2023

Dataset for Gendered Language

Shweta Soundararajan
Technological University Dublin, shweta.x.soundararajan@mytudublin.ie

Recommended Citation

Using ChatGPT to Generate Gendered Language

Datasets for Gendered Language generated using ChatGPT

Files description:
Three datasets(set 1, set 2, set 3) generated using Gaucher lexicon are:
1. Gaucher_dataset_set_1.csv
2. Gaucher_dataset_set_2.csv
3. Gaucher_dataset_set_3.csv

Three datasets(set 1, set 2, set 3) generated using Cryan lexicon are:
1. Cryan_dataset_set_1.csv
2. Cryan_dataset_set_2.csv
3. Cryan_dataset_set_3.csv

Each CSV file has the following columns:
1. Index: Index of the sentence.
2. Sentences: List of sentences.
3. Labels: Labels assigned to the sentences. The labels are 'consistent' and 'contradictory'.
4. about_Male: Has two values, either 0 or 1. 1 indicates that the sentence is about male, and 0 indicates it's about female.
5. Masc_terms: Has two values, either 0 or 1. 1 indicates that the sentence is generated using male-coded words from the lexicon, and 0 indicates it's generated using female-coded words from the lexicon.

If you would like to use the dataset, please cite the paper mentioned below: