Page 5212 - 5221 **DOI:** *https://doij.org/10.10000/IJLMH.119611*

INTERNATIONAL JOURNAL OF LAW MANAGEMENT & HUMANITIES

[ISSN 2581-5369]

Volume 8 | Issue 2

2025

Algorithm bias and Discrimination bias in AI-Assisted Legal Processes

NEHA BHARTI[1] AND MOHD IMRAN[2]

ABSTRACT

The integration of Artificial Intelligence into the legal system has brought about both opportunities for enhanced efficiency and impartiality as well as new challenges, particularly in the form of algorithm and discrimination biases. Algorithm bias stems from inherent errors in AI model creation, training, or implementation, often mirroring the existing societal inequities found in training data or decision-making frameworks. Conversely, discrimination bias leads to the unfair treatment of individuals or groups due to these algorithmic biases, resulting in unjust or prejudiced outcomes within legal contexts. In the realm of legal processes, biases can surface in tools employed for predictive policing, sentence suggestions, and risk evaluations, potentially reinforcing societal inequalities. Tackling these issues is crucial to ensure that AI-assisted legal processes enhance justice rather than perpetual inequalities, thereby aligning technological progress with ethical and legal norms. This paper highlights the issues in AI-assisted legal system algorithm bias and discrimination bias, and also states the feedback loops. This study explores the implications of and mitigation strategies for algorithm and discrimination biases in an AI-driven Legal System.

***Keywords**: Artificial Intelligence, legal system, bias, discrimination.*

I. INTRODUCTION

An important development in the subject of law is the incorporation of artificial intelligence into legal systems, which marks a move away from antiquated procedures and toward more cutting-edge technological approaches. To automate legal reasoning, expert systems were developed in the late 20th century. These systems were designed to capture the knowledge and thought processes of legal experts and provide automated decision-making tools.[3]Artificial intelligence has its origin back from 1950, when an English polymath, Alan Turing, devised a

[1] Author is a Research Scholar at School of Law & Constitutional Studies, Shobhit Institute of Engineering & Technology (Deemed to be University), Meerut, India.

[2] Author is a Professor and Director Academics, University Institute of Legal Studies, Chandigarh University, India.

[3] Rachid Ejjami, *AI-Driven Justice: Evaluating the Impact of Artificial Intelligence on Legal System*, 6 IJFMR, 1 (2024),https://www.researchgate.net/profile/Rachid-Ejjami/publication/381926291_AI-Driven_Justice_Evaluating_the_Impact_of_Artificial_Intelligence_on_Legal_Systems/links/66adeda051aa0775f264db66/AI-Driven-Justice-Evaluating-the-Impact-of-Artificial-Intelligence-on-Legal-Systems.pdf.

test to see if a machine could be making human cognitive function to identify patterns. It becomes popular in 1956, McCarthy invited people from different fields to discuss the importance of computers that consume data and mimic human behavior. These discussions and data sharing have created the possibility of developing new advancements in high-computing systems across the globe.[4]Contemporary AI technologies in the legal field are revolutionizing the industry by expediting voluminous tasks with unparalleled speed and efficiency and introducing sophisticated analytical capabilities. These advancements are poised to revolutionize every aspect of legal operations from the ground up, questioning conventional methodologies and fundamentally altering legal practices. As AI continues to evolve, it has the capacity to revolutionize the legal system, paving the way for more efficient, precise, and accessible legal frameworks in the years to come. Modern technology, especially artificial intelligence (AI), is perceived and used, and is greatly impacted by the interaction between social factors such as gender and technical elements such as algorithms.[5]

(A) Algorithm Bias

Algorithms are fundamentally a precise and comprehensive set of instructions, crafted to generate solutions to the problems they were created to address. They offer an automated approach to performing calculations and computational operations. Basic algorithms operate using a linear approach to problem solving, whereas more intricate algorithms incorporate conditional functions that enable them to perform more complex tasks, now referred to as *automated decision-making* (ADM), aimed at achieving automation for tasks that previously necessitated oversight or additional human resources. Currently, numerous Artificial Intelligence Software (AIs) have included Automated Decision-Making systems (ADMs) to enhance their operational capabilities in judicial systems worldwide.[6]

There are a number of factors that can lead to algorithmic bias, including human prejudice, which is normally addressed by the courts, cognitive bias, which is often addressed by behavioural scientists, and design and data errors, which are typically addressed by computer engineers.[7]Biases in algorithmic systems can lead to discrimination and exacerbate it due to their large scale and feedback loops. However, distinguishing between prejudice and

[4] Dr. Varsha P. S, *How can we manage biases in artificial intelligence systems – A systematic literature review*, 3 IJIMDI, 100165 (2023), https://www.sciencedirect.com/science/article/pii/S2667096823000125

[5] Anna M. Gorska & Dariusz Jemielniak, *The invisible women: Uncovering gender bias in AI-generated images of professionals*, 23 Feminist Media Studies 4370–4375 (2023).

