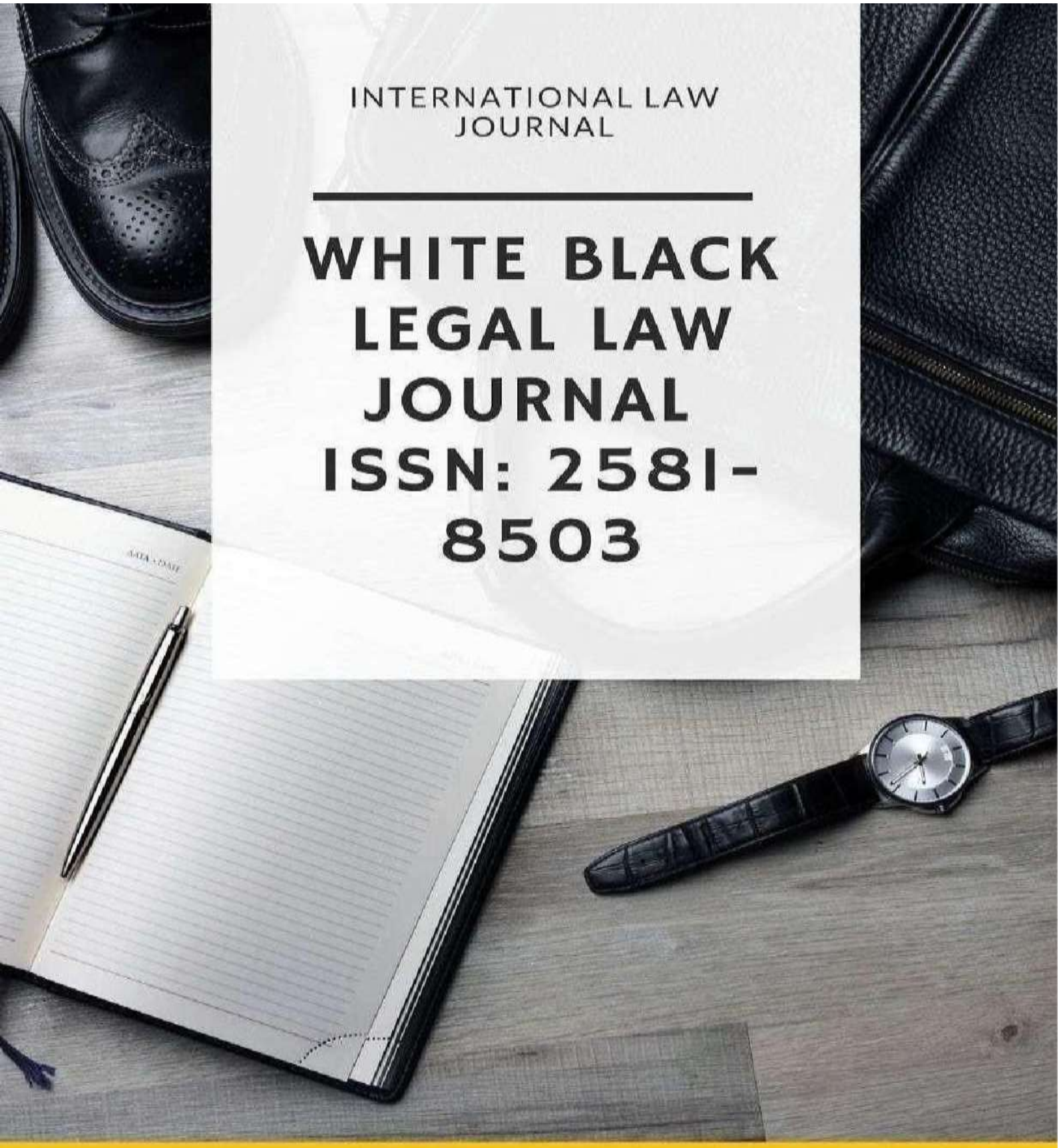

INTERNATIONAL LAW JOURNAL

WHITE BLACK LEGAL LAW JOURNAL

ISSN: 2581-8503

Peer - Reviewed & Refereed Journal

The Law Journal strives to provide a platform for discussion of International as well as National Developments in the Field of Law.

WWW.WHITEBLACKLEGAL.CO.IN

DISCLAIMER

EDITORIAL TEAM

Raju Narayana Swamy (IAS) Indian Administrative Service officer

Dr. Raju Narayana Swamy popularly known as Kerala's Anti Corruption Crusader is the All India Topper of the 1991 batch of the IAS and is currently posted as Principal Secretary to the Government of Kerala . He has earned many accolades as he hit against the political-bureaucrat corruption nexus in India. Dr Swamy holds a B.Tech in Computer Science and Engineering from the IIT Madras and a Ph. D. in Cyber Law from Gujarat National Law University . He also has an LLM (Pro) (with specialization in IPR) as well as three PG Diplomas from the National Law University, Delhi- one in Urban Environmental Management and Law, another in Environmental Law and Policy and a third one in Tourism and Environmental Law. He also holds a post-graduate diploma in IPR from the National Law School, Bengaluru and a professional diploma in Public Procurement from the World Bank.

Dr. R. K. Upadhyay

Dr. R. K. Upadhyay is Registrar, University of Kota (Raj.), Dr Upadhyay obtained LLB , LLM degrees from Banaras Hindu University & Phd from university of Kota.He has succesfully completed UGC sponsored M.R.P for the work in the ares of the various prisoners reforms in the state of the Rajasthan.

Senior Editor

Dr. Neha Mishra

Dr. Neha Mishra is Associate Professor & Associate Dean (Scholarships) in Jindal Global Law School, OP Jindal Global University. She was awarded both her PhD degree and Associate Professor & Associate Dean M.A.; LL.B. (University of Delhi); LL.M.; Ph.D. (NLSIU, Bangalore) LLM from National Law School of India University, Bengaluru; she did her LL.B. from Faculty of Law, Delhi University as well as M.A. and B.A. from Hindu College and DCAC from DU respectively. Neha has been a Visiting Fellow, School of Social Work, Michigan State University, 2016 and invited speaker Panelist at Global Conference, Whitney R. Harris World Law Institute, Washington University in St.Louis, 2015.

