The code in this folder is used in the following order shown by their filenames. For clarification find the following:

1-Preprocessing is used to load the data from huggingface and preprocess it for the purposes of this thesis. It stores the datasets.

2-Selection is used to select the relevant datapoints for the following steps in the pipeline from the previous stored data and stores the new datasets.

3-News\_raining and 3-Twitter\_training are used to train Phi2 on both finalized datasets. It stores the models and the logs of the training.

4-Logging is used to visualize the logs of the training for the models from the stored logs.

5-GPT\_test, 5-Phi\_test, 5-Phi\_q\_test are used to generate the results with the trained and GPT models. The results are stored

6-Scoring is used to retrieve the metrics for the responses, the confusion matrices are stored.