

GRAMMAR AUTO-CORRECT

DOMAIN: AI/ML

MENTOR: ANUSHKA
VASHNEY

TEAM MEMEBERS:

1. SHREEYUT MAHESWARI
2. SRIKANTH CHEBOLU
3. PIYUSH TIWARI
4. VIKASH KUMAR SINGH



PROBLEM STATEMENT

Given an input paragraph, detect the grammatical inconsistencies in it and rectify them. Bonus: Add suggestions to improve vocabulary and tone improvements by understanding context.

PROJECT TIMELINE

WEEK: 1-2

**Learnt neccesary
tech-stacks.**

WEEK: 3-4

**DATASET
FINDING AND
EXPLORATION**

WEEK: 5-6

**WORKING ON
MODIFICATIONS**

WEEK: 7

**Deploy Gradio
Website.**

DATASET FINDING AND TRAINING

Firstly, we researched various hugging face transformers and read various research papers on grammar auto-correct to generate precise corrected sentence. In our process we tried fine-tuning various hugging face models. After checking Metrics like Accuracy Score , analysing the quality of results, we chose grammarly coedit dataset for grammar related task and for finding the error in the sentences. We created our custom dataset with help of ai, we created of csv file and both the models were fine-tuned on t5-small transformer.



EXPLORATION

First we learn basic machine learning algorithms and libraries, then we studied about transformers, nlp, hugging face, then we tried various models, some of which were transformers trained from scratch on datasets related for the work and also tried using models directly from hugging face using pipeline which were already fine-tuned for the related work.



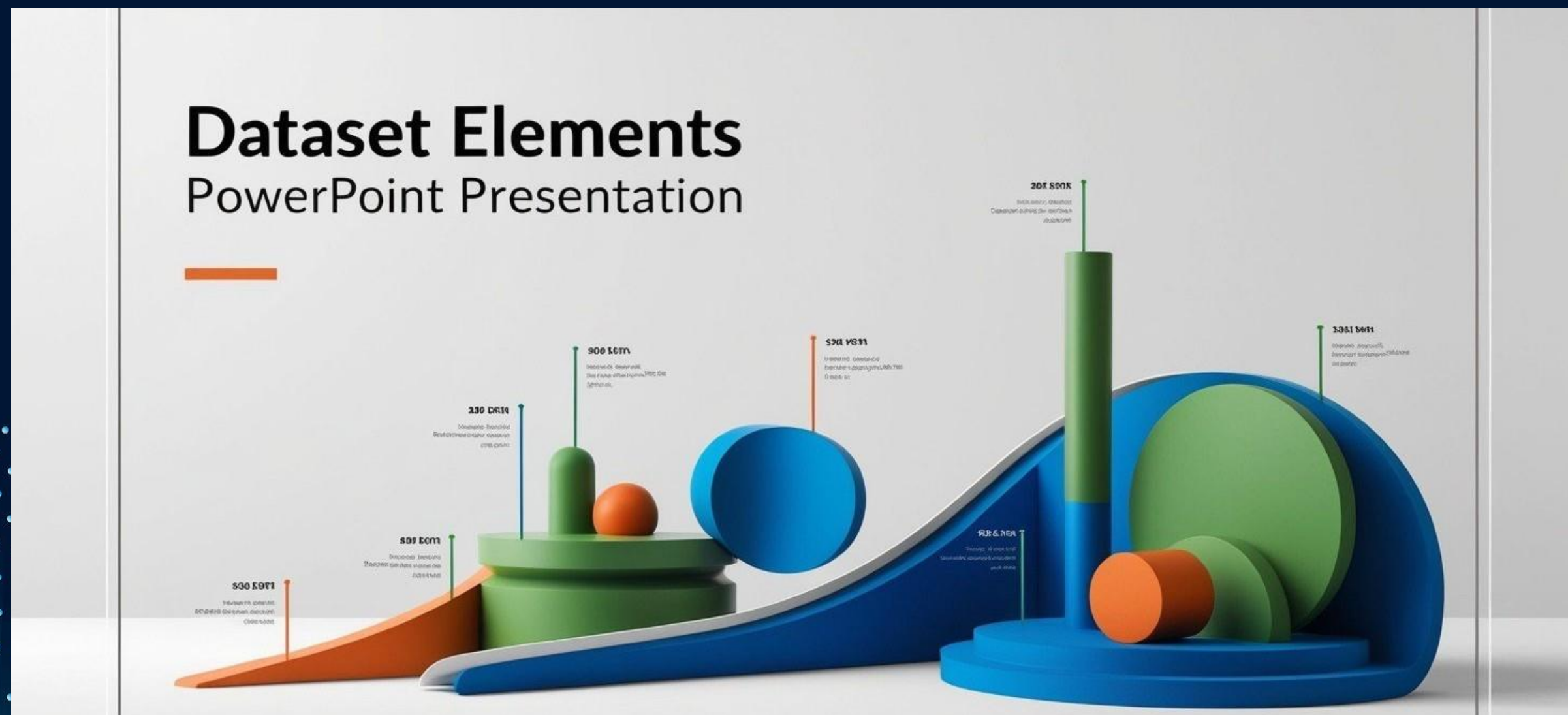
WHY GRAMMARLY COEDIT DATASET IS CHOSEN?