[6] Kartik Pendharkar, *Algorithmic bias and discrimination: Legal and policy considerations*, SSRN Electronic Journal (2023).

[7] Judge James E. Baker, Laurie Hobart, et al., *An introduction to artificial intelligence for federal Judges*, Goodreads (2023), https://www.goodreads.com/book/show/123689739-an-introduction-to-artificial-intelligence-for-federal-judges (last visited March 20, 2025).

discrimination is crucial. The United Nations Institute for Disarmament Research has suggested several categories and sources of algorithmic bias. Statistical bias, moral bias, training data bias, unsuitable focus, inappropriate deployment, interpretation bias, unintentional human prejudice, and purposeful bias are the eight types of potential AI bias highlighted in the study.[8]

Statistical bias may arise when an algorithm's anticipated results diverge from statistical norms, such as the actual occurrence rates of real-world outcomes. This discrepancy may be due to poor statistical modelling or inadequate data. When the results of an algorithm deviate from the recognized standards (legal, ethical, societal, etc.), *moral bias* is present. Data from the existing prison population can be used by a predictive crime algorithm to "predict" recidivism rates. The algorithm would probably provide biased results by overtly or unintentionally weighing "race" or its proxies, and possibly even within the black box, given the high number of persons of color who were imprisoned. AI does not think or comprehend the world way humans do, and unless given other instructions, its output may demonstrate a lack of awareness of the standards outlined in the due process and equal protection sections.[9]Contrary to humans, AI learns from experience in *Training data bias;* however, this experience is solely dependent on data, which is frequently hand-picked by a human developer. These data errors or misrepresentations have the potential to reinforce prejudices by incorporating them into codes using algorithms. *Inappropriate focus* occurs when the training data of an algorithm are unsuitable. Facial recognition data disparities between men and women may increase false positives, such as innocent female travellers being selected for extra screening or questioning at airport. However, failing to identify known subjects or risks, such as a missing person or an Amber Alert kidnap victim on CTV camera feeds, is dangerous.[10] When a system is put to use in a situation where it is not intended, tested, or validated, it is called *inappropriate deployment*. For example, an autonomous vehicle that has been trained to drive in the US might not be able to drive in the UK on the left. While a human adjusts to such a shift, an algorithm for an autonomous vehicle would require further training. When using AI applications in court, judges and litigators want to confirm that they are made for the particular purpose for which they are being used. When the results of an algorithm are unclear or misinterpreted by those utilizing the technology, this is known as *interpretation bias*, which denotes the inadvertent incorporation of human preferences, stereotypes, values, fears, or knowledge into an application. Examine an algorithm designed to forecast risks. *In Intentional bias*, researchers, practitioners, and policymakers may

[8] *Ibid at page no. 32.*
[9] *Ibid at page no. 33.*
[10] *Supra note 7 at page no. 35.*

 [ISSN 2581-5369]

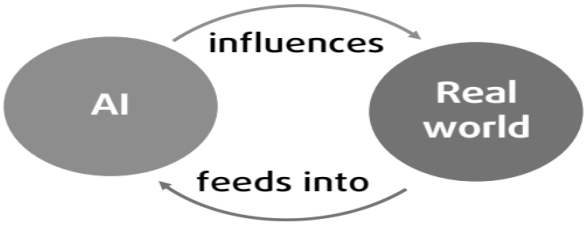
employ AI facial recognition technologies or algorithmic predictive policing to identify marginalized or at-risk populations. Algorithms can be developed to recognize and pick specific real and perceived social attributes related to "race," gender, sexuality, country of origin, religion, handicap, and others. Facial recognition technology may identify and monitor specific ethnic groups, as seen by the identification of "Uighur characteristics" in China. The profiling of Uighurs, specifically through the physical traits linked to them by Chinese state security services, raises the question of whether the intentional use of social identity descriptors is a suitable search criterion. The response may partially hinge on intent, the definition of "search parameter," and the degree of human oversight.[11]

The United Nations Institute for Disarmament Research's explanation of many forms of bias makes it abundantly evident that algorithmic bias can be minimized when AI is employed in legal proceedings alongside human assistance. Courts may need to assess whether AI applications blur the distinction between individual suspect identification and group profiling. AI applications such as facial recognition databases have the potential to broaden their scope. Unlike traditional human law enforcement investigations, group profiling might lead to the suspicion and scrutiny of innocent individuals.

