PORSESH POLICY
RESEARCH INSTITUTE
Fact-based Analysis & Solutions

HOW DO AI BIASES CONTRIBUTE TO DISCRIMINATION AGAINST MINORITIES?

Hussain Rezai

ABOUT THE PUBLISHER

PORSESH POLICY RESEARCH INSTITUTE (PR) is an independent, non-profit, community-centered research institute based in Washington, dedicated to producing fact-based analyses that drive impactful interventions, programs, services and solutions. Committed to impartiality, PR employs a human-centered design approach, emphasizing the value of contextualized knowledge to better serve communities. PR primarily focuses on underserved communities and vulnerable populations.

Founded in 2022 in Washington, PR was established with a vision of a world where no one's dignity or rights are compromised in social decision-making and service delivery due to misinformation, disinformation, or lack of systematic information. Today, PR remains steadfast in its mission to deliver evidence-based insights that inform policies, programs, and services to tackle critical community challenges and needs.

Published in September 2025 by **PORSESH POLICY RESEARCH INSTITUTE (PR)**
www.prresearch.us

Published September 2025
Community Centered Tech Series

Keywords: Artificial intelligence; bias; system; discrimination; minority

1. Introduction

Artificial Intelligence (AI) systems are increasingly being used in decision-making processes across many areas, including hiring, healthcare, criminal justice, and financial services. Predictive AI, in particular, has become a part of everyday life. However, the use of AI in such processes can have serious consequences for human rights, especially for minorities and marginalized groups. AI systems sometimes reproduce bias due to issues in training data, algorithm design, and broader social and structural factors. As a result, they can discriminate against, marginalize, or exclude minorities, reinforcing existing prejudices and sometimes creating new forms of discrimination.

This paper examines how AI bias contributes to discrimination against minorities and explores potential solutions to address this issue. In the context of AI, minority groups are those that are underrepresented in data, making them more vulnerable to biased outcomes. In other words, minority groups in the context of AI can be thought of as "minority data" because their experiences are often missing or not proportionally represented in datasets. Underrepresentation in AI data can therefore lead to harm, marginalization, and exclusion of minorities. Minorities can be defined by race, ethnicity, language, sexual orientation, disability, or other social characteristics.

AI bias is not only a technical issue but also a human rights concern. It touches on fundamental rights such as the right to work, education, health, security, privacy, and, most importantly, the right to equality and non-discrimination. As article 2(1) of the Universal Declaration of Human Rights states: "Everyone is entitled to all the rights and freedoms outlined in this Declaration, without distinction of any kind, such as race, color, sex, language, religion, political or other opinion, national or social origin, property, birth or other status." Addressing AI bias which leads to discriminatory decisions therefore requires more than technical solutions. It also calls for legal, policy, social, and political measures at both national and international levels to ensure that AI systems respect everyone's rights without any distinction.

2. How AI Bias Leads to Discrimination

According to the Oxford Dictionary, bias is "an inclination or prejudice for or against one person or group, especially in a way considered to be unfair." In the context of AI, bias is defined as a systematic error in decision-making processes that produces unfair outcomes[1]. AI bias refers to systematic errors in AI systems that disadvantage certain groups, often reproducing or amplifying existing social inequalities. AI bias can lead to AI-driven discrimination, also known as "Algorithmic Discrimination". This occurs when decisions made by AI result in unfair or unequal outcomes for some groups, even if there is no discriminatory intent.

[1] Ferrara, "Fairness and Bias in Artificial Intelligence."

Psychologists have identified approximately 180 cognitive biases that can affect how people think and make decisions. These biases can appear when people design Machine Learning models[2]. If the design of a model, including the data used, the feature selected, the algorithms, or the way the evaluation method is applied, does not fairly represent all groups, bias can emerge. For example, feature selection bias occurs when important information is missing for certain groups. Another example is algorithmic bias, which is not only a technical flaw but also reflects the broader social context, which is often discriminatory against minorities.[3] Therefore, different causes are contributing to how AI biases lead to discrimination, which can be defined in three main categories: 1) data problems, 2) algorithm design, and 3) broader societal factors.

Data Problems

First, data problems play a central role in how AI contributes to discrimination against minorities. The quality of training data strongly influences how algorithms behave. If datasets contain prejudice, imbalance, or historical bias, these flaws shape the decision-making process of AI systems[4]. Additionally, training data are often incomplete or fail to represent certain groups. When some groups are overrepresented or underrepresented, the resulting models produce biased and discriminatory outcomes[5]. Minority groups, in particular, are frequently underrepresented in datasets, which puts them in a vulnerable situation.

