

# Bias

MMI Module 12

Activity 5.1: What are ethical issues concerning AI

Author: Krzysztof Dąbrowski

## 1. Definition

**prompt:** definition of bias in AI

### 1.1. Perplexity answer

The definition of bias in Artificial Intelligence (AI) refers to a systematic error in decision-making processes that leads to unfair outcomes. In the context of AI, bias can stem from various sources such as data collection, algorithm design, and human interpretation. It can manifest as **data bias**, **algorithmic bias**, or **user bias**, impacting the fairness and equity of AI systems. Data bias arises from unrepresentative or incomplete training data, algorithmic bias from biased assumptions in algorithms, and user bias from users introducing their biases into the system. To mitigate bias in AI, strategies like dataset augmentation, bias-aware algorithms, and user feedback mechanisms are employed. Addressing bias in AI is crucial to ensure fair and unbiased AI systems for all users.

### 1.2. MS Copilot answer

**AI bias** refers to biased results due to human biases that skew original training data or AI algorithms—leading to distorted and potentially harmful outputs. It's also known as machine learning bias or algorithm bias. Let me provide you with some real-world examples of AI bias:

**Healthcare:** Underrepresentation of data from women or minority groups can skew predictive AI algorithms. For instance, computer-aided diagnosis (CAD) systems have been found to return lower accuracy results for African-American patients than white patients.

**Recruitment Tools:** AI tools used for resume scanning during job searches can inadvertently result in disproportionate outcomes across groups. For example, if a job ad uses the word “ninja,” it might attract more men than women, even though that is not a job requirement.

**Image Generation:** Studies have shown that AI-generated images can reinforce gender bias. For instance, one study requested AI-generated images of people in specialized professions, and older people were always depicted as men, reinforcing gender stereotypes.

