

2023

## Dataset for Gendered Language

Shweta Soundararajan

*Technological University Dublin*, [shweta.x.soundararajan@mytudublin.ie](mailto:shweta.x.soundararajan@mytudublin.ie)

Follow this and additional works at: <https://arrow.tudublin.ie/datas>

 Part of the [Computer Sciences Commons](#)

---

### Recommended Citation

S. Soundararajan, M. N. Jeyaraj and S. J. Delany, "Using ChatGPT to Generate Gendered Language," 2023 31st Irish Conference on Artificial Intelligence and Cognitive Science (AICS), Letterkenny, Ireland, 2023, pp. 1-8, doi: 10.1109/AICS60730.2023.10470830

This Dataset is brought to you for free and open access by the Data@TU Dublin at ARROW@TU Dublin. It has been accepted for inclusion in Datasets by an authorized administrator of ARROW@TU Dublin. For more information, please contact [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie), [aisling.coyne@tudublin.ie](mailto:aisling.coyne@tudublin.ie), [vera.kilshaw@tudublin.ie](mailto:vera.kilshaw@tudublin.ie).



This work is licensed under a [Creative Commons Attribution-Share Alike 4.0 International License](#).  
Funder: Technological University Dublin

# 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:

```
@INPROCEEDINGS{10470830,
  author={Soundararajan, Shweta and Jeyaraj, Manuela Nayantara and Delany, Sarah Jane},
  booktitle={2023 31st Irish Conference on Artificial Intelligence and Cognitive Science
(AICS)},
  title={Using ChatGPT to Generate Gendered Language},
  year={2023},
  volume={},
  number={},
  pages={1-8},
  keywords={Natural languages;Chatbots;Cognitive science;Artificial intelligence;natural
language processing;machine learning;large language models;ChatGPT;gendered
language;prompt engineering;zero-shot prompting},
  doi={10.1109/AICS60730.2023.10470830}}
```