OUR WORKING DEFINITIONS OF BIAS

1 IMPLICIT BIAS

A disproportionate weight in favor of or against an idea or thing, usually in a way that is closed-minded, prejudicial, or unfair. Implicit biases can be innate or learned. People may develop biases for or against an individual, a group, or a belief.

2 SYSTEMIC BIAS

The inherent tendency of a structure to favor certain bodies or outcomes, as created and reinforced by social, governmental, political, and economic systems in place.

3 STATISTICAL BIAS

This can involve either bias in data collection or in the difference between the expected value and actual value of an estimate. With respect to the former, this is when the sample isn't representative of the population of interest. The latter is with respect to "results that are systematically off the mark."

OUR WORKING DEFINITION OF BIAS

4 MACHINE LEARNING BIAS

Machine learning bias refers to when a machine learning model's erroneous assumptions results in systemically biased predictions.

It reflects and amplifies implicit biases, and reflects and fuels the structures that allow systemic biases

a INDUCTIVE BIAS

In machine learning, inductive bias refers to the idea that an algorithm must make assumptions and therefore have biases in order to generalize from training data to novel data. This is because for any training dataset, there are an infinite number of functions to choose from that would fit this dataset. Therefore, in order to select which function to use to generalise from training to novel data, assumptions and therefore biases must come into play.

b UNDERFITTING

Underfitting in machine learning is related to a property called the bias-variance tradeoff. In this context, bias refers to when there are erroneous assumptions being made by the model which lead the model to miss important relations between the features and target labels. This bias error is known as underfitting.