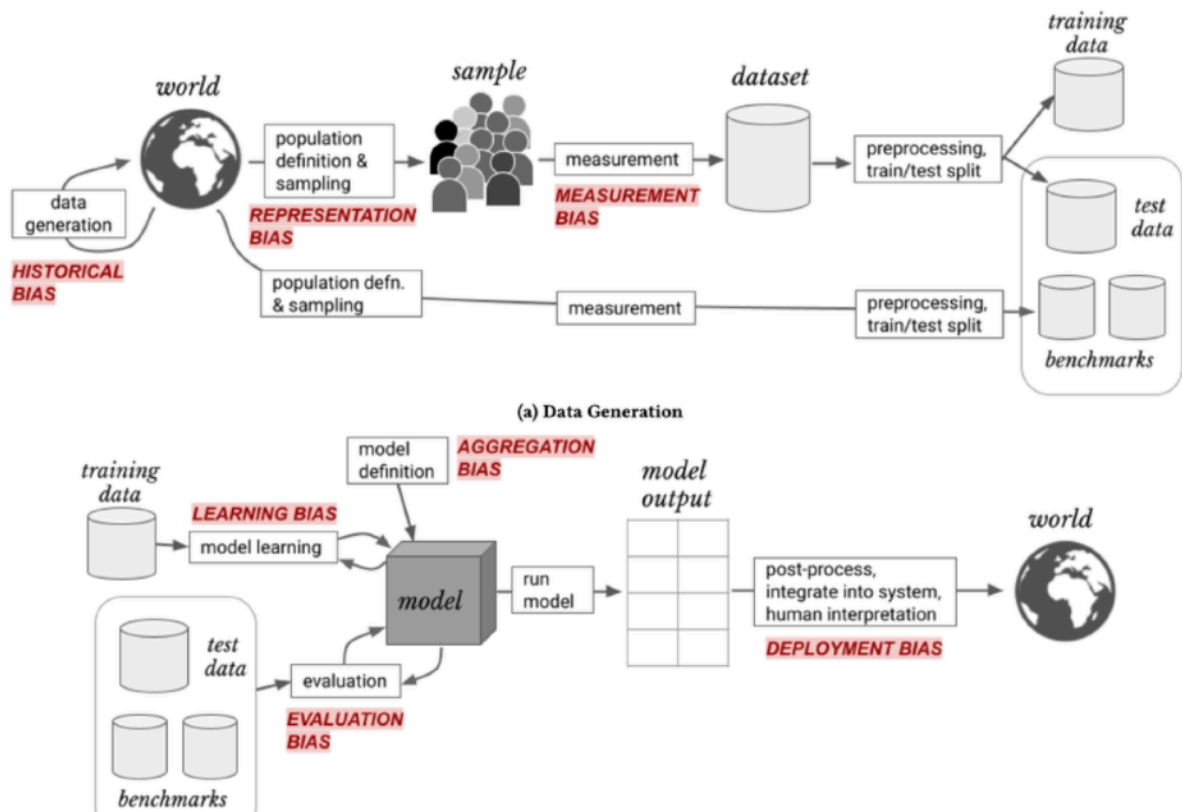


Figure 1: Types of biases. Source: <https://www.kaggle.com/code/alexisbcook/identifying-bias-in-ai>

## 2. Examples

**prompt:** examples of bias in machine learning, give concise answer

### 2.1. Perplexity answer

Examples of bias in machine learning include:

1. Sampling bias: Training data not sampled randomly enough from the collected data, leading to non-representative data.
2. Measurement bias: Differences in how certain features are assessed and measured, resulting in uneven conclusions.
3. Popularity bias: Models favoring popular items or conclusions, even if they are not the best.
4. Emergent bias: Biases that develop over time due to user interactions and changing habits.
5. Evaluation bias: Biases in model evaluation due to ill-suited or disproportionate benchmarks.
6. Population bias: Non-representative data from platforms like social media, affecting model outcomes.
7. Social bias: User-generated data reflecting social influences and biases.
8. Behavioral bias: Users reacting differently to the same information, leading to contrasting communication directions.

9. Automation bias: Blindly trusting AI decisions without verifying their accuracy or fairness.
  10. Implicit bias: Biases that are not consciously intended but still affect AI outcomes.
  11. Systemic bias: Biases rooted in institutional practices and underrepresentation of certain groups.
  12. Randomization bias: Data not properly randomized, leading to non-representative samples.
- These biases can significantly impact the fairness and accuracy of machine learning models, making it essential to address them through various strategies like data augmentation, bias-aware algorithms, and user feedback mechanisms.

## 2.2. MS Copilot answer

Here are concise examples of bias in machine learning:

1. **Gender Bias in Hiring Algorithms:** Some AI-based hiring tools have been found to favor male candidates over equally qualified female candidates due to biased training data or algorithm design.
2. **Racial Bias in Predictive Policing:** Predictive policing algorithms may disproportionately target minority neighborhoods, leading to biased law enforcement actions.
3. **Age Bias in Loan Approval Models:** Loan approval models might unfairly discriminate against older applicants based on age-related features.

Remember that addressing bias is crucial for creating fair and equitable AI systems. If you'd like more examples or have further questions, feel free to ask!

## 3. Limitations

**prompt:** What are examples for limitations using AI?

### 3.1. Perplexity answer

Examples of limitations using AI include:

1. Inaccurate Data Analysis: AI systems are only as good as the data they are trained on. Biased or incomplete data can lead to skewed, inaccurate, or even discriminatory outcomes.
2. Lack of Emotional and Social Understanding: AI still falls short in understanding human emotions, cultural norms, and social cues, posing challenges in applications that require a deep understanding of human behavior.
3. Ethical and Privacy Concerns: AI applications raise important ethical questions, including concerns about data privacy, informed consent, and algorithmic bias.
4. Complexity and Resource Intensity: Advanced AI algorithms can be resource-intensive, requiring high computing power and specialized hardware, making them inaccessible for many small and medium-sized enterprises.
5. Lack of General Intelligence: AI lacks the ability to transfer knowledge from one domain to another, a hallmark of human intelligence, and is far from achieving Artificial General Intelligence (AGI).
6. Dependence on Data Quality: AI algorithms are only as good as the data they are trained on. Biased or incomplete data can lead to unfair or inaccurate outcomes.
7. Algorithmic Bias: Biases primarily result from the partial design of the algorithm by programmers, who favored some desirable or self-serving criterion. Large platforms with algorithms such as search engines and social media sites frequently have algorithmic bias.
8. Lack of Creativity: AI systems are excellent at pattern recognition but lack creativity and the ability to think "outside the box." Creative tasks that require originality, imagination, and emotional understanding are challenging for AI.
9. High Costs: Implementing and maintaining AI systems can be expensive, requiring significant computing power and data, and continuous updates and monitoring to stay relevant and accurate.
10. Amplifying Biases: AI systems learn from historical data and can perpetuate biases in decision-making if that data contains biases or prejudices, reinforcing existing social inequalities.