First we tried using datasets from kaggle but they did not contain large data and were not giving expected output. Then we used datasets CoNLL-2014, lesyaran c4 200m gec, liweli c4 200m, train 100k test 25k, jfleg, but we didn't choose this because ram was crashing while training them. Plus, this grammarly coedit dataset contained other task also like neutralization, simplification, paraphrasing the text.



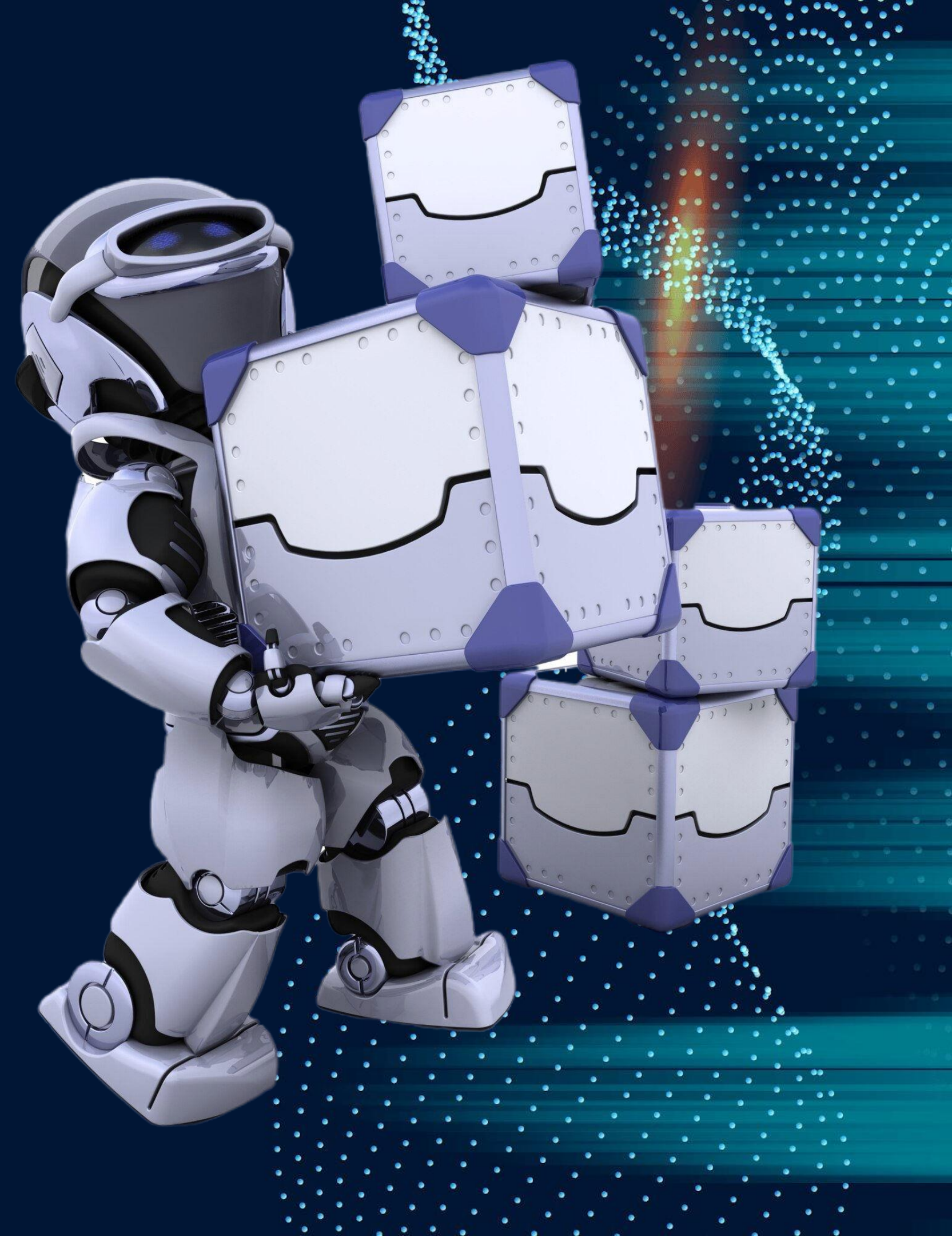
HOW CUSTOM DATASET WAS MADE

The sentences were searched from various english books, There were exercise regarding finding the grammatical error in the sentences and then it was made into csv file and then used for training purposes.