Ms. Sumiti Ahuja

Ms. Sumiti Ahuja, Assistant Professor, Faculty of Law, University of Delhi,

Ms. Sumiti Ahuja completed her LL.M. from the Indian Law Institute with specialization in Criminal Law and Corporate Law, and has over nine years of teaching experience. She has done her LL.B. from the Faculty of Law, University of Delhi. She is currently pursuing Ph.D. in the area of Forensics and Law. Prior to joining the teaching profession, she has worked as Research Assistant for projects funded by different agencies of Govt. of India. She has developed various audio-video teaching modules under UGC e-PG Pathshala programme in the area of Criminology, under the aegis of an MHRD Project. Her areas of interest are Criminal Law, Law of Evidence, Interpretation of Statutes, and Clinical Legal Education.

Dr. Navtika Singh Nautiyal

Dr. Navtika Singh Nautiyal presently working as an Assistant Professor in School of law, Forensic Justice and Policy studies at National Forensic Sciences University, Gandhinagar, Gujarat. She has 9 years of Teaching and Research Experience. She has completed her Philosophy of Doctorate in ‘Intercountry adoption laws from Uttranchal University, Dehradun’ and LLM from Indian Law Institute, New Delhi.

Dr. Rinu Saraswat

Associate Professor at School of Law, Apex University, Jaipur, M.A, LL.M, Ph.D,

Dr. Rinu have 5 yrs of teaching experience in renowned institutions like Jagannath University and Apex University. Participated in more than 20 national and international seminars and conferences and 5 workshops and training programmes.

Dr. Nitesh Saraswat

E.MBA, LL.M, Ph.D, PGDSAPM
Currently working as Assistant Professor at Law Centre II, Faculty of Law, University of Delhi. Dr. Nitesh have 14 years of Teaching, Administrative and research experience in Renowned Institutions like Amity University, Tata Institute of Social Sciences, Jai Narain Vyas University Jodhpur, Jagannath University and Nirma University.
More than 25 Publications in renowned National and International Journals and has authored a Text book on Cr.P.C and Juvenile Delinquency law.

Subhrajit Chanda

BBA. LL.B. (Hons.) (Amity University, Rajasthan); LL. M. (UPES, Dehradun) (Nottingham Trent University, UK); Ph.D. Candidate (G.D. Goenka University)

Subhrajit did his LL.M. in Sports Law, from Nottingham Trent University of United Kingdoms, with international scholarship provided by university; he has also completed another LL.M. in Energy Law from University of Petroleum and Energy Studies, India. He did his B.B.A.LL.B. (Hons.) focussing on International Trade Law.

ABOUT US

WHITE BLACK LEGAL is an open access, peer-reviewed and refereed journal providededicated to express views on topical legal issues, thereby generating a cross current of ideas on emerging matters. This platform shall also ignite the initiative and desire of young law students to contribute in the field of law. The erudite response of legal luminaries shall be solicited to enable readers to explore challenges that lie before law makers, lawyers and the society at large, in the event of the ever changing social, economic and technological scenario.

With this thought, we hereby present to you

www.whiteblacklegal.co.in
Volume 3 Issue 1 | April 2025 ISSN: 2581-8503

AI BIAS AND DISCRIMINATION: LEGAL REMEDIES AND POLICY SOLUTIONS

AUTHORED BY - KOPAL TEWARI

ABSTRACT

Artificial Intelligence (AI) is increasingly shaping critical decision-making processes across various domains, including hiring, criminal justice, financial services, and healthcare. While AI offers efficiency and data-driven insights, it also introduces significant risks of bias and discrimination. This paper explores the nature of AI bias, its origins, and its impact on marginalized groups. AI bias often stems from biased training data, flawed algorithms, and human intervention, leading to discriminatory outcomes that reinforce societal inequalities. Through case studies, this paper examines real-world instances of AI bias, including Amazon's hiring algorithm, which exhibited gender discrimination; the COMPAS algorithm, which disproportionately labelled Black defendants as high-risk offenders; and the Apple Card, which allocated lower credit limits to women compared to men. These cases underscore the pressing need for regulatory frameworks to mitigate AI bias. Existing legal responses are analysed, including the European Union's AI Act and General Data Protection Regulation (GDPR), the OECD AI Principles, and AI regulatory proposals in India. While these frameworks aim to enhance transparency, accountability, and fairness, they face challenges in enforcement and adaptability to evolving AI technologies. The paper highlights gaps in legal structures and the need for harmonized global standards to ensure AI accountability. To address these challenges, the paper explores policy solutions, including stronger regulatory compliance mechanisms, bias detection audits, and public-private sector collaborations. Policymakers must integrate ethical AI governance principles, ensuring fairness and non-discrimination in AI applications. Additionally, interdisciplinary cooperation between technologists, legal experts, and policymakers is essential to create robust oversight mechanisms. This paper concludes that while AI holds transformative potential, unchecked bias can perpetuate discrimination. A combination of legal reforms, ethical AI development, and proactive governance is necessary to ensure that AI serves as an equitable tool rather than a discriminatory force. Addressing AI bias requires a dynamic legal approach that evolves alongside AI advancements to safeguard fundamental rights and social justice.

Keywords: *AI Bias, Algorithmic Discrimination, AI Regulation, Ethical AI, AI and Human Rights, Legal Frameworks for AI.*

I. INTRODUCTION

Artificial Intelligence ("AI") is a branch of computer science that seeks to create machines capable of simulating human intelligence. AI systems leverage data-driven algorithms to process information, recognize patterns, and make decisions with minimal human intervention. Over the past decade, AI has evolved from simple rule-based automation to sophisticated machine learning and deep learning models capable of handling complex tasks, including language processing, medical diagnosis, and autonomous decision-making. AI is categorized into different levels based on its capabilities. Narrow AI (*Weak AI*) is designed to perform specific tasks, such as virtual assistants, fraud detection, and facial recognition. General AI (*Strong AI*), which remains theoretical, would exhibit human-like intelligence, capable of reasoning and problem-solving across various domains. As AI capabilities expand, its influence on decision-making processes in critical sectors continues to grow, necessitating careful consideration of its ethical, legal, and social implications.[1]

