**Machine Learning (WS 2025/2026)**

**Assignment Sheet 1 – Task 1.2**

**Submitted by: Suvansh Shukla Matriculation No: 256245**

**Assignment 1.2 Answer:**

Survivor bias: is a type of bias where focus on the “survivors” or observations which pass a certain test causes us to miss out on the observations which failed to pass. An example of this type of bias is looking at airplanes that returned successfully from air raids during world war 2. The people studying these airplanes initially aimed to strengthen areas of the planes with bullet holes, believing that securing these spaces would lead to better chances of surviving the hail of bullets. It was in fact the opposite that held true, i.e. areas of the planes which the bullets had missed that were areas which required strengthening.

Annotator bias: is a type of bias that stems from the source of data that is used to train models, i.e. if there is bias in the way the data source has been labelled then training a model on that data would reflect the bias when the model performs its task. An example of this type of bias would be that when training a model for perceived attractiveness. The bias may rear its head if the data used to train model only has data labelled by people who think only a certain type of person is attractive. If the model comes across a person that is not conventionally attractive, then the model would echo the bias of the individuals that had done the categorization.

Selection bias: is a type of bias that arises due to shortcomings in the selected data used for training a model. If a model uses a subset of data that doesn’t properly reflect the diversity of possibilities, then the model will be less accurate on data that was selected. An example of this would be training a facial recognition system using only faces of one ethnic background. Doing so would reduce the model’s accuracy when trying to identify faces of different ethnic backgrounds.

Biases like these need to be properly considered when training a model to ensure fairness, robustness and the reliability of the model. Not doing so can create blind spots in the model’s judgement and thus hampering its accuracy and functioning.