

1. Personalized Writing Coach is a way users can write better. It will help correct the grammatical errors as well as offer some recommendations for improvement. Users can either upload a document or just type in words and the model will give them suggestions on what to do. It will also have the feature to provide alternative word choices which can help change tone of the sentence. It is trained using LLM and fine tuned especially for this grammar correction task. Several LLM models will be used and whichever performs the best will be chosen as the model for the app.

2. JFLEG: English Grammatical Error Benchmark

Collector(s): Bryant, Christopher; Felice, Mariano; and Briscoe, Ted

Year: 2017

Title of Dataset: *JFLEG: The Johns Hopkins Fluency-Extended GEC Corpus*

Version Number: Not explicitly versioned (Initial Release)

Publisher: Johns Hopkins University

DOI or URL:

<https://www.kaggle.com/datasets/thedevastator/jfleg-english-grammatical-error-benchmark?resource=download>

Study/Paper/Reason: The dataset was collected to address limitations of existing grammatical error correction (GEC) datasets by emphasizing fluency corrections over minimal edits. It provides multiple human-annotated corrections for each sentence, focusing on fluency and naturalness

3. Python, Pytorch , transformers
4. I will be writing the code to preprocess the data as well as tuning the model and evaluating the model too.
5. I think GPT-J should be the best choice here.
6. Different LLMS will be used which will also be a parameter as well as learning rate and weight decay will be the hyperparameters. Grid Search can be used to find the optimal one
7. Based on accuracy, it will be evaluated
8. <https://github.com/codets1989/mahto-B00115087-Spring-2025>