A. AI's increasing role in decision-making

AI is transforming decision-making across multiple industries, automating complex tasks that were traditionally performed by humans. In healthcare, AI assists in disease diagnosis, drug discovery, and treatment recommendations. In finance, AI-driven credit scoring and risk assessment models streamline lending processes. Law enforcement agencies use AI for crime prediction, facial recognition, and surveillance, while AI-driven recruitment tools help companies select candidates based on predictive hiring models.[2]

Despite its efficiency, AI-driven decision-making presents significant challenges. One major concern is the "*black box*"[3] nature of AI models, where decision-making processes remain opaque and difficult to interpret. The lack of transparency raises accountability issues, especially in high-stakes fields such as criminal justice, healthcare, and hiring. AI systems also

[1] Russell S and Norvig P, Artificial Intelligence: A Modern Approach (4th edn, Pearson 2021)
[2] Wang X and others, 'Algorithmic Discrimination: Examining Its Types and Regulatory Measures with Emphasis on US Legal Practices' (2024) 7 Frontiers in Artificial Intelligence 1320277 <https://www.frontiersin.org/articles/10.3389/frai.2024.1320277/full> accessed 3 January 2025
[3] 'AI's Mysterious "Black Box" Problem, Explained University of Michigan-Dearborn' <https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained> accessed 7 January 2025

risk reinforcing existing biases present in training data, leading to discriminatory outcomes.

For instance, AI-powered tools used in hiring may favor certain demographics over others due to historical biases embedded in training datasets. Similarly, AI-based sentencing algorithms in the legal system may disproportionately classify individuals from marginalized communities as high-risk offenders. Given these challenges, ensuring that AI is fair, transparent, and accountable is essential for its responsible integration into society.[4]

B. Explanation of AI bias and its sources

AI bias refers to systematic and unfair discrimination that occurs when an AI system produces skewed results that disadvantage certain individuals or groups. Bias in AI is not an inherent flaw of the technology itself but rather a consequence of flaws in data collection, algorithm design, and human intervention.[5] The primary sources of AI bias include:

a) Data Bias

AI models learn from vast amounts of data, and if the data used for training is incomplete, imbalanced, or reflective of historical prejudices, the AI system may replicate and even amplify these biases. For example, a hiring algorithm trained on historical data where men were predominantly hired for tech roles may systematically disadvantage female candidates. Similarly, if facial recognition technology is trained primarily on lighter-skinned individuals, it may perform poorly in identifying individuals with darker skin tones.

b) Algorithmic Bias

The way AI models are designed and structured can introduce bias. Some machine learning algorithms prioritize efficiency over fairness, optimizing for accuracy without considering ethical implications. In predictive policing algorithms, for example, over-reliance on historical arrest data can result in racial profiling, reinforcing existing patterns of discrimination. Algorithmic bias can also arise from poorly calibrated weighting mechanisms, where certain variables disproportionately influence decision outcomes.

[4] Nathalie A. Smuha, An Introduction to the Law, Ethics, and Policy of Artificial Intelligence (Cambridge University Press 2025)

[5] Christina P., 'Ethical concerns mount as AI takes bigger decision making role in more industries' The Harvard Gazette (26 October 2020) <https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/> accessed 7 January 2025

c) Human Intervention and Implicit Bias

AI systems are created, trained, and implemented by human developers, who may inadvertently introduce their own biases into the model. Decisions made during feature selection, data labeling, and algorithm tuning can embed human prejudices into AI systems. Furthermore, the lack of diverse perspectives in AI development teams can lead to blind spots, where the system fails to account for the experiences of underrepresented communities.

d) Feedback Loops and Self-Reinforcing Bias

AI systems that continuously learn from their environment can develop self-reinforcing biases. If a recommendation algorithm is trained on biased user interactions, it may continue to suggest similar biased outcomes, exacerbating discrimination over time. For instance, AI-driven content recommendation systems on social media can create "*filter bubbles*,"[6] where users are repeatedly exposed to the same types of content, reinforcing pre-existing opinions and biases.

C. The Need for Legal and Policy Interventions

Given the increasing reliance on AI in decision-making, addressing AI bias is crucial to ensure fairness, equity, and accountability. Regulatory frameworks must evolve to keep pace with AI advancements, establishing clear guidelines for bias detection, algorithmic transparency, and ethical AI development.[7] Policymakers must consider implementing, firstly, bias auditing mechanisms to assess AI systems for discriminatory patterns before deployment. Secondly, transparency requirements to ensure AI decisions are interpretable and explainable. Thirdly, diverse and inclusive training data to mitigate bias at the data collection stage. Lastly, AI ethics committees to oversee the responsible deployment of AI in critical sectors.

While AI presents opportunities for innovation, unchecked bias can have serious consequences, exacerbating societal inequalities and eroding public trust. As AI continues to shape decision-making, legal and policy interventions must proactively address these risks, ensuring that AI systems align with ethical and legal standards.

[6] Chinmay B., 'How Filter Bubbles are biasing your opinions on Social Media' Medium (8 July 2023) <https://medium.com/data-and-beyond/how-filter-bubbles-are-biasing-your-opinions-on-social-media-9469b940154> accessed 7 January 2025

[7] O'Neil C, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (Crown Publishing Group 2016).

II. UNDERSTANDING AI BIAS AND DISCRIMINATION

AI bias refers to systematic and unfair discrimination that arises when AI systems produce skewed or prejudicial results, disproportionately affecting certain individuals or groups. AI bias is not merely a technical flaw but a reflection of broader social and historical inequities embedded in datasets, algorithms, and decision-making processes. It manifests in various forms, such as racial bias in facial recognition, gender bias in hiring algorithms, and socioeconomic bias in credit scoring systems. At its core, AI bias occurs when an algorithm favours one group over another in ways that deviate from fairness and objectivity. It can be explicit (arising from intentional discrimination) or implicit (resulting from unconscious biases embedded in training data or algorithmic logic). AI bias is particularly concerning because it can scale discrimination at unprecedented levels, reinforcing structural inequalities across domains such as employment, healthcare, law enforcement, and finance.[8] Bias in AI is not limited to isolated incidents; rather, it is an endemic issue rooted in historical prejudices and flawed design choices. It becomes especially problematic when AI systems are used in high-stakes environments where fairness, transparency, and accountability are paramount. Understanding the sources and impacts of AI bias is crucial to devising effective legal and policy measures to mitigate its harmful effects.

