UOT: 004.8 (025.4.03)
DOI: https://doi.org/10.30546/09090.2025.210.001

ALGORITHMIC BIAS: INJUSTICE IN ARTIFICIAL INTELLIGENCE AND HOW TO PREVENT IT

GULBICHA ZAHIDOVA[1], LEYLA ZEYNALLI – HUSEYNZADA[2], AYTAKIN VALIYEVA[3], ALI BAGIRZADA[4]

[1]guabdullayeva@beu.edu.az
[2]lzeynalli@beu.edu.az
[3]aveliyeva4@std.beu.edu.az
[4]ebagirzade1@std.beu.edu.az
Baku Engineering University
Xhirdalan, Azerbaijan

ARTICLE INFO

Article history:
Received:2025-04-18
Received in revised form: 2025-04-18
Accepted:2025-05-29
Available online

Keywords:
"Artificial intelligence"
"Algorithmic bias"
"AI transparency"
"Ethical issues"
"Unbiased data"
"Legal regulation"
2010 Mathematics Subject Classifications: 68T05

ABSTRACT

With the proliferation of applications and the impact of artificial intelligence (AI) on decision-making, algorithmic bias and discriminatory outcomes are emerging as a critical ethical issue. Algorithmic bias is the data generated by the software used to train AI models, design choices, and behavior. These biases have the potential to exacerbate social vulnerabilities in areas such as employment, law, medicine, and social media.

This article explores the nature of algorithmic bias and its social impact, and explores existing techniques and approaches to combat this problem. Key ways to mitigate bias include creating a more participatory data base, informed AI models, ethical practices, and legal enforcement. The article highlights the importance of designing more balanced and transparent mechanisms in AI systems, and suggests the potential solutions for their implementation.

INTRODUCTION

Algorithmic bias refers to systematic and repeatable errors in AI/ML systems that result in unequal outcomes. Typically, these biases arise due to incomplete or unrepresentative input data, flaws in algorithm design, or the exclusion of certain groups during technological innovation. Algorithmic bias has the potential to favor certain user populations at the expense of others, leading to discriminatory outcomes.

As artificial intelligence increasingly influences more aspects of human life, bias in algorithms has become a widespread issue. One of the main causes of algorithmic bias is the inaccuracy present in training datasets. Inaccurate or incomplete data can lead to erroneous findings and impair the algorithm's ability to make impartial judgments. Human input plays a crucial role in determining how data is collected, which data is used, and what level of performance the AI system is expected to achieve. Without effective regulation, AI-driven decisions can lead to discrimination and disproportionate treatment of certain groups.

Moreover, programmers and algorithmic design significantly contribute to the creation of bias. AI systems, during the process of model training and coding, can replicate human prejudices. As a result, AI algorithms may develop judgments that lead to unequal and discriminatory treatment of specific social groups. If these systems are not properly tested and fine-tuned, they can operate based on incorrect or incomplete data, thereby perpetuating bias.

In general, algorithmic bias is an urgent issue that must be explored and thoroughly addressed to ensure fairness and trust in AI systems. Reducing bias should involve creating diverse and comprehensive datasets, developing explainable AI models, and strengthening regulatory frameworks with an emphasis on ethics. Otherwise, the societal impact of artificial intelligence will continue to raise critical concerns and may even exacerbate existing disparities.

There are three major and most common types of algorithmic bias: [1]

1. Data Bias
2. Algorithmic Design Bias
3. User Bias

1. Data Bias – This occurs when the data is inaccurate, incomplete, limited in certain ways, or fails to provide a full representation of the population. As a result, a distorted view is formed or vital context is omitted. This type of bias has the potential to affect individuals, organizations, and even entire societies.

AI systems make decisions based on data. However, data may not be objective, and systems trained on biased data can lead to unfair and inaccurate outcomes. Data bias can cause AI systems to discriminate against certain groups of people based on disproportionate or incorrect information.

There are several different types of Data Bias, each arising for different reasons. These biases hinder the unbiased functioning of AI and create challenges in many areas of society. The main types of data bias include the following:

Selection Bias – Selection bias occurs when the data used to train AI systems favors a specific group. When training data does not cover all the diverse segments of society and is skewed in its selection, AI systems may end up making unfair and biased decisions against certain groups. These biases can result in the production of misinformation and discriminatory practices by AI in various fields such as healthcare, law, finance, and social media.