Historical bias exacerbates the problem, as machine learning requires large amounts of data, all of which are often drawn from the recent or distant past —some of which may have been generated during an earlier period with different priorities or contexts and may be misleading or outdated for current purposes. If past data reflect prejudice, AI models not only reproduce past biases but also amplify that bias into the present. In other words, past biases become future biases[6].

In a broader debate, the question of how data itself is produced demonstrates the complexity of addressing bias in AI systems. Is data neutral? Some argue that data is shaped by historical patterns of injustice[7]. Data is influenced and shaped within and by power structures. Data is produced as a result of interactions between humans in a specific culture and context. Data in AI are certainly not raw materials to feed algorithms: they are inherently political interventions.[8] If the data is intrinsically biased, does this raise a critical question: Is it even possible to have an AI system that makes fair decisions? From this perspective, the answer is no because the data, knowledge, and culture is

[2] You Chen et al., “Human-Centered Design to Address Biases in Artificial Intelligence.”
[3] You Chen et al., “Human-Centered Design to Address Biases in Artificial Intelligence.”
[4] Samala and Rawas, “Bias in Artificial Intelligence.”
[5] Ashwini K.P., *Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*.
[6] Ashwini K.P., *Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*.
[7] Crawford, *Atlas of AI*.
[8] Crawford, *Atlas of AI*.

biased. Of course, this bias is not limited to AI systems but also extends to non-AI decisions for the same reasons.

Anyway, bias is not limited to how the data is gathered. For example, data annotation and data labeling can also introduce bias, as annotators may have different interpretations of the same data. Subjective labels such as facial expressions can be influenced by cultural or personal biases[9]. Some scholars go further, arguing that the very processes of collecting data, categorizing, and labeling data, and subsequently using it to train systems, are political acts[10]. So, again, if we assume that even data is not biased, flaws in the training data are corrected, the result may still be problematic because bias can also occur in later stages of the AI lifecycle, including the designing phase, which will be discussed in the next paragraphs.

Algorithm Design Choices

Second, algorithm design choices can embed bias. Even if the data are representative, the way algorithms are designed still affects outcomes. In other words, the background or perspective of designers may cause them to embed biases in the algorithm design.[11] Designers make decisions such as which model to use, how to optimize it, and which variables to prioritize, which can all lead to bias[12]. First, the model of the algorithm selected influences how patterns are understood. For example, a simple model like linear regression may not capture complex realities and can disadvantage groups whose data does not fit in that model. Secondly, the way an algorithm is optimized also shapes its outcomes. Optimization means tuning an algorithm to meet goals like accuracy, profit, or speed, but if fairness is not included, it can disadvantage minorities. Optimization techniques often favor predictions for the majority groups over minorities[13]. Most systems are trained to maximize accuracy, efficiency, or profit, but not fairness. A hiring tool optimized only for "job performance" may reproduce past patterns where men were favored, leading to the exclusion of qualified women. Third, the choice of which variables to prioritize can indirectly introduce sensitive factors. For instance, using "zip code" in loan approvals can serve as a proxy for race or income level, reinforcing segregation and economic inequality.

Bias can also arise from feedback loops in algorithm design. Feedback loops in AI systems mean that when a system's outputs become inputs that influence its future decisions, it creates a self-reinforcing cycle. So, if an AI system that receives biased input could make biased decisions, it would ultimately reinforce bias in a self-perpetuating cycle[14]. For example, Netflix recommends movies based on what you have watched, and your choices from those recommendations further train the algorithm to suggest similar content. In another example, a policing algorithm that labels

[9] Chapman University, "Bias in AI."
[10] Crawford, *Atlas of AI*.
[11] Ashwini K.P., *Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*.
[12] Samala and Rawas, "Bias in Artificial Intelligence."
[13] Chapman University, "Bias in AI."
[14] DAN HENDRYCKS, *INTRODUCTION TO AI SAFETY, ETHICS, AND SOCIETY*.

certain neighborhoods as “high risk” will send more officers to those areas. This increased presence leads to more arrests, not necessarily because crime is higher, but because more policing occurs. Many of these design choices are hidden within “black box” systems, making them difficult to examine or challenge. These examples demonstrate that bias does not come only from data but also from design choices made by developers and from interactions between human and AI systems. Therefore, it is important to recognize that bias can happen in various stages of the AI lifecycle[15].