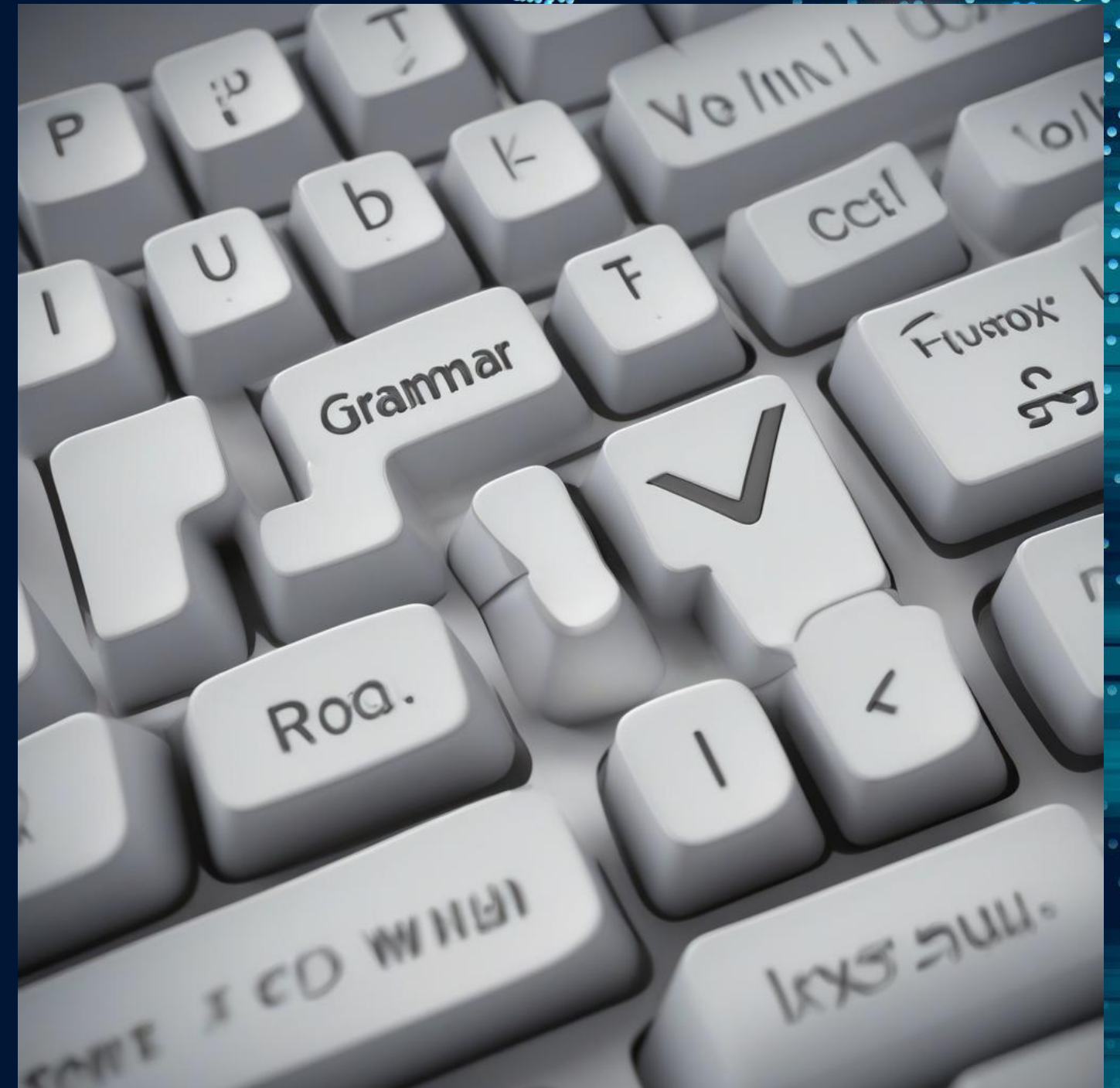
FINE- TUNING

For fine-tuning the coedit dataset was directly imported from hugging face , then data was preprocessed and padded. Then the tokenizer and model was imported from hugging face library and after fine-tuning, the model was saved, for using it for future purposes.



WORKING ON MODIFICATINS

The t5 small model was fine-tuned on the following parameters like standard learning rate was set and we adjusted batch size to process multiple inputs and improve efficiency. We adjusted the weight decay to avoid overfitting. Then saved to output directory named results.



RESULTS for the results Transformer

TOTAL TRAINING TIME: 2 HOURS

EPOCH	TRAINING LOSS	VALIDATION LOSS
1.	0.1686	0.245516
2.	0.1583	0.234669
3.	0.1556	0.232516

RESULTS for the gennnie Transformer

Training Loss	Epoch	Step	Validation Loss
0.1981	1.0	1280	0.0820
0.0933	2.0	2560	0.0617
0.084	3.0	3840	0.0576

[link to this model](#)

CREATING THE FINAL CODE

The fine-tuned model was reused on google collab with the help of hugging face and then the final code was prepared in which text and audio input was added and the model was returning precise and concise results.



Deploy Gradio Website

We deployed our model through gradio it takes text and audio input and gives fast and precise results.

Paragraph Processor

How to Use:

1.Enter Your Paragraph: Type or paste the paragraph you want to process into the text box provided., 2. Select Processing Type: Choose the type of processing you want to apply from the dropdown menu. The options include: (a)Correct and Paraphrase: Automatically corrects grammatical errors in your paragraph and