Selection bias is especially evident when working with small and inaccurate datasets. If AI systems are trained on non-representative samples, they may produce unrealistic, incorrect outcomes and make unjust decisions against certain groups within society.

There are several key reasons for the emergence of selection bias:

➢ *Preference for certain groups during data collection*

If the data used to train AI systems is collected from specific regions or groups, other groups not included in that dataset may be misrepresented, leading to incorrect outcomes.

➢ *Historical data lacking representation of certain groups*

If, in the past, sufficient data was not collected about certain groups within a field, then AI systems trained on such historical data will also be unable to provide accurate results about those groups.

- *Algorithms trained on limited and non-representative samples*

When some AI systems are trained only on specific samples, they may fail to properly analyze groups that do not belong to those samples.

Selection bias can be observed in many real-world sectors. Examples of selection bias include the following:

a. *In Facial Recognition Systems:*

Most modern facial recognition systems are primarily trained on images of white individuals. As a result, they tend to misclassify people from other ethnic groups more frequently. For instance, Black and Asian individuals are more likely to be incorrectly identified by these systems.

b. *In Medical Diagnostic Systems:*

If an AI medical system is trained only to analyze male health data, it may misdiagnose diseases and symptoms in women. This can lead to incorrect diagnoses and inadequate treatment.

c. *Recruitment Systems and Gender:*

When specific data is fed into AI recruitment systems, historical male dominance in certain professions can cause bias. For example, AI systems trained on historical data showing higher hiring rates for men in technology and engineering fields may automatically undervalue new female applicants.

Selection bias can lead to social and economic discrimination in decisions made by artificial intelligence. As a result, some individuals may face unequal treatment in areas such as employment, credit access, and medical services. Selection bias can cause incorrect medical diagnoses and errors in judicial decisions, potentially leading to serious consequences, including death. When AI systems operate in a discriminatory manner, certain social groups are disproportionately harmed. For example, facial recognition software may misidentify individuals from specific ethnic groups, which can lead to legal issues.

There are several ways to prevent selection bias. Firstly, the data used to train AI systems should be collected from as many diverse sources as possible. This will help create a balanced dataset that includes a mix of social, ethnic, and gender categories. In fact, AI systems should be tested across various groups and validated for their results. AI systems developed without considering diversity and inclusion may generate discriminatory outcomes. Therefore, the dataset used in an AI system's decision-making should be tested and normalized for potential biases.

Finally, ethical standards must be applied to AI developers and experts. During the development of AI systems, adherence to ethical guidelines and principles of inclusion should be standard practice. In this way, AI systems will be more likely to make objective and fair decisions.

2. Algorithmic Design Bias – Algorithmic design bias is identified as a source of bias introduced during the design phase of AI systems, especially during programming and model development. This type of bias arises from flaws in the operational logic of AI algorithms, evolving model processes, and control parameters that influence the decision-making process.

In the real world, algorithmic design bias can lead to AI systems making non-objective decisions across various sectors. Here are some examples: [2]

- Credit Scores and Banking

Artificial intelligence algorithms are widely used in credit approval processes by financial institutions. However, if such algorithms are built on criteria that favor certain income levels or jurisdictions, individuals' chances of receiving credit can be significantly reduced. For example, some algorithms may classify people from certain socio-economic classes as bad customers, thereby subjecting them to higher interest rates. This limits their access to the financial system and has the potential to perpetuate economic inequality.[3]

- Recruitment Systems

If AI filters job listings based on historical data where a certain gender or ethnicity was favored, it may result in discrimination against new applicants. Instead of eliminating bias, AI may perpetuate existing prejudices. For example, Amazon's AI-based recruitment system consistently downgraded female applicants because the majority of previous hires in technical roles were men. As a result, AI systems may replicate past biases and reinforce gender inequality.