11. **Requires Monitoring:** AI systems require constant monitoring and fine-tuning to ensure they perform as intended and align with ethical guidelines.
12. **Limited Understanding of Context:** AI often struggles to comprehend the broader context of a situation, missing subtle and complex aspects that are essential for effective decision-making.

These limitations highlight the need for careful consideration of AI's capabilities and constraints in various applications to ensure fair and effective outcomes.

### 3.2. MS Copilot answer

Here are some examples of limitations when using AI:

1. **Data Dependence:** AI models heavily rely on the quality and quantity of training data. If the data is biased, incomplete, or unrepresentative, the AI's performance can suffer.
2. **Overfitting:** AI models can become too specialized on the training data, leading to poor generalization to new, unseen examples.
3. **Ethical and Bias Challenges:** AI systems can inadvertently perpetuate biases present in the training data, leading to unfair outcomes.
4. **Interpretability:** Some AI models, like deep neural networks, lack transparency, making it challenging to understand their decision-making process.
5. **Resource Intensiveness:** Training complex AI models requires significant computational resources and energy.
6. **Security Risks:** AI systems can be vulnerable to adversarial attacks or misuse.

Remember that understanding these limitations helps us use AI effectively and responsibly! If you'd like more details or have other questions, feel free to ask!

## 4. Limitations

**prompt:** What are use cases where AI may not be appropriate? concise please.

### 4.1. Perplexity answer

Here are some use cases where AI may not be appropriate:

1. **Sensitive or high-stakes decisions:** AI should not be used for high-risk decisions that significantly impact people's lives, such as in healthcare, criminal justice, or financial lending, without human oversight.
2. **Analyzing restricted data:** AI should not be used to analyze sensitive or restricted data unless the system offers robust privacy protection and all identifying information is removed.
3. **Evaluating competitive bids:** AI should not be used to review vendor proposals in competitive bidding processes as it can introduce bias and be factually incorrect.
4. **Generating legal documents:** Using AI-generated legal language without expert review is incredibly risky as the output may sound convincing but be inaccurate or inappropriate.
5. **Replacing human judgment:** AI should augment human intelligence, not replace it entirely. Overreliance on AI can diminish critical thinking skills and lead to mistakes.
6. **Analyzing research data:** Using AI to analyze research data depends on the data's classification and the AI system's privacy protection. Restricted data should not be entered into AI tools.
7. **Reviewing grant proposals:** Using AI to review grant proposals is not appropriate, and some agencies have strictly forbidden it.
8. **Performing literature reviews:** While AI can be helpful in literature reviews, the results should be verified, and AI should not be used as the sole source.

In general, AI should not be used for tasks that require creativity, emotional understanding, or complex reasoning without human oversight. It is crucial to carefully consider the limitations and potential risks of AI in each use case.

## 4.2. MS Copilot answer

Here are some concise examples of use cases where AI may not be appropriate:

1. **Creative or Emotional Tasks:** For projects requiring more right-brain activity (such as creativity or emotional understanding), a hybrid human-AI solution might be better than relying solely on machine-led design<sup>1</sup>.
2. **Safety-Critical Processes:** In situations where AI could jeopardize existing processes or workforce safety, caution is necessary. Ensuring safety through proper levers is crucial<sup>2</sup>.
3. **Judicial Decision-Making:** While AI can aid judicial systems, ethical challenges arise due to lack of transparency, susceptibility to bias, and privacy concerns<sup>3</sup>.

Remember that thoughtful consideration of context and ethical implications is essential when deciding whether to use AI. If you have more questions, feel free to ask!