A. Sources and Impact of AI Bias

AI bias arises from multiple sources, often interacting in complex ways to produce discriminatory outcomes. These sources can be broadly categorized into three main areas: biased data, algorithmic flaws, and human intervention.[9]

a) Data Bias

AI systems rely on vast datasets to learn patterns and make predictions. However, if these datasets are unrepresentative, incomplete, or contain historical discrimination, the AI model will inherit and perpetuate these biases.[10] Several forms of data bias contribute to AI discrimination:

[8] Wachter S, Mittelstadt B and Russell C, 'Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI' (2021) Computer Law & Security Review 41, 1-14

[9] O'Neil C, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (Crown Publishing Group 2016)

[10] James Manyika and Brittany Presten, 'What Do We Do About the Biases in AI?' (2019) Harvard Business Review <https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai> accessed 4 February 2025

i. Historical Bias

AI models trained on historical data reflect societal prejudices. For example, if past hiring data shows a preference for male employees in tech jobs, an AI-driven hiring system may learn to replicate this pattern, disadvantaging female candidates.

ii. Sampling Bias

When training data does not accurately represent all demographic groups, AI models may produce skewed results.[11] A facial recognition system trained primarily on lighter-skinned individuals, for instance, may struggle to accurately identify people with darker skin tones.

iii. Labelling Bias

The process of categorizing and labelling data can introduce bias if subjective human judgments influence data annotation. For example, law enforcement datasets that label certain neighbourhoods as "high crime areas" based on past police activity can reinforce racial profiling.[12]

b) Algorithmic Bias

Even if data is unbiased, the way algorithms process information can introduce discriminatory patterns. Algorithmic bias can emerge from:

i. Feature Selection Bias

AI models prioritize certain variables over others when making decisions. For example, an AI loan approval system may heavily weigh an applicant's ZIP code, indirectly leading to racial or socioeconomic discrimination.

ii. Optimization Bias

AI models are typically optimized for accuracy, efficiency, or profitability, rather than fairness.[13] This can lead to unintended consequences, such as favouring high-income individuals in credit scoring while systematically disadvantaging lower-income applicants.

[11] 'Bias in AI' <https://azwww.chapman.edu/ai/bias-in-ai.aspx> accessed 8 February 2025
[12] Rejmaniak R, 'Bias in Artificial Intelligence Systems' (2021) 26 Białostockie Studia Prawnicze 25 <https://www.sciendo.com/article/10.15290/bsp.2021.26.03.02> accessed 8 February 2025
[13] Lepri B, Oliver N and Pentland A, 'Ethical Machines: The Human-Centric Use of Artificial Intelligence' (2021) 24 iScience 102249 <https://linkinghub.elsevier.com/retrieve/pii/S2589004221002170> accessed 8 January 2025

iii. Feedback Loop Bias

AI systems that continuously learn from real-world interactions can reinforce pre-existing biases.[14] For instance, predictive policing systems that recommend increased law enforcement presence in certain communities may create a cycle where more arrests occur, further validating the AI's original (biased) predictions.

c) Human Intervention and Implicit Bias

AI is designed and implemented by humans, who may unknowingly embed their own biases into the system. Human biases can manifest at various stages:

i. In AI Model Design

Developers make choices about which variables to include in AI models, potentially reinforcing biased outcomes.

ii. In Training Data Selection

If AI developers fail to ensure diverse and representative datasets, the model will produce biased results.

iii. In Policy Implementation

Organizations using AI may fail to critically evaluate their AI systems for fairness, leading to unintentional discrimination in hiring, lending, and legal decision-making.

B. Impact of AI Bias

AI bias has far-reaching consequences, disproportionately affecting marginalized groups. Some key impacts include:

i. AI-driven hiring tools may favour certain demographics, reinforcing workplace inequality;
ii. AI-based risk assessment tools may unfairly classify individuals based on race, leading to harsher sentencing for minority groups;
iii. AI-powered credit scoring systems may systematically disadvantage women and lower-income individuals;

[14] Lee-St. John TJ and others, 'Towards Artificial Intelligence-Based Disease Prediction Algorithms That Comprehensively Leverage and Continuously Learn from Real-World Clinical Tabular Data Systems' (2024) 3 PLOS Digital Health

iv. AI diagnostic tools may underperform on underrepresented populations, leading to misdiagnoses and inadequate treatment.

Given these significant impacts, addressing AI bias is critical to ensuring fairness, accountability, and trust in AI systems.[15] The following case studies illustrate real-world instances where AI bias led to discriminatory outcomes.

C. Case Studies: Real-World Examples of AI Bias

The following case studies illustrate real-world instances where AI bias led to discriminatory outcomes.

a) Amazon's Hiring Algorithm: Gender Bias[16]

i. Background

In an attempt to streamline its hiring process, Amazon developed an AI-powered recruitment tool that screened resumes and identified top candidates. The system was trained on ten years of past hiring data, primarily consisting of resumes submitted for technical positions.

ii. Bias and Discrimination

Since the tech industry has historically been male-dominated, the AI system learned to favor male applicants. It systematically downgraded resumes that contained words associated with women, such as "women's chess club captain" or "graduated from a women's college." Even if female candidates had comparable or superior qualifications, the algorithm assigned them lower scores.

iii. Outcome

Amazon ultimately abandoned the AI hiring tool after discovering its bias. This case highlights how historical biases in training data can translate into AI-driven discrimination, reinforcing gender disparities in employment.

