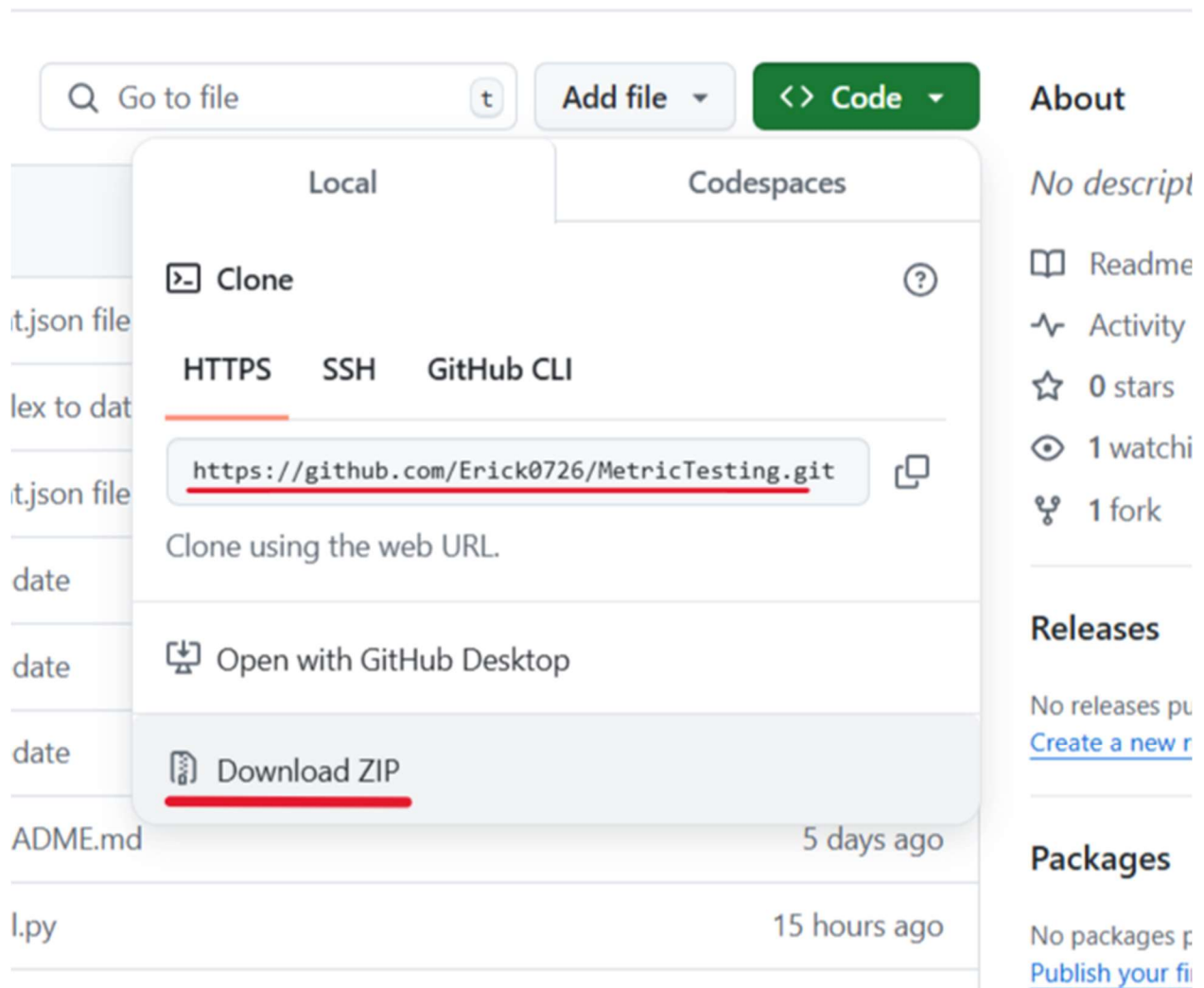
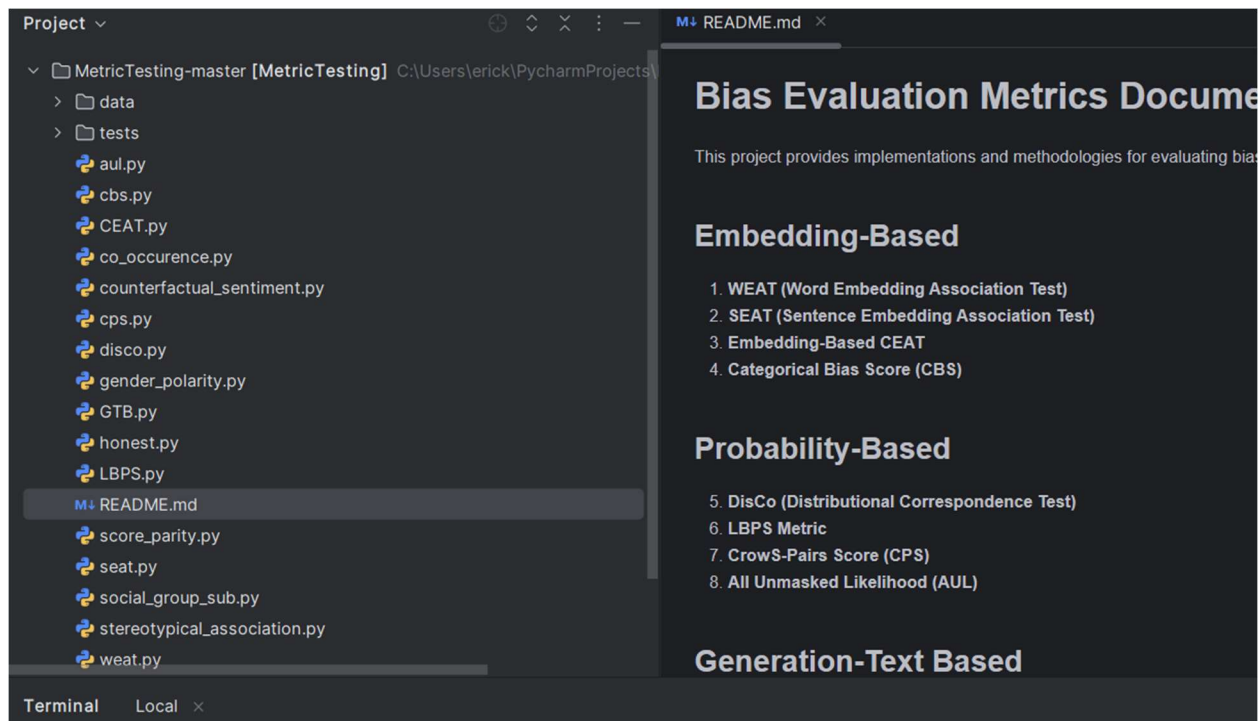


