. Legal reform will form the first pillar of this model,

where there will be stronger laws to bind the accountability of AI developers and companies behind discriminatory outcomes. In other words, through proposed requirements of transparency, mandatory audits, and expansion of anti-discrimination laws, this pillar makes it legally tenable to challenge and correct gender bias in AI. This framework allows individuals to demand redress and puts pressure on the corporations to engage in more just and equitable practices while filling an existing gap within current AI governance that ignores algorithmic bias legal implications.

Algorithmic governance

This framework probes into both technical and ethical elements of AI bias. In this light, it makes an argument to introduce ethics-by-design into AI systems while encouraging more inclusive data practices along with an AIA of algorithms. This pillar ensures that gender bias is prevented from being ingrained into AI systems by embedding fairness at the design stage and regularly auditing AI systems for bias. It directly addresses the research question of how technical solutions can reduce algorithmic bias and create a more equitable technological landscape.

Systemic societal interventions

These include actions that address the need for change at a societal level. They include increased diversity in the development teams for AI, as well as greater public awareness of the risks associated with biased AI systems. These interventions are fundamental to addressing the root causes of gender bias in AI because they address the structural inequities that influence how AI systems are developed and deployed. The diversity quota is one of these and this pillar is aimed to explain that solving AI bias entails more than a mere fix in the law or at the technical levels but at society's level toward more inclusionary and fair society.

This paper provides a very comprehensive approach towards gender bias in AI demonstrating how a combination of legal reforms, technical solutions, and societal interventions could build the way toward creating an equity-promoting AI ecosystem. The model focuses on human rights, accountability, and inclusivity, filling the research gaps in the existing literature to address algorithmic bias comprehensively. The model thus lays out a framework for meaningful progress on reducing gender bias in AI systems, with an emphasis on transparency, accountability, and diversity.

Future Research Directions

The model presented here is comprehensive but needs testing and refinement of its components in further research. Another avenue of research is the feasibility of mandatory algorithmic audits and impact assessments across industries, especially those where AI makes the biggest difference, like healthcare and criminal justice. Cost-effective audit frameworks accessible to both small companies and large corporations will play a critical role in ensuring that the model scales.

Further empirical research may also be required to explore how well diversity quotas and inclusive data practices work within AI development teams. It may thus be possible to study the effects of such interventions on long-term changes in reducing gender bias in AI systems, yielding valuable data on its effectiveness and feasibility for broader-scale application.

Finally, future research may focus on how AI bias intersects with one another, especially in how it discriminates against not just women but also other minoritised groups, such as racial minorities, disabled individuals, and LGBTQ+ communities. A greater understanding of how these intersecting identities shape algorithmic bias will be essential in developing more inclusive and equitable AI systems. Comparative studies across jurisdictions may even add insights to best practices and shape the global standard by which to mitigate algorithmic bias. In summary, although this paper provides a sound base for addressing gender bias in AI, further research will be critical to test the proposed model in practice, tune its components, and extend it in order to ensure that AI technologies contribute to a fairer and more just society for all.