[15] Ferrara E, 'Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies' (2024) 6 Sci 3 <https://www.mdpi.com/2413-4155/6/1/3> accessed 5 January 2025

[16] Jeffrey Dastin, 'Insight - Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women' *Reuters* (11 October 2018) <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/> accessed 8 January 2025.

b) Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) Algorithm: Racial Disparities in Sentencing Predictions[17]

i. Background

The COMPAS algorithm is used in the U.S. criminal justice system to assess the likelihood of defendants reoffending. Judges use COMPAS risk scores to make sentencing and parole decisions.

ii. Bias and Discrimination

A 2016 investigation by ProPublica found that COMPAS disproportionately classified black defendants as high-risk offenders, even when their criminal histories were similar to white defendants. In contrast, white defendants were more likely to be classified as low risk, even if they had a higher likelihood of reoffending.

iii. Outcome

The racial bias in COMPAS raised serious concerns about fairness in AI-driven sentencing. Critics argue that using AI in criminal justice without transparency and oversight exacerbates systemic racial disparities rather than mitigating them.

c) Apple Card: Gender Discrimination in Credit Limits[18]

i. Background

Apple launched its AI-powered Apple Card in partnership with Goldman Sachs, promising a data-driven credit approval system. However, shortly after its release, reports emerged that women were receiving significantly lower credit limits than men, even when they had similar financial backgrounds.

ii. Bias and Discrimination

Several high-profile cases revealed gender discrimination in Apple Card's credit scoring algorithm. For instance, a well-known tech entrepreneur reported that his

[17] Saachi Dhingra, 'COMPAS - Correctional Offender Management Profiling for Alternative Sanctions: A Global and Comparative Perspective' (Record Of Law, 30 November 2024) <https://recordoflaw.in/compas-correctional-offender-management-profiling-for-alternative-sanctions-a-global-and-comparative-perspective/> accessed 12 January 2025

[18] Reuters, 'Apple Card Issuer Investigated after Claims of Sexist Credit Checks' The Guardian (10 November 2019) <https://www.theguardian.com/technology/2019/nov/10/apple-card-issuer-investigated-after-claims-of-sexist-credit-checks> accessed 12 January 2025

wife, despite having a higher credit score, was offered a credit limit 20 times lower than his. The AI model likely relied on biased historical data that systematically favoured male applicants over female applicants.

iii. Outcome

Following public scrutiny, Goldman Sachs promised to investigate its credit evaluation process. This case underscores the risks of AI bias in financial services, where discriminatory algorithms can restrict economic opportunities for marginalized groups.[19]

AI bias is a complex and multifaceted issue that stems from biased data, algorithmic flaws, and human intervention. As AI continues to shape decision-making across industries, its potential to reinforce discrimination must be addressed through robust legal and policy interventions.

The case studies of Amazon's hiring algorithm, COMPAS sentencing predictions, and Apple Card credit limits illustrate how unchecked AI bias can lead to real-world harm. Ensuring fairness in AI requires ongoing efforts in bias detection, transparency, and ethical AI governance to create systems that promote equality and justice for all.

III. LEGAL FRAMEWORKS ADDRESSING AI BIAS

The rapid adoption of Artificial Intelligence (AI) across critical domains such as healthcare, finance, law enforcement, and employment has necessitated robust legal frameworks to mitigate AI bias and discrimination. While AI has the potential to enhance efficiency and decision-making, biased algorithms can reinforce societal inequities, leading to unlawful discrimination. Recognizing these risks, governments and international organizations have begun developing legal and regulatory mechanisms to ensure AI operates within ethical and justifiable limits.[20]

This chapter explores the existing international and national legal frameworks designed to address AI bias, focusing on major regulatory initiatives such as the EU Artificial Intelligence Act ("EU AI Act"), General Data Protection Regulation ("GDPR"), Organisation for Economic Co-operation and Development ("OECD") AI Principles, and India's evolving AI regulatory

[19] Tallberg J and others, 'The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research' (2023) 25 International Studies Review viad040 <https://academic.oup.com/isr/article/doi/10.1093/isr/viad040/7259354> accessed 18 January 2025

[20] European Commission, 'Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' COM (2021) 206 final

landscape, including the Information Technology ("IT") Act and Digital Personal Data Protection Bill ("DPDPB"). Additionally, this chapter examines the challenges inherent in current legal frameworks, highlighting enforcement gaps, regulatory inconsistencies, and the difficulty of keeping pace with AI advancements.

A. International and National Legal Approaches

Given AI's global impact, multiple regulatory frameworks have emerged at both international and national levels. Some governments have established direct AI laws, while others rely on data protection regulations and anti-discrimination laws to govern AI applications.

a) European Union (EU) AI Regulations

The European Union has taken a proactive approach to AI regulation, focusing on risk-based categorization, transparency, and accountability. Two key legislative efforts addressing AI bias in the EU are:

i. EU Artificial Intelligence Act

The EU Artificial Intelligence Act (EU AI Act)[21] is one of the most comprehensive AI regulatory proposals globally. Introduced by the European Commission in 2021, the Act seeks to regulate AI systems based on their potential risk to human rights and safety. The risk-based approach in AI regulation means that AI systems are classified based on their potential harm to people and society. The higher the risk, the stricter the regulations. It's like rating AI based on how dangerous it could be and applying rules accordingly. AI risk based systems are classified into four categories[22]:

1) Unacceptable Risk AI (banned): AI applications that violate fundamental rights, such as social credit scoring or real-time biometric surveillance.
2) High-Risk AI (strictly regulated): AI used in hiring, credit scoring, healthcare, and law enforcement, where bias could lead to serious harm.
3) Limited Risk AI (transparency requirements): AI-based chatbots, which must disclose that they are not human.
4) Minimal Risk AI (unregulated): Basic AI applications such as spam filters.

The EU AI Act is expected to set a global precedent for AI regulation, influencing AI

[21] European Parliament and Council, Artificial Intelligence Act (Regulation (EU) 2024/1689)
[22] 'High-Level Summary of the AI Act | EU Artificial Intelligence Act' <https://artificialintelligenceact.eu/high-level-summary/> accessed 8th February 2025

governance models in other countries.

ii. General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR) [23] is the EU's cornerstone legislation on data protection and privacy. While not specifically designed for AI, the GDPR plays a crucial role in regulating AI bias by ensuring transparency and fairness in automated decision-making. Key provisions relevant to AI bias include – i) Article 22: Grants individuals the right to contest AI-driven decisions that have legal or significant impacts on them (e.g., AI-based credit scoring or hiring decisions)[24]; ii) Right to Explanation: Requires organizations to provide meaningful explanations of AI-driven decisions, ensuring transparency; iii) Data Minimization and Fairness: Organizations must ensure that AI training data is accurate, unbiased, and legally obtained, reducing the risk of discriminatory AI models; iv) GDPR's emphasis on data protection, fairness, and accountability forms a strong legal foundation for addressing AI bias.

b) Organisation for Economic Co-operation and Development AI Principles

The Organisation for Economic Co-operation and Development (OECD) has established a global framework for responsible AI governance, influencing policy approaches in several countries.[25] The OECD AI Principles (2019) outline five key pillars[26]:

1) Inclusive Growth & Fairness: AI should benefit all individuals and reduce discrimination rather than reinforcing it.
2) Transparency & Explainability: AI decision-making must be understandable and auditable to ensure accountability.
3) Robustness & Safety: AI should operate reliably and prevent harmful biases.
4) Human-Centered AI: AI must respect human rights and ensure human oversight in critical decision-making.
5) Accountability: Governments and companies must be legally responsible for AI-driven harms.
6) While not legally binding, the OECD AI Principles provide guidance for governments seeking to regulate AI bias while fostering innovation.

