Dependencies

T5\_00\_EDA\_LO.ipynb

* Seaborn
* sentencePiece
* Need to remove lines 46 and 47, “loss\_per\_10\_steps” not defined or purposefully used

covar\_00\_bbcBatch\_LO.ipynb

* No issues ( so far)

triplets\_01\_clean\_LO.ipynb

* Textacy
* En\_core\_web\_lg (python -m spacy download en\_core\_web\_lg)
* Need to create a folder named ”Generation” in “data/BBC/”

triplets\_02\_tripleFormatting\_LO.ipynb

* (Error) Multiple uses of “tag\_and\_tokenize\_keywords” with 3 parameters even though function defined to only take one.
* (Fix) remove the 2 other arguments

GPT\_00\_fineTuning.ipynb

* [Errno 2] No such file or directory: '/kaggle/input/bbc-triplets-to-sentence/business\_gpt2\_truncated.txt'
* Fix, change data set file path to “./data/BBC/training/business\_gpt2/business\_gpt2\_truncated.txt”

T5\_01\_train8epoch.ipynb

* No issues (so far)

T5\_01\_train16epoch.ipynb

* No issues (so far)

GPT\_01\_textGenDistil\_LO.ipynb

* Fix, FOLDER\_TRIPLES = "./data/BBC/Training/business\_triples" instead
* Fix, FOLDER = "./results/Generated\_DistilGPT/"

GPT\_01\_textGenDistilFiltered\_LO.ipynb

* Fix, FOLDER\_TRIPLES = "./data/BBC/Training/business\_triples" instead
* En\_core\_web\_md (python -m spacy download en\_core\_web\_md)
* Fix, line 60, “idx” instead of “IDX”

GPT\_01\_textGenGPT2med\_LO.ipynb

* Fix, FOLDER\_TRIPLES = "./data/BBC/Training/business\_triples" instead
* Fix, FOLDER = "./results/Generated\_GPT2/"

ChatGPT\_01\_promptedTextGen.ipynb

* Openai
* Fix, sys\_prompt= SYS\_PROMPT
* Requires Api key for openai

T5\_02\_textGen8epoch.ipynb

* No issues

T5\_02\_textGen16Epoch\_LO.ipynb

* No issues

scoring\_BatchSimplicityIndex\_LO.ipynb

* No issues