II. FEEDBACK LOOPS; HOW ALGORITHM EFFECT ALGORITHM

A feedback loop emerges when a system's predictions affect the data used to update the system. This phenomenon occurs because algorithms influence other algorithms with their recommendations and forecasts shaping real-world conditions. An example is when an algorithm's crime predictions alter police officer behavior, subsequently affecting crime detection. Newly detected crimes were then incorporated into the system. Such feedback loops are prevalent, and many machine learning systems inherently include them.[12]A simplified representation of a feedback loop is depicted in Figure 1.

FIGURE 1: SIMPLISTIC ILLUSTRATION OF A FEEDBACK LOOP



Source: FRA, 2022 13

[11] *Supra note 6 at page no. 36.*
[12] Report on *Bias in Algorithm Artificial Intelligence and Discrimination*, European Union Agency for Fundamental Rights,2022 https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf
[13] *Supra note 10.*

(A) Discrimination Bias

When artificial intelligence systems used in judicial proceedings generate results that unfairly penalize particular people or groups, this is known as discrimination bias. This is frequently the result of structural problems in system development, training, or implementation. In AI-driven legal systems, where decisions can significantly affect people's lives, rights, and freedom, this prejudice is especially worrisome. Although research[14] indicating gender bias in textual General Artificial Intelligence (GAI) systems like ChatGPT has led to proposed mitigation techniques by regulators, analogous studies in visual GAI systems have been considerably more restricted. Research on gender bias in visual GAI identified its prevalence.[15]Gorska and Jemielniak (2023) conducted an experiment involving various visual GAI systems, instructing them to generate illustrations of individuals participating in law, medicine, engineering, and scientific research. It was reported that men constituted 76% of the photos, whereas women comprised only 8%.[16]

a. Origin of discrimination bias

- **Biased Training Data**

Artificial intelligence systems learn from historical information, which may reflect societal and institutional prejudices. For instance, crime statistics that disproportionately represent certain communities could result in unfair outcomes for these groups. In a system, if the date entered is biased, then the output for the same date is biased. Numerous studies have shown that human bias is a reaction to technological advancement, whereas cognitive biases affect every facet of human decision-making through artificial intelligence and robotic creation.[17]

- **Surrogate Characteristics**

AI may depend on features that correlate with protected attributes (such as postal codes associated with race), resulting in indirect discrimination. Even when explicit characteristics, such as race or gender, are omitted, these surrogates can introduce unfairness. For example, Amazon discriminated against female applicants when using AI technology to measure and rate job applicants.[18]

---

[14] Anna M. Gorska & Dariusz Jemielniak, *The invisible women: Uncovering gender bias in AI-generated images of professionals*, 23 Feminist Media Studies 4370–4375 (2023).
[15] Larry G. Locke & Grace Hodgdon, *Gender bias in visual generative artificial intelligence systems and the socialization of ai*, AI & SOCIETY (2024).
[16] *Supra note 6.*
[17] K. Letheren, R. Russell-Bennett, L. Whittaker, *Black, white or grey magic? Our future with artificial intelligence*, 36 JMM, 216 (2020)
[18] J. Weissman, *Amazon created a hiring tool using AI it immediately started discriminating against women*, 2018 https://slate.com/business/2018/10/amazon-artificial-intelligence-hiring-discrimination-women.html

- **Non-representative Data Sampling**

If certain demographic groups are underrepresented in the training data, the AI may underperform for those populations, leading to unequal outcomes. By forecasting increased crime risk based on skewed data, AI systems may unfairly target underprivileged populations.