[23] European Parliament and Council, General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679)
[24] Kettas JC Muhammed Demircan, Kalyna, 'Europe: The EU AI Act's Relationship with Data Protection Law: Key Takeaways' (*Privacy Matters*, 25 April 2024) <https://privacymatters.dlapiper.com/2024/04/europe-the-eu-ai-acts-relationship-with-data-protection-law-key-takeaways/> accessed 2 February 2025
[25] OECD, 'OECD Principles on Artificial Intelligence' (2019)
[26] 'AI Principles Overview' <https://oecd.ai/en/principles> accessed 2 February 2025

c) India's Evolving AI Legal Framework

India, as one of the fastest-growing AI markets, is in the process of developing a regulatory framework for AI governance. While India currently lacks a dedicated AI law, existing regulations such as the IT Act and proposed Digital Personal Data Protection Bill (DPDPB) provide partial legal coverage for AI-related discrimination.

i. Information Technology (IT) Act, 2000

The IT Act[27] governs electronic transactions, cybercrime, and digital data security in India. Though not AI-specific, it applies to AI-related violations, such as:

1) Cybersecurity and AI-based fraud detection.
2) Data protection and AI-driven privacy concerns.
3) Liability of AI service providers for biased outcomes.

However, the IT Act does not directly regulate AI bias, highlighting the need for dedicated AI legislation.

ii. Digital Personal Data Protection Bill (DPDPB), 2023

The Digital Personal Data Protection Bill (DPDPB) [28] aims to modernize India's data protection regime, with provisions addressing AI-driven discrimination. Key provisions:

1) Fair AI Practices: Ensures that AI systems processing personal data do so in a lawful, fair, and transparent manner.
2) Right to Explanation: Grants individuals the right to understand how AI-based decisions affect them, promoting AI accountability.
3) Bias Prevention Measures: Organizations must ensure that AI training datasets do not intentionally or unintentionally reinforce discrimination.

While DPDPB is a step towards AI regulation, India still lacks comprehensive AI laws to govern bias, fairness, and algorithmic accountability.

B. Challenges in Existing Legal Frameworks

Despite growing regulatory efforts, several challenges persist in enforcing AI bias regulations effectively:

i. Lack of Global Standardization

[27] Government of India, Information Technology Act, 2000 (IT Act)
[28] Government of India, *Digital Personal Data Protection Bill, 2023* (DPDPB)

AI regulations vary significantly across jurisdictions, creating regulatory fragmentation.[29] The EU AI Act imposes strict AI risk assessments, whereas India and the U.S. lack equivalent laws. Companies operating internationally must comply with multiple, sometimes conflicting, AI regulations.

ii. AI's Rapid Evolution Outpacing Regulation

AI technology evolves faster than legal frameworks can adapt. Deep learning and generative AI introduce new forms of bias that existing laws do not explicitly address.[30] Regulators struggle to keep pace with new AI applications in finance, healthcare, and surveillance.

iii. Difficulty in AI Auditing and Transparency

Many AI models, especially neural networks, function as "black boxes," making it difficult to identify bias or explain decisions. The lack of AI audit standards hinders enforcement of fairness laws. Many AI developers resist disclosing proprietary AI decision-making processes, citing trade secrets.[31]

iv. Weak Enforcement Mechanisms

Even in regions with AI laws, enforcement remains weak: GDPR's Right to Explanation is often ignored by companies using AI in hiring and finance. Bias audits are not legally required in many countries, making AI discrimination hard to detect.

While legal frameworks such as the EU AI Act, GDPR, OECD AI Principles, and India's DPDPB provide a foundation for addressing AI bias, significant challenges remain. Regulatory inconsistencies, lack of enforcement, and AI's rapid evolution hinder effective governance. Moving forward, governments must implement stronger AI bias auditing mechanisms, global regulatory cooperation, and mandatory transparency requirements to ensure fairness and accountability in AI-driven decision-making.[32]

[29] 'AI Watch: Global Regulatory Tracker - European Union | White & Case LLP' (2025) <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-european-union> accessed 28 February 2025
[30] Cohen JE, The Law of Robots: The Ethical and Legal Implications of AI in Law (Cambridge University Press 2022)
[31] Zarsky T, 'Transparency and the Role of AI in the Legal System' (2020) Stanford Law Review Online 72, 1-15
[32] Marr B, 'What is AI Accountability?' (Forbes, 5 June 2022)

IV. AI POLICY SOLUTIONS TO AI BIAS

Artificial Intelligence (AI) has become an indispensable tool in shaping decision-making processes across various sectors, offering unprecedented efficiency, speed, and predictive capabilities. However, the growing recognition of AI bias and discrimination poses significant challenges to social equality, human rights, and justice. While legal frameworks aim to regulate AI applications, they often fall short in addressing the dynamic and complex nature of AI bias.[33] Therefore, beyond legal mechanisms, policy solutions play a critical role in mitigating bias, enhancing fairness, and ensuring that AI technologies promote inclusivity and social justice. This chapter explores policy-based solutions to AI bias through two key avenues: strengthening regulatory and compliance mechanisms and fostering public-private sector collaboration. Both approaches emphasize the importance of proactive governance, ethical AI development, and the need for interdisciplinary cooperation to create a more equitable AI landscape.[34]

A. Strengthening Regulatory and Compliance Mechanisms

Strengthening regulatory and compliance mechanisms is a fundamental step in addressing AI bias. Regulatory policies must not only prohibit discrimination but also provide practical tools for preventing and detecting biased outcomes in AI systems.[35] This section outlines various policy interventions to improve accountability, transparency, and fairness in AI technologies.