Fairness, among many other approaches, is often presented as a solution to mitigate AI bias, but it is a complicated and contested idea with no single agreed-upon definition.[16] Different perspectives pursue different goals: Individual fairness focuses on treating similar individuals similarly. Group fairness emphasizes that all groups, including minorities, should receive similar outcomes. Procedural fairness focuses on improving the process itself. Distributive fairness highlights the equal distribution of resources. Counterfactual fairness suggests a model is fair if its predictions remain the same even when a protected characteristic, such as race, is changed.[17] Finally, fairness as justice adds another layer of complexity, as the definition of justice itself is debated. These approaches tell us that while mitigation solutions such as fairness seem straightforward, they are extremely complicated. Using one definition may violate others. Creating fair AI systems also depends on the field of application, such as recruitment, health, criminal justice, and on the views of different stakeholders in each field.

Societal Factors

Societal factors also play an important role in shaping AI bias, which can deepen existing inequalities and discrimination against minorities. AI systems are not developed in isolation; they are created within wider social, historical, and cultural contexts. These contexts influence both the data collected and the way algorithms are designed and trained[18]. Historically, knowledge and cultural norms have largely been shaped by dominant groups, and the perspectives of minorities have often been excluded. This exclusion reflects what Miranda Fricker calls “epistemic injustice,” where the knowledge and experiences of certain groups are discredited because of prejudice against their identity. Such injustice particularly affects communities with less social power[19].

Since dominant groups have historically shaped knowledge and how the world is understood, the knowledge and data used to train AI reflect their perspectives. Algorithms are then built on the same foundation, often overlooking the experiences of minorities. Because these social and cultural contexts are unjust, they tend to reproduce themselves through AI systems and human interactions. In other words, AI systems are surrounded by social, political, cultural, and economic worlds, shaped by humans, institutions, and driving forces. Those who design and develop AI systems are

[15] Chapman University, “Bias in AI.”

[16] DAN HENDRYCKS, *INTRODUCTION TO AI SAFETY, ETHICS, AND SOCIETY*.

[17] DAN HENDRYCKS, *INTRODUCTION TO AI SAFETY, ETHICS, AND SOCIETY*.

[18] Samala and Rawas, “Bias in Artificial Intelligence.”

[19] Jackie Kay et al., *Epistemic Injustice in Generative AI*.

influenced by this context and the knowledge that represents the dominant perspective. These result in a continuing cycle of discrimination against minorities. In the following section, we present some examples of how biases through the mechanisms that we have just explained can lead to such outcomes in AI systems.

3. **Real Examples of AI Discriminatory Decisions Against Minorities**

In real-world cases, there are hundreds of examples of bias throughout the tech ecosystem. Many more have either never been detected or publicly admitted.[20] In 2025, for example, the University of Melbourne found that the AI system in the recruitment process failed to accurately evaluate candidates with speech disabilities or non-native accents. [21] In 2024, for example, the University of Washington carried out a study on gender and racial bias in AI resume-screening tools. The researchers tested a large language model on identical resumes, changing only the names to reflect different genders and races. The findings showed that the AI consistently favored names associated with white men, while resumes with black male names were never ranked first. Asian female names performed slightly better, but overall, the system reflected historical inequalities in hiring[22]. In another example, the U.S Equal Employment Opportunity Commission revealed that an AI system rejected female applicants aged 55 and older and male applicants aged 60 and above.[23]

Other research shows similar concerns. In the healthcare and welfare sector, for instance, a widely used health risk-prediction algorithm covering over 200 million U.S. citizens showed racial bias because it relied on healthcare spending as a proxy for medical need. Since less money is often spent on Black patients, the algorithm underestimated their care requirements compared to those of white patients[24]. Similarly, AI diagnostic tools for skin cancer perform less accurately on people with darker skin due to limited diversity in training datasets[25]. According to Fricker, this constitutes epistemic injustice because the AI systems do not exactly understand and recognize black skin.

In criminal justice, predictive policing and risk assessment tools have amplified racial disparities. Facial recognition systems frequently misidentify individuals with darker skin tones, leading to wrongful arrests. One of the most prominent examples is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) in the United States, used by judges to predict whether defendants should be detained or released on bail pending trials. A study found that COMPAS incorrectly labeled Black defendants as high risk nearly twice as often as white defendants[26].

