This folder contains CSV files used in the evaluation of semantic similarity metrics in the context of our **RQ2**.

Below is a brief description of each file:

* RQ2\_Sampled\_Prompts: Contains the sample of test cases selected for evaluation in RQ2.
* RQ2\_Responses: Contains the LLM responses to all the test cases sampled.
* RQ2\_Responses\_with\_similarity: Contains the LLM responses to all the test cases sampled alongside results for the three (BERT cosine similarity, LSA, and LDA) similarity metrics for each.
* global\_fairness\_bug\_rates.csv: Contains the global number of fairness bugs detected, total test cases, and fail rates for each metric–threshold combination.
* Bias\_specific\_fairness\_bug\_rates.csv: Provides fairness bug counts and fail rates for each bias category, broken down by metric–threshold combination.
* best\_combinations\_per\_bias.csv: Lists the best-performing metric and threshold for each bias type based on the number of FAILs.
* Summary\_all\_testcases.csv: Aggregated global performance results, including total PASS and FAIL counts, for all tested metric–threshold configurations.
* Summary\_by\_bias\_type.csv: Detailed performance results split by bias type, including PASS and FAIL counts for each metric–threshold pair.
* Grouped\_stats.csv: Summary statistics (mean, median, std, etc.) of semantic similarity scores grouped by bias type and metric.
* Overall\_stats.csv: Statistical summary of similarity scores across all test cases, aggregated by metric.
* The folder “Visualizations” provides additional charts containing the distributions and bias-specific barplots for the three similarity metrics.