- Medical Diagnostics

AI-based diagnostic systems trained on medical data from specific ethnic or gender groups may misdiagnose diseases prevalent in other groups. This can lead to incorrect diagnoses and unequal access to healthcare. For instance, studies have shown that some AI models used to detect skin conditions struggle to diagnose dark-skinned individuals, as they were primarily trained on images of light-skinned patients. These biases can negatively affect the quality and accessibility of care.

- Legal and Policing Systems

AI tools used to predict the likelihood of crime can discriminate against certain social groups. If the model reflects biases present in the training data, it can lead to unequal treatment in the criminal justice system. For example, the COMPAS AI system used in the U.S. assigned higher recidivism risk scores to Black individuals. As a result, they were often unfairly given harsher penalties, leading to more unjust court rulings. Such biased algorithms can further exacerbate existing social imbalances within the justice system.

Algorithmic Design Bias primarily arises due to the following reasons:

- Functional limitations of the algorithm – In such cases, the model may fail to consider or correctly process certain factors.
- Incorrect selection of hyperparameters – Bias can occur when the learning methods and hyperparameters of the AI are improperly set.
- Developer bias – The personal beliefs and perspectives of the programmer or researcher can influence the decisions embedded in the algorithm.
- Lack of transparency in the algorithm – In systems known as "black-box" models, it becomes difficult to analyze the causes of bias when they are observed.

These factors contribute to the emergence of algorithmic design bias. However, it is certainly possible to prevent such biases when they arise. The following recommendations should be implemented to reduce and eliminate algorithmic design bias:[4]

1. Algorithmic decisions should be reviewed by humans – Instead of relying solely on AI systems, decisions should be examined by experts.
2. Transparent algorithms must be developed – To achieve this, explainable AI models should be created instead of AI systems that function like black boxes. Specific research should be conducted to understand the AI decision-making process, and the functionality of these systems should be clearly explained.
3. Developers should be required to design fair and unbiased models – This means taking into account various ethnic and social groups. The data used to train AI must represent all layers of society and must not contain any bias. While developing a model, equality must be ensured across gender, race, and social class to achieve fair and objective results.
4. Members from diverse social and cultural backgrounds should be involved in AI development – Neutral developers and team members should be hired to be part of the process, and more objective and fair algorithms should be created by considering multiple perspectives.

By implementing these recommendations, algorithmic design bias can be effectively prevented.

3. User Bias – refers to the distortion in an AI system's decision-making process caused by individuals' personal experiences, preferences, and beliefs. This type of bias is related to the way users search for and select certain types of information — in other words, individuals often prefer information that aligns with their existing beliefs. User bias can cause AI systems to learn from user behavior and subsequently present more similar information in the future.

Here are some examples of user bias:

- Search engine bias:

People tend to select results that align with their own opinions and give them more importance. For example, a person searching for positive or negative political news will likely choose results that support their views. AI systems learn this behavior and continue to provide more of the same.

- Bias in recommendation systems:

On shopping websites and social media platforms, users are shown product suggestions based on what they've already selected or shown interest in. This limits users to information and content within their existing interests and prevents them from being exposed to a broader range of perspectives.

- Bias in social media and news content:

People read and share news and content that match their own views. AI systems track this tendency and provide users with more similar content. This leads to a situation called a filter bubble, where individuals are exposed only to information that supports their beliefs.

Such bias can limit the user's access to diverse content and different perspectives on the same issue, eventually leading to social fragmentation and a narrowing of cognitive processes.

Here are some key reasons behind the emergence of user bias:[5]

- Personal Beliefs and Interests of the User:

Users tend to consume information that aligns with what they already believe or want to learn more about. As a result, AI models begin to mimic this behavior by reinforcing only the accepted viewpoints, leading to biased outcomes.

- Previous Experiences and Selective Preferences:

The human mind recalls and focuses on information it has previously encountered. This influences how AI processes and recommends content, as the system, based on learned patterns, will keep suggesting similar types of information — causing biased recommendations.

- Information Selection and Searching Behavior:

People often use search engines and social media platforms to access specific information. Their search filters and queries reflect their personal interests and preferences. This leads systems to prioritize certain types of data, neglecting broader or more general topics.

- Algorithmic Selection and Recommendation Systems:

Algorithms deliver and adapt content based on user behavior. If a user is only interested in a certain type of content, the system will generate more of it, reinforcing the user's bias.

To eliminate user bias, the following actions should be taken:

- *Trackable and Transparent Algorithms:*

AI programs should be more transparent so that users can understand how the system operates and on what basis it makes decisions. This increases trust in the system and allows for fairer and more balanced outcomes.

- Diverse Information and Perspectives:

Users should be presented not only with information based on their past preferences but also with a variety of viewpoints and topics. This approach helps them develop broader and more unbiased thinking and become familiar with new ideas.

- New Filters:

Users should be given more freedom in search and selection to access a wider range of sources. This ensures more comprehensive and balanced information and helps prevent misleading or narrow exposure.

- Diverse Recommendation Systems:

Social networks and other recommendation platforms should not rely solely on users' behavioral history. Instead, they should suggest content on various topics. This helps users discover new subjects and engage with different ways of thinking.

- Promoting User Knowledge and Education:

Users should be encouraged to explore a broader base of knowledge and learn about new topics. This leads to more balanced and unbiased decision-making.

- Algorithmic Optimization:

Algorithms should consider the needs of wider user groups rather than focusing solely on individual preferences. This results in more accurate and universal outcomes.

After noting these points, it is important to address the issue of transparency and accountability in algorithms, as this is one of the key principles necessary to understand and manage the impact of artificial intelligence and algorithms on society. Algorithmic transparency and accountability aim to explain how algorithms function, what decisions they make, how these decisions are made, and why those particular decisions were reached.

Transparency refers to clearly demonstrating how an algorithm works, what data it uses, and which methods are employed to produce results. Algorithmic transparency includes the use of data, the structure of the algorithm, open-source algorithms, and clear explanations.[6] In terms of data usage, it is important to know what kind of data the algorithm relies on. Where transparency exists, users and the public should be aware of why the data is collected and how it is processed. From the perspective of algorithmic structure, there should be complete information about how the algorithm is built and how it operates. This helps in understanding why the algorithm produces specific outcomes. Open-source algorithms allow external researchers and experts to study the code when it is publicly available. This helps to understand how the algorithm functions, including any possible flaws or biases it may contain. Users and the public should be provided with clear and understandable information about the algorithm or AI application. This is especially important due to the complexity of such technical systems.

Accountability refers to whether an individual or an organization is held responsible for algorithmic actions. When algorithms are used in areas that significantly affect us—such as education, healthcare, or employment—it is essential that someone or some organization is held accountable to ensure that the outcomes are correct and fair. Accountability mechanisms involve identifying the individuals or organizations responsible for the algorithmic decisions. If an algorithm produces an error or an unfair result, responsibility should be assigned to correcting those mistakes. Monitoring and auditing are crucial after the deployment of the algorithm, as they allow the ongoing evaluation of its performance and help assess the fairness of the results produced. These processes also provide opportunities to respond to suspicious or harmful outcomes. In situations where people are negatively affected by algorithmic decisions, responsibility and compensation mechanisms should be put in place. Under accountability, there should be clear information on when and how affected individuals can receive compensation for incorrect outcomes. Finally, before an algorithm is implemented, proper testing and evaluation procedures must be conducted to verify the accuracy and fairness of the data used. Transparency and accountability enable algorithms to function in a more equitable and just manner. This helps prevent discrimination and unfair treatment. Providing correct and transparent information about algorithms builds trust among citizens. People want to trust the systems that make decisions on their behalf.

Transparency also allows for the detection and correction of bias within algorithms, helping them become less biased and fairer. To protect individual rights, algorithms must be transparent, and users must understand how their data is being used and how their personal rights are safeguarded.[7] Therefore, transparency and accountability in algorithms are essential. In finance, credit scoring algorithms may be biased against certain ethnic groups or genders. Transparency allows such biases to be identified and eliminated. In law enforcement, facial recognition algorithms used by authorities may disproportionately affect specific ethnic groups. Transparency and accountability enable these algorithmic outcomes to be reviewed and assessed. In healthcare, transparent diagnostic algorithms can prevent misdiagnoses and biased treatment recommendations. However, a lack of transparency is common because certain organizations and companies prefer to keep the source codes of their algorithms confidential as trade secrets. To address this issue, appropriate legislation and regulatory policies must be developed. Additionally, when algorithms produce errors or harmful outcomes, determining accountability becomes a major concern. This issue must be clearly defined within a legal framework. Moreover, since algorithms are often trained by humans, they may inherit existing biases. To prevent this, it is necessary to use unbiased data sets and ensure periodic auditing of the algorithms.[6]