[20] Crawford, *Atlas of AI*.
[21] Cem Dilmegani, "Bias in AI: Examples and 6 Ways to Fix It."
[22] Stefan Milne, "AI Tools Show Biases in Ranking Job Applicants' Names According to Perceived Race and Gender."
[23] Cem Dilmegani, "Bias in AI: Examples and 6 Ways to Fix It."
[24] Cem Dilmegani, "Bias in AI: Examples and 6 Ways to Fix It."
[25] Cem Dilmegani, "Bias in AI: Examples and 6 Ways to Fix It."
[26] Samala and Rawas, "Bias in Artificial Intelligence."

A case against the Navy Federal Credit Union revealed that 52 percent of Black borrowers were denied loans, compared with 23 percent of white borrowers. Research also shows that AI systems are up to 80 percent more likely to reject loan applications from Black Americans.[27]

In another instance in India, the facial recognition system used by the Delhi Police was reportedly accurate in only 2 percent of cases, which put minority groups at a disproportionate risk of misidentification and false arrests. In Brazil, according to a study in 2019, 90 percent of people arrested by law enforcement officials based on faulty facial recognition tools are of African descent.[28] In one specific example of facial recognition in an educational setting in the Netherlands, students of African descent have had to shine lights in their faces to be recognized by the AI systems used for examinations.[29]

In addition to image recognition, voice recognition systems often struggle to identify female voices, and social media platforms tend to show higher-paying job advertisements to men more than to women.[30] The lack of transparency in how algorithms make decisions shows how AI can commit discrimination by relying on biased or stereotype-based classifications. These examples indicate how AI biases reinforce cycles of discrimination, disparity, and stereotypes against certain groups, particularly those who are marginalized and excluded in society, and in particular, from power and knowledge generation.

4. Addressing the Problem: Solutions and Pathways

Understanding and acknowledging the causes of bias is the first step in addressing it. As noted previously, bias in AI systems is not only a technical problem. It is also shaped by broader social and cultural factors, since these systems are developed within existing societal contexts. Therefore, solutions for reducing biases cannot rely on technical measures alone. They must also include policy, legal, social, and ethical approaches at both national and international levels.

A. Technical/Data Level Solutions

- Ensure that data is complete, cleaned, diverse, representative, balanced, and inclusive.
- Conduct data sampling, annotation, and labeling carefully and accurately.
- Audit datasets to identify problematic patterns, using red-teaming and third-party reviewers[31].
- Regularly monitor AI systems for bias and discriminatory outcomes.

[27] *Human Centered Artificial Intelligence*, "AI's Impact on Black Americans."
[28] Ashwini K.P., *Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*.
[29] Ashwini K.P., *Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*.
[30] Crawford, *Atlas of AI*.
[31] Cem Dilmegani, "Bias in AI: Examples and 6 Ways to Fix It."

- Apply bias detection methods such as adversarial testing (change inputs to see if it behaves unfairly) and explainable AI techniques (to know why it made such a decision)[32].
- Promote algorithmic transparency to understand how AI systems make decisions.
- Include participatory design processes that involve diverse stakeholders throughout the development of AI systems.
- Build diverse teams to develop and design AI systems.

B. Policy and Legal Solutions

- Develop regulatory frameworks that prioritize transparency, accountability, and inclusiveness.
- Apply and update existing anti-discrimination laws to address AI-related harms, as these laws have been developed in a different time for different needs.
- Extend the International Convention on the Elimination of All Forms of Racial Discrimination to cover AI technologies. This convention covers employment, health, and public services, but it predates AI and does not cover AI decision-making.
- Conduct bias impact assessments to evaluate the social consequences of AI systems.
- Establish liability frameworks to hold organizations accountable for discriminatory outcomes.
- Promote a multidisciplinary approach to AI by encouraging collaboration among experts in computer science, ethics, sociology, philosophy, and human rights.
- Adopt a human right–based approach to AI governance systems to ensure that fundamental rights are respected and protected.

C. Social and Ethical Solutions

- Increase AI literacy among communities affected by AI systems.
- Strengthen civil society oversight and advocacy efforts.
- Develop public education programs on the responsible and human rights–focused use of AI.
- Promote international cooperation to adopt anti-discrimination legal frameworks to protect minority rights.
- Facilitate global dialogue on the intersection of AI and human rights, with special attention to minority rights.
- Create platforms for sharing knowledge, experiences, and best practices. The establishment of the Global Dialogue on AI Governance by the United Nations is a bold step in this direction.
- Continue research and public awareness efforts related to bias in AI systems.
- Engage a broader range of communities in the design and implementation of AI systems to generate authentic feedback.

[32] Chapman University, “Bias in AI.”