a) Mandatory Bias Auditing and Impact Assessments

Governments should implement mandatory AI bias audits and algorithmic impact assessments before AI systems are deployed in high-risk sectors such as:

1) Hiring and recruitment
2) Credit scoring and financial services
3) Criminal justice
4) Healthcare

These audits would involve evaluating training datasets, testing the algorithm's outcomes for disparate impacts, and identifying any discriminatory patterns before deployment. The audits should be conducted by independent third-party organizations to ensure impartiality. Example

[33] Kim B-P, 'Privacy-Enhancing Technologies for AI and the Challenges of Legal Frameworks' (2023) 20 Korean Journal of Law and Economics 113
[34] Wachter S, Mittelstadt B and Russell C, 'Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI' (2021) Computer Law & Security Review 41, 1-14.
[35] 'AI Bias Examples' (2023) IBM <https://www.ibm.com/think/topics/shedding-light-on-ai-bias-with-real-world-examples> accessed 8 February 2025

Policy Model: The EU AI Act mandates conformity assessments and bias mitigation measures for high-risk AI systems, setting a global precedent for bias auditing.

b) Algorithmic Transparency and Explainability Standards

One of the most significant challenges in regulating AI bias is the opacity of AI algorithms, particularly deep learning models. Policymakers must mandate that AI systems used in critical decision-making processes are:

i. *Interpretable:* AI models must provide clear and understandable explanations of how decisions are made.
ii. *Auditable:* Developers must create algorithmic logs to allow regulators and users to trace how decisions were reached.
iii. *Fair by Design:* Algorithms must be explicitly designed to avoid disparate impacts on protected groups.

Transparency requirements can be enforced through Algorithmic Transparency Certificates that certify AI systems as compliant with fairness and non-discrimination principles.

c) Ethical Design and Inclusive Datasets

Governments should promote the development of ethical AI systems by integrating diversity and inclusivity standards into the AI design process. This involves:

i. Requiring developers to train AI models on diverse datasets that accurately represent different genders, races, socioeconomic groups, and marginalized communities.
ii. Establishing Diversity Data Guidelines that mandate the inclusion of minority populations in AI datasets.
iii. Encouraging the use of Fairness-Aware Algorithms (FAA)[36], which automatically adjust AI outputs to ensure equal treatment across demographic groups.

d) Certification and Accountability Frameworks

To improve public trust in AI systems, governments should create AI Certification Programs that verify whether AI systems meet fairness, transparency, and accountability standards.[37] These certifications would work similarly to International Organization to Standardization

[36] Cheng M and others, 'Social Norm Bias: Residual Harms of Fairness-Aware Algorithms' (2023) 37 Data Mining and Knowledge Discovery
[37] European Commission, 'Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' COM (2021) 206 final

(ISO) standards in other industries.

i. Possible Certification Models:[38]

Certification Type	Description	Mandatory/Voluntary
AI Fairness Certification	Confirms that an AI model meets fairness and bias detection standards	Mandatory for High-Risk AI Systems
Transparency and Explainability Certification	Certifies that AI models provide transparent and explainable decisions	Mandatory for Government and Public Services AI
Privacy and Data Protection Certification	Certifies compliance with GDPR or equivalent data protection laws	Mandatory for AI systems processing personal data

e) Whistleblower Protection and Algorithmic Redress Mechanisms

To empower individuals affected by biased AI systems, governments should establish Algorithmic Redress Mechanisms that allow people to challenge AI-driven decisions. Additionally, whistleblower protection policies should encourage employees and researchers to report unethical AI practices without fear of retaliation. Example: The GDPR's Article 22 already grants individuals the right to contest automated decisions that significantly affect them. This model could be expanded globally.[39]

B. Public-Private Sector Collaboration

Addressing AI bias requires a collaborative effort between governments, technology companies, civil society organizations, and academic institutions.[40] The public-private partnership model can foster innovation while ensuring that AI systems are developed and deployed responsibly.

a) AI Ethics Committees and Advisory Boards

Governments should mandate that companies working on high-risk AI applications establish

[38] Marr B, 'What is AI Accountability?' (Forbes, 5 June 2022)
[39] OECD, 'OECD Principles on Artificial Intelligence' (2019) https://www.oecd.org/going-digital/ai/principles/ accessed 19 February 2025
[40] Zarsky T, 'Transparency and the Role of AI in the Legal System' (2020) Stanford Law Review Online 72, 1-15.

AI Ethics Committees or Algorithmic Fairness Boards. These committees would include diverse stakeholders, including; AI Developers, Human Rights Experts, Data Scientists, Legal Professionals, Civil Society Representatives etc.

The committees would be responsible for:

i. Reviewing AI models for fairness and transparency.
ii. Conducting regular bias audits.
iii. Issuing public reports on AI system performance.

b) *Open Data and Algorithmic Transparency Initiatives*

Public-private collaborations can promote open datasets and algorithmic transparency by: Creating Public Data Trusts where companies and governments share anonymized datasets for training fairer AI models. Developing Open AI Auditing Platforms that allow civil society organizations to independently test AI models for discrimination.[41] Encouraging algorithmic transparency standards through corporate social responsibility (CSR) policies. Example: Google's AI Principles include commitments to fairness and transparency, setting a model for other companies.[42]

c) *AI Ethics Training and Capacity Building*

Governments and companies should jointly invest in AI ethics training programs to raise awareness among developers, policymakers, and regulators. These programs should cover-ethical AI design principles, bias detection methods, legal obligations under data protection laws, inclusive dataset creation etc. AI capacity-building programs would help bridge the gap between technical experts and legal professionals, fostering interdisciplinary collaboration.

d) Funding for Inclusive AI Innovation

Public-private partnerships should promote Inclusive AI Innovation Funds to support startups and research initiatives developing bias-mitigation technologies and fair AI systems.

e) Global AI Governance Networks

Governments and corporations should collaborate on the creation of Global AI Governance

[41] 'Beyond the Algorithm: OpenAI's Commitment to Responsible AI Development – Quantilus Innovation' <https://quantilus.com/article/beyond-the-algorithm-openais-commitment-to-responsible-ai-development/> accessed 17 February 2025