Algorithmic bias has a widespread impact across many areas of society and tends to deepen social inequality. If artificial intelligence is applied in a biased manner in areas such as recruitment, credit approval, and law enforcement, certain groups may be subjected to unfair discrimination, their economic opportunities may be limited, and social integration may become more difficult. In social media, the manipulation of algorithms and informational bias can lead people to consume only content that aligns with their existing beliefs, which in turn increases polarization in society. In the healthcare sector, biased algorithms may cause individuals from certain ethnic and social groups to receive incorrect diagnoses or substandard medical care. To solve such issues, transparency and accountability are essential—without them, addressing these problems becomes increasingly difficult. To ensure that AI systems function fairly and reliably, it is crucial to make their decision-making processes transparent, detect and eliminate biases, and adhere to ethical principles. To maintain public trust in AI technologies and ensure their operations align with human rights, effective regulation and monitoring mechanisms must be implemented. [10]

Conclusion

Algorithmic bias is one of the main issues responsible for unequal outcomes in artificial intelligence (AI) systems. It is primarily divided into three broad categories: data bias, algorithmic design bias, and user bias. Data bias occurs when training data is incomplete or skewed, leading AI systems to discriminate against certain groups. Algorithmic design bias refers to the impact of developers' intentional or unintentional biases during the coding and structuring of a model, which can influence the algorithm's decision-making process. User bias arises when individuals tend to select information that aligns with their own perspectives. In such cases, AI reflects those patterns, contributing to the formation of filter bubbles.

These biases can create profound injustices across various sectors such as recruitment, finance, law, healthcare, and social media. For example, in facial recognition systems, bias in the data can lead to the misclassification of certain ethnic groups. Similarly, AI used in hiring programs may reject applications based on gender or racial differences. To prevent these issues, it is essential to create more diverse and representative datasets, build transparent and explainable AI models, ensure human oversight and regulation of AI-related decisions, enforce ethical standards for developers, and conduct regular audits of AI systems. Transparency and accountability mechanisms must be implemented in algorithmic systems to address these problems. Without such measures, AI systems may perpetuate existing societal biases and exacerbate social injustices.

In conclusion, algorithmic bias is not only a technological issue but also a social and ethical one. To prevent it, both technological solutions and legal and ethical standards must be developed. Otherwise, the impact of AI will not reduce social inequalities but rather intensify them.

REFERENCE LIST

1. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning: Limitations and Opportunities. MIT Press.
2. O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing.
3. Binns, R. (2018). "Fairness in Machine Learning: Lessons from Political Philosophy." Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT).
4. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). "A Survey on Bias and Fairness in Machine Learning." ACM Computing Surveys (CSUR), 54(6), 1-35.
5. Zliobaite, I. (2017). "Measuring Discrimination in Algorithmic Decision Making." Data Mining and Knowledge Discovery, 31(4), 1060-1089.
6. "A Survey of Recent Methods for Addressing AI Fairness and Bias in Biomedical NLP and Computer Vision" Journal of Biomedical Informatics, 2024.
7. Smith, E., & Johnson, M. (2024). "Ethical and Bias Considerations in Artificial Intelligence/Machine Learning Applications in Pathology." Pathology, 56(2), 123-130.
8. Zhang, H., & Patel, R. (2024). "A Survey of Recent Methods for Addressing AI Fairness and Bias in Biomedical NLP and Computer Vision." Journal of Biomedical Informatics, 139, 104273.
9. Seymour, J., & Nguyen, T. (2023). "Survey on Machine Learning Biases and Mitigation Techniques." Machine Learning and Knowledge Extraction, 4(1), 1.
10. Wang, L., & Li, F. (2023). "Algorithmic Bias in Artificial Intelligence: A Problem—And the Root Cause is More Insidious." Nursing Outlook, 71(5), 103-109.