Conclusion

AI bias represents one of the most significant challenges of AI governance in today's AI-driven society, with profound implications for minorities. This paper demonstrated how AI biases contribute to discrimination, marginalization, and exclusion, and reinforce existing prejudices against minorities. The causes go beyond merely technical problems. They stem from the complex interactions between flawed and biased data, algorithm design choices, and broader societal factors. The examples from hiring, healthcare, and criminal justice demonstrate how AI systems can reproduce and amplify historical discrimination and sometimes can create new forms of discrimination against minorities.

Reducing AI bias requires a multifaceted and comprehensive approach that combines technical solutions with policy changes, legal reform, and social action. Encouragingly, some companies have recognized this problem and have begun taking some actions to ensure their system behave relatively unbiased[33]. However, putting these solutions into practice will require long-term efforts by different stakeholders, such as AI companies, citizens, minority and marginalized groups, civil society organizations, academic and educational institutions, states, and international organizations. This is especially challenging today as many governments and major players focus more on advancing AI innovation than addressing AI-based discrimination against minorities and marginalized communities, hoping to have an advantage and upper hand in the global AI competition.

[33] *OpenAI*, "The Power of Personalized AI."

Bibliography

Ashwini K.P. *Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*. A/HRC/56/68. Human Rights Council, 2024. https://docs.un.org/en/A/HRC/56/68.

Cem Dilmegani. "Bias in AI: Examples and 6 Ways to Fix It." *AI Multiple Research*, August 25, 2025. https://research.aimultiple.com/ai-bias/.

Chapman University. "Bias in AI." *Chapman University AI Hub*, n.d. https://www.chapman.edu/ai/bias-in-ai.aspx#:~:text=For%20example%2C%20if%20a%20facial,tones%2C%20leading%20to%20discriminatory%20outcomes.

Crawford, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.

DAN HENDRYCKS. *INTRODUCTION TO AI SAFETY, ETHICS, AND SOCIETY*. Center for AI Safety, 2024. https://www.aisafetybook.com/.

Ferrara, Emilio. "Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies." *Sci* 6, no. 1 (2023): 3. https://doi.org/10.3390/sci6010003.

Human Centered Artificial Intelligence. "AI's Impact on Black Americans." July 22, 2024. https://hai.stanford.edu/news/ais-impact-black-americans.

Jackie Kay, Atoosa Kasirzadeh, and Shakir Mohamed. *Epistemic Injustice in Generative AI*. August 21, 2024. https://arxiv.org/html/2408.11441?utm_source=chatgpt.com.

OpenAI. "The Power of Personalized AI." January 17, 2025. https://openai.com/global-affairs/the-power-of-personalized-ai/.

Samala, Agariadne Dwinggo, and Soha Rawas. "Bias in Artificial Intelligence: Smart Solutions for Detection, Mitigation, and Ethical Strategies in Real-World Applications." *IAES International Journal of Artificial Intelligence (IJ-AI)* 14, no. 1 (2025): 32. https://doi.org/10.11591/ijai.v14.i1.pp32-43.

Stefan Milne. "AI Tools Show Biases in Ranking Job Applicants' Names According to Perceived Race and Gender." *University of Washington*, October 31, 2024. https://www.washington.edu/news/2024/10/31/ai-bias-resume-screening-race-gender/?utm_source=chatgpt.com.

You Chen, Ellen Wright Clayton, Laurie Lovett Novak, Shilo Anders, and Bradley Malin. "Human-Centered Design to Address Biases in Artificial Intelligence." *Journal of Medical Internet Research*, March 24, 2023. https://pmc.ncbi.nlm.nih.gov/articles/PMC10132017/?utm_source=chatgpt.com.

Partner With Us for Impact

At **PORSESH POLICY RESEARCH INSTITUTE (PR)** we don’t just produce community-centered data and research—we turn insights into action, always putting the community at the heart of our work. We collaborate with government agencies, foundations, NGOs, and community-based organizations to co-design and implement data-informed strategies that address real-world challenges and strengthen community services. If your organization is seeking support in:

- Conducting community-centered research
- Translating data into meaningful policy and practice
- Designing and evaluating programs
- Facilitating inclusive stakeholder engagement
- Delivering data-informed training tailored to community needs.
- Co-organizing public talks, seminars, and learning events.
- Providing strategic consulting and advisory services

We invite you to collaborate with us.
Our team brings rigorous research methods, trusted community relationships, and deep cultural fluency to every engagement.

Email: president@prresearch.us
Website: https://prresearch.us/