

Bias in AI

Introduction:

AI has become an essential part of modern systems, from employment and healthcare to security and digital services. While it can improve efficiency and decision-making, one of its most significant challenges is bias.

Bias occurs when AI algorithms produce unfair or unbalanced decisions towards a particular group of people, often due to the data used to train models or design the algorithms themselves.

Problem Statement:

Despite the significant advancement in AI technologies, many systems still exhibit different levels of bias, whether it be in:

- Facial recognition systems that go the wrong way for darker skin.
- Recruitment algorithms that favor a particular gender or background.
- Criminal prediction systems that show bias against specific ethnic groups.

This bias often arises from:

- Unbalanced training data that represents certain groups more than others.
- Non-neutral mathematical models give more weight to certain attributes.
- biased human decisions are passed on to the algorithms as they are designed.

This bias leads to unfair decisions, reduces the reliability of AI, and affects the rights of individuals, making addressing the problem an urgent necessity.

Research Objectives:

1. Analyze the sources of bias in AI systems

The research aims to identify the types of bias (data bias, algorithmic bias, human bias) and understand how they transition to models.

2. Assess the impact of bias on the accuracy and reliability of systems

This includes studying applied cases such as recruitment, facial recognition, and healthcare, and measuring the impact of bias on outcomes.

3. Develop strategies to reduce bias

Such as:

- Improve data quality and balance
- Use fair algorithms (Fairness Algorithms)
- Apply debias techniques before, during, and after training

4. Propose a normative framework for assessing the fairness of systems

The research aims to establish clear criteria for measuring justice, such as:

- Equal Opportunity
- Demographic Parity
- Predictive Equality

5. Promote awareness of the importance of ethics in AI

By highlighting the social and legal risks of biased systems.

References:

You can use these well-known academic references in the field of bias and artificial intelligence:

- 1. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning.**
 - 2. Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research.**
 - 3. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys.**
 - 4. O’Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.**
 - 5. Mitchell, M. et al. (2019). Model Cards for Model Reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency.**
-