- **Biased Speech Detection**

Whether it takes the form of hate speech, harassment, or encouragement of violence or hatred, online hatred has become a common occurrence for many demographic groups worldwide. There are numerous demographic groups in the EU; hate speech, harassment, and encouragement of violence or hatred are all examples of everyday reality that is online hatred. Internet hate speech is becoming a big problem, according to many specialists who assist victims of hate crimes. According to FRA's 2018 survey on antisemitism perceptions, cyberharassment had the highest reported incidence rate of antisemitic harassment. Additionally, according to the FRA's 2012 survey on violence against women, one in 20 EU women reported having been the victim of cyber-harassment, indicating that this type of harassment is pervasive in the EU.[19]

Facebook has significantly enhanced its use of AI to identify hate speeches. In the first quarter of 2022, Facebook recognized 96% of the hate speech posted on the network, compared to only 38% in the same period in 2018. AI systems that assist in flagging content are probably what drove this proportion, also known as the *"pro-active rate,"* after which humans determine what to do, such as post-deletion. Facebook users' reports of hate speech were the source of the remaining hate speech.[20] Automated content moderation of hate speech, especially regarding unlawful hate speech, is not yet viable with AI. Numerous scholars have refused to claim that AI might readily address hate speech problems and have cautioned against the drawbacks of employing AI and algorithms for online content control.

III. CASE STUDIES OF ALGORITHM BIAS AND DISCRIMINATION BIAS

(A) COMPAS algorithm bias: - The Correctional Offender Management Profiling for Alternative Sanctions, or COMPAS, is an artificial intelligence system utilized in correctional facilities to assess the likelihood of offenders reoffending following their release. The data derived from these forecasts or profiles were subsequently submitted to the courts to determine the potential release, parole, or bond amounts for incarcerated

[19] Report on Bias in Algorithm Artificial Intelligence and Discrimination, European Union Agency For Fundamental Rights,2022https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf.
[20] Supra note 6 at page no. 54.

individuals. In their study, an NGO discovered that the COMPAS algorithm exhibited bias, as it was observed that the AI discriminated based on race, categorizing black prisoners as significantly more likely to reoffend than individuals of other races, while white offenders were assessed by the AI as less likely to recidivist. Further analysis revealed that the AI had inadvertently acquired biases based on the training data.[21]

(B) The Amazon Recruitment Engine's Bias towards Women

A further example of discrimination stemming from the training data was identified in the AI utilized by Amazon for staff recruitment, in which the AI's role was to evaluate applications and recommend the most qualified candidates for the position. It was shown that AI has cultivated an inbuilt prejudice against women as a result of its data processing, which was influenced by Amazon's historical employment practices. The given data were found to originate from hiring officers who exhibited bias against women, resulting in the AI adopting the same bias. Consequently, Amazon discontinued AI for recruitment.[22]

IV. A WAY FORWARD MITIGATING STRATEGIES

Upon recognizing the elements that generate and influence Algorithmic Bias, it is clear that global laws and regulations must evolve accordingly in their continuous endeavours to govern AI operations comprehensively. Several organizations worldwide have undertaken significant initiatives in this regard.[23]

One of the notable steps taken by European Union is the General Data Protection Regulation. The GDPR was enforced in May 2018. GDPR does not use the term "*algorithm bias"* directly but there are some provisions which indirectly applies to algorithm bias. European Union (EU)data protection authorities have the power to investigate and fine companies if the algorithm causes harm or discrimination based on gender, sex, place of birth, and so on. Article 22 of the GDPR gives individual rights not to be subject to a decision solely based on automatic processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.[24] This system can use a biased algorithm. Article 5 of the GDP discusses principles related to the processing of personal data, which includes personal data that shall be processed lawfully and in a transparent manner. It also expresses that data that are collected shall be used for legitimate purposes and in public interest, scientific or historical

[21] Kartik Pendharkar, *Algorithmic bias and discrimination: Legal and policy considerations*, SSRN Electronic Journal (2023).
[22] *Ibid.*
[23] *Ibid.*
[24] The General Data Protection Act, 2018. Art. 22.

research purposes, or statical purposes. If a biased algorithm violates the fairness principle, it leads to infringement of the rights of an individual, which can be considered when applying GDPR. Articles 13 to 15 of the GDPR include when information as to personal data is collected from the data subject shall process transparently and ensure fairness in protecting personal data.[25] The data subject can ensure the information of logic involved in automated decision making and also ensure that any decision that is based on the biased data shell is restricted and the data that is free from buyers shall be used.