Installation/Maintenance

Go to <https://github.com/Erick0726/MetricTesting> and press the green button that says “code.” Then, either copy the link and clone the repository through git, or simply press “Download ZIP” to download a zip file of the project.



Open the project folder with a Python IDE (Our team used PyCharm) and it should look like this.



Then go to this link in line 30 of the `weat.py` file or in the `README.md` file under the WEAT section above “Steps”:

https://drive.google.com/file/d/1L3_k2SwdAyB3YPzlnzDrscjOjcv9jj0E/view?usp=sharing

Overview

This repository contains tools for bias evaluation in pre-trained language models. These metrics allow researchers and practitioners to detect and quantify biases related to social group representation in natural language processing.

WEAT: Word Embedding Association Test

The **WEAT** metric measures bias by taking two sets of target words and two sets of attribute words and measuring the associations between them.

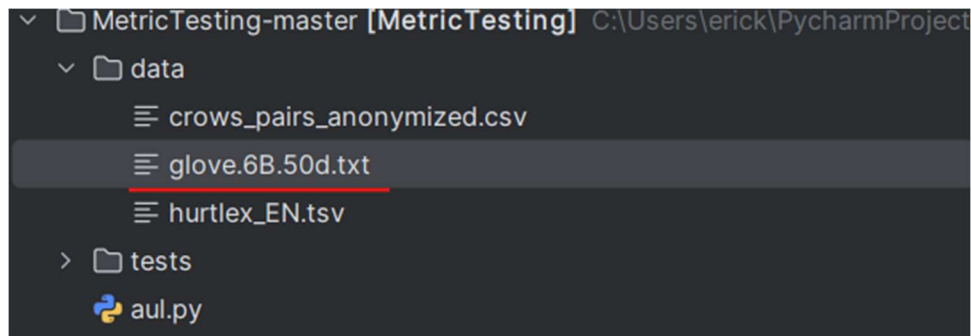
This link has the `glove.6B.50d.txt` file which is crucial for the WEAT metric:

https://drive.google.com/file/d/1L3_k2SwdAyB3YPzlnzDrscjOjcv9jj0E/view?usp=sharing

Steps

1. Prepare target and attribute word sets.
2. Generate word embeddings using a pre-trained language model.

Then, place the downloaded file inside the “data” folder in the repository



(Optional: you can use this link

<https://drive.google.com/file/d/0B7XkCwpI5KDYNINUTTISS2lpQmM/view?usp=sharing&resourcekey=0-wjGZdNAUop6WykTtMip30g> that's under the steps in the WEAT section of the README.md file and line 43 of the weat.py file and place it in the same folder. Then, comment out the code on lines 44-46 to test for the Word2Vec model)

Go to the terminal by pressing the button in blue shown below and install all of the following using the same method used in the screenshot: `pip install ...`

transformers, torch, pandas, numpy, scipy, regex, scikit-learn, matplotlib, seaborn, sentence-transformers, gensim

