# Examples of Data Bias in Large Language Models (LLMs)

## Handedness Bias (Right-Handed Assumption)

Example: When asked to describe a person writing, ChatGPT may assume the person is right-handed unless explicitly stated otherwise.  
Bias Type: Representation Bias (more data about right-handed individuals than left-handed ones).

## Gender Bias in Professions

Example: When generating text about a doctor, ChatGPT may assume the doctor is male, while it may assume a nurse is female.  
Bias Type: Stereotype Bias (reinforcing gender roles based on historical data).

## Cultural or Regional Bias in Names

Example: Asking for a "common American name" may return "John Smith," overlooking the diverse cultural backgrounds in the U.S.  
Bias Type: Selection Bias (favoring certain demographic data over others).

## Language Bias (English-Centric Assumptions)

Example: When asked about global news, ChatGPT may prioritize English-language sources and perspectives.  
Bias Type: Monolingual Bias (favoring dominant languages in training data).

## Socioeconomic Bias (Tech-Centric Jobs Overrepresented)

Example: When asked about high-paying careers, the model may primarily list tech-related jobs while neglecting skilled trades and other industries.  
Bias Type: Exposure Bias (favoring data from well-documented industries over others).

## Racial or Ethnic Bias in Descriptions

Example: Descriptions of criminals may disproportionately use terms associated with certain ethnicities due to biased datasets.  
Bias Type: Implicit Bias (reflecting biases present in training data).

## Age Bias in Work and Technology

Example: When discussing technology adoption, the model may assume younger individuals are more tech-savvy, ignoring experienced older users.  
Bias Type: Age Stereotype Bias (favoring younger demographics in data).

## Western-Centric Knowledge Bias

Example: When asked about history or global issues, ChatGPT may provide Western perspectives more prominently than non-Western viewpoints.  
Bias Type: Cultural Bias (overrepresentation of Western narratives in datasets).

## Conclusion

Bias in LLMs arises due to the nature of the training data, which reflects societal biases. Addressing these biases requires diverse data representation, ethical AI development, and continuous model evaluation.