In 2022, the Equal Employment Opportunity Commission (EEOC) of the United States commenced enforcement of guidelines targeting algorithmic bias in the workplace. A collaborative endeavour among the non-profit, legal, and employment sectors is being developed to provide a model of industry self-regulation for artificial intelligence (AI) in recruitment and hiring procedures akin to the GDPR.[26]

On May 17, 2024, Colorado became an inaugural state to implement legislation that targeted algorithmic bias. The legislation, referred to as the Colorado AI Act, aims to provide *"Consumer Protections in Interactions with Artificial Intelligence Systems."* The statute delineates responsibilities for both *"developers"* and *"deployers"* of AI systems to resolve issues exemplified by *Mobley v. Workday, Inc.*[27] Thirty The term *"deployer"* refers specifically to an entity conducting business in Colorado that implements a *"high-risk artificial intelligence system," defined as "any artificial intelligence system that, upon deployment, influences or significantly contributes to a consequential decision."*[28]

Providers of high-risk AI systems must apply suitable data governance and management strategies to their training, validation, and testing data to ensure that it is pertinent, representative, accurate, and comprehensive. In terms of content moderation algorithms, this approach can aid in addressing the bias that might lead to discriminatory practices. *"AI systems shall be designed and developed in such a way, including with appropriate human–machine interface tools, that they can be effectively overseen by natural persons,"* with the aim of *"preventing or minimising the risks to health, safety or fundamental rights."*[29]

"Overfitting," which happens when a *Machine Learning model* is overly specific to the data it

[25] The General Data Protection Act, 2018. Art. 13-15.
[26] *Supra note 21*.
[27] 23-cv-00770-RFL, 2024 U.S. LEXIS 126336 (N.D. Ca.).
[28] John Meridith, *Regulation by the EEOC and the States of Algorithmic Bias in High-Risk Use Cases*, 80 ABA (2025) HTTPS://WWW.AMERICANBAR.ORG/GROUPS/BUSINESS_LAW/RESOURCES/BUSINESS-LAWYER/2024-2025-WINTER/EEOC-STATES-REGULATION-ALGORITHMIC-BIAS-HIGH-RISK/
[29] *Supra note 6 at page no.53.*

has been trained on and fails to take ambiguities or deviations into account, is one risk associated with machine learning. Making sure that the data the ML system is trained on is distinct from the data it will encounter in use is often how this issue is resolved. Thus, a "generalized" model needs to be adaptable enough to accurately understand the data that it has not come across. The results could be skewed if data separation was not performed.[30]

Regulatory tools can help address the issue of biased algorithms to some extent. First, several nations have enacted moral guidelines and criteria pertaining to scientific advancements in artificial intelligence. Based on the knowledge at hand, ethics aids in making appropriate choices in a given situation. Therefore, analysing ethical values and principles while considering their applications is essential.

V. CONCLUSION

The integration of artificial intelligence into the legal system presents significant advancements in efficiency and accuracy, yet it simultaneously exposes the legal framework to complex issues of algorithmic bias and discrimination. The potential for AI to perpetuate societal prejudices underscores the urgent need for careful implementation and oversight to ensure fairness and equity within judicial proceedings. From the COMPAS algorithm's racial bias to Amazon's recruitment engine discriminating against women, these case studies underscore the urgent need for vigilance and proactive measures. The potential for AI to perpetuate and amplify existing societal biases through feedback loops and biased training data is a stark reminder that technology, no matter how advanced, is not immune to human fallibility. Yet, the future is not bleak. The global legal community is responding with innovative strategies to mitigate these risks. The European Union's GDPR, the United States's EEOC guidelines, and Colorado's pioneering AI Act represent significant strides towards responsible AI governance. These regulatory frameworks, coupled with industry self-regulation and ethical guidelines, offer a multi-pronged approach to addressing algorithmic bias. As we stand at the crossroads of technological innovation and legal tradition, the path forward demands a delicate balance. We must harness the transformative power of AI while steadfastly safeguarding the principles of fairness, transparency, and equal protection under the law. This requires not only robust legislation and ethical guidelines but also a commitment to diverse, representative data sets and ongoing human oversight.

To ensure that there is no bias in the AI-Assistant legal system, they must be people from all geographical locations, including men and women. Gender and diversity will help companies

[30] IBM, What is overfitting? IBM (2024), https://www.ibm.com/think/topics/overfitting (last visited Apr 20, 2025).

 [ISSN 2581-5369]

to encourage their products and facilitate the reduction of bias in the ai-assistant legal process. The government has to self-regulate and ensure policies and regulations for AI-assisted legal processes so that they can be used in a legal manner to remove bias. Use of AI should also protect and promote the human order in society not to harm them.

In conclusion, as AI continues to reshape the legal landscape, our vigilance in addressing algorithmic bias and discrimination will be crucial. The challenges are significant, but so too are the opportunities. With careful navigation, thoughtful regulation, and unwavering commitment to ethical principles, we can harness the power of AI to create a more efficient, accessible, and equitable legal system for all.

*****

 [ISSN 2581-5369]