

The following folder has 3 python scripts and then the train jsonl and cleaned csv file.

- The **train\_codeT5.py** file is where the model is trained and where things like the loss function can be changed
- The **data\_loader.py** is what was used to create the csv from the jsonl file. This file output the file named **output.csv**. It took about an hour and a half for 5,000 entries to run, so we can rerun this but the API to NVD severely limits. It can be run in batch or parallel. At least I couldn't figure it out
- The **Import Requests.py** file was used for the previous files debugging
- The **T5** folder is a virtual environment I was using
- The **codet5\_cvss\_model** is a folder where the trained model is saved out to from the **train\_codeT5.py** script. When I've been running test I just rm -r this directory so I don't run into any issues
- The **primevul\_train.jsonl** file is the data.
  - Previous 2 files had to be removed as they were too big to push to github