provides a paraphrased version. (b)Neutralize: Adjusts the tone of your paragraph to be more neutral. (c)Cohesive: Enhances the coherence of your paragraph, making it flow better. (d)Simplify: Simplifies complex sentences for better readability and understanding., 3.Process and View Results: Click the submit button to process your paragraph. The results will be displayed, showing the refined version of your text.

Text Input

Select Processing Type

Select what you want to do with your text...

Select to see the error in text

☐ Grammatical Error

Clear

Submit

output

Flag

Examples

Text Input	Select Processing Type	Grammatical Error
Guy Leech is still a celebrity today. Guy Leech regularly appears on television as an advocate for health and fitness.	Cohesive	true
There was a girl standing there, staring into the room.	Simplify	true
chloroform "the molecular lifesaver" an article at oxford university providing interesting facts about chloroform.	Neutralize	true
melina is currently dating john cena, known as johnny nitro.	Neutralize	true
Injury, bad luck and inconsistency prevented him from ever winning the series. He was certainly good enough.	Cohesive	true
That's about 1.9 trillion by today's standards.	Simplify	true

Use via API

Built with Gradio

Conclusion

1.Error Detection: Identifies various types of grammatical errors such as punctuation mistakes, subject-verb agreement issues, incorrect word usage, etc

2.Performance: Achieves high accuracy in error detection and correction through pre-trained fine-tuning.

3.Modifications: Paraphrase text, Improved grammar text, Simplified text, Cohesivetext, and Neutralized text.

Thanks!