[42] Banawan M and others, 'Transformative Approach to Fairness and Transparency in Classroom Participation Assessment', *Proceedings of the Eleventh ACM Conference on Learning* (ACM 2024)

Networks to share best practices, develop international AI fairness standards, and create unified regulatory frameworks. Example: The Global Partnership on AI (GPAI) is an emerging initiative that encourages multi-stakeholder cooperation on AI governance.[43]

AI bias presents profound challenges to social justice, equality, and human rights. While legal frameworks play a critical role in regulating AI, they must be complemented by proactive policy solutions that promote fairness, transparency, and inclusivity. Strengthening regulatory and compliance mechanisms through bias auditing, algorithmic transparency, and ethical design standards can create a foundation for responsible AI deployment. At the same time, public-private sector collaboration will accelerate the development of innovative solutions that mitigate AI bias and ensure that AI technologies serve the interests of all members of society. Moving forward, policymakers must adopt a holistic governance model that combines legal, technical, and social interventions to build equitable and accountable AI systems. Only through dynamic and collaborative approaches can AI fulfil its transformative potential without perpetuating discrimination.[44]

V. CONCLUSION

A. Summary of Findings

The integration of AI in decision-making processes across various sectors has introduced efficiency, speed, and predictive capabilities, but it has also highlighted significant concerns regarding bias and discrimination. This paper explored the fundamental causes of AI bias, its legal and ethical implications, and the various international and national regulatory approaches to addressing these concerns.[45] AI bias arises from multiple sources, including historical discrimination in training data, algorithmic design flaws, and human intervention. Case studies, such as Amazon's biased hiring algorithm, the COMPAS risk assessment tool, and Apple Card's gender discrimination, demonstrate how unchecked AI systems can reinforce societal inequalities. These examples highlight the urgent need for stronger legal and policy interventions to ensure fairness and accountability in AI systems.[46] At the legal level, several

[43] Helsinki Process on Globalization and Democracy and Finnish Institute of International Affairs (eds), *Multi-Stakeholder Cooperation in Global Governance* (Ministry of Foreign Affairs 2008)
[44] Pérez C, 'Legal Accountability for AI Decision-Making: A Review of Current Frameworks' (2021) Journal of Technology and Law 15(1), 25-40
[45] Cohen JE, The Law of Robots: The Ethical and Legal Implications of AI in Law (Cambridge University Press 2022).
[46] O'Neil C, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (Crown Publishing Group 2023).

frameworks have been introduced to mitigate AI bias, including the EU AI Act, GDPR, and OECD AI Principles, as well as India's IT Act and Digital Personal Data Protection Bill. While these frameworks provide a foundation for addressing AI-related discrimination, they face challenges such as regulatory inconsistencies, lack of enforcement mechanisms, and the rapid evolution of AI technology.[47] To address these gaps, this paper outlined policy solutions, including mandatory AI bias audits, transparency requirements, ethical AI design principles, and interdisciplinary collaboration between the public and private sectors. Strengthening compliance mechanisms and fostering public-private partnerships are crucial steps toward ensuring that AI operates fairly, ethically, and within legal boundaries.[48] Ultimately, the findings emphasize that AI bias is not merely a technical issue—it is a societal challenge requiring a multifaceted approach. Without appropriate regulatory oversight, ethical considerations, and continuous monitoring, AI has the potential to exacerbate systemic inequalities rather than mitigate them.

B. Recommendations

Based on the findings of this study, the following recommendations are proposed to address AI bias effectively:

a) Strengthening Legal and Regulatory Oversight

Governments should establish clear and enforceable AI regulations that mandate bias detection, impact assessments, and accountability measures for AI developers and deployers.[49] AI regulations must be dynamic and adaptable to keep pace with rapid technological advancements.

b) Implementing Mandatory AI Bias Audits

AI models used in high-risk decision-making sectors (e.g., hiring, credit scoring, law enforcement) should be subject to mandatory bias audits before deployment. AI audits should be conducted by independent third-party organizations to ensure neutrality and credibility.

[47] Pérez C, 'Legal Accountability for AI Decision-Making: A Review of Current Frameworks' (2021) Journal of Technology and Law 15(1), 25-40

[48] Publishing O, Fostering Public-Private Partnership for Innovation in Russia (Organisation for Economic Co-operation and Development 2005)

[49] Sikombe M, 'Regulating the Future: AI and Governance' (Artificial intelligence, 10 May 2024) <https://nationalcentreforai.jiscinvolve.org/wp/2024/05/10/regulating-the-future-ai-and-governance/> accessed 15th February 2025

c) *Promoting Algorithmic Transparency and Explainability*

AI developers should be required to document and disclose how their algorithms make decisions, particularly in high-risk applications.[50] Explainability standards should be integrated into AI governance frameworks to prevent black-box decision-making.

d) Enhancing Data Diversity and Fairness

AI models should be trained on diverse and representative datasets to prevent systematic discrimination against marginalized groups. Policymakers should introduce data diversity standards to guide AI development.

e) Encouraging Public-Private Collaboration

Governments, private companies, and civil society organizations should work together to create AI fairness benchmarks and responsible AI deployment frameworks. AI Ethics Committees should be established in organizations deploying AI in critical decision-making processes.

f) *Creating AI Redress Mechanisms and Whistleblower Protections*

Individuals affected by biased AI decisions should have legal avenues to challenge AI-driven outcomes. Strong whistleblower protection laws should be introduced to encourage AI developers and employees to report unethical AI practices.

g) *Supporting AI Ethics Education and Capacity Building*

AI ethics and fairness should be incorporated into educational curricula for data scientists, policymakers, and business leaders. Governments should invest in training programs to equip regulators with the technical knowledge needed to assess AI fairness and compliance.

h) *Developing Global AI Governance Standards*

International cooperation is essential to establish global AI regulatory frameworks that ensure consistency in AI fairness laws across different jurisdictions. Countries should collaborate on cross-border AI research initiatives to create fair and unbiased AI models.

[50] 'What Does Transparency Really Mean in the Context of AI Governance?' (*OCEG*, 8 November 2024) <https://www.oceg.org/what-does-transparency-really-mean-in-the-context-of-ai-governance/> accessed 26 February 2025