GEN AI UNIT 1 PROJECT

NAME:DELISHA RIYONA DSOUZA
SRN: PES2UG23CS166
SECTION: C

PROJECT TITLE:
Category 5: Optimization & Tools (The "Utility" Agents)
Grammar & Spell Fixer Goal:
"I am go to school" -> Detects error and suggests "going".
Tech: pipeline('fill-mask') to predict the mathematically correct word in a sentence structure.

https://github.com/pes2ug23cs166/GENAI_UNIT1/tree/main/PROJECT

ABSTRACT
This project implements a simple Grammar and Spell Fixer using a pre-trained masked language model from Hugging Face. By masking a word in a sentence, the model analyzes the surrounding context and predicts the most suitable replacement. The system demonstrates how transformer-based models like BERT understand language structure and grammar without requiring any additional training or datasets. The solution is lightweight, reliable, and well-suited for beginner-level NLP applications.

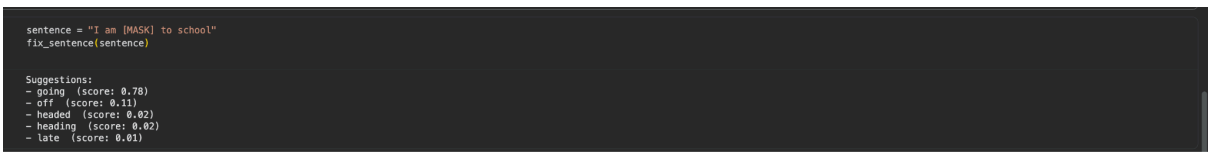
Short Documentation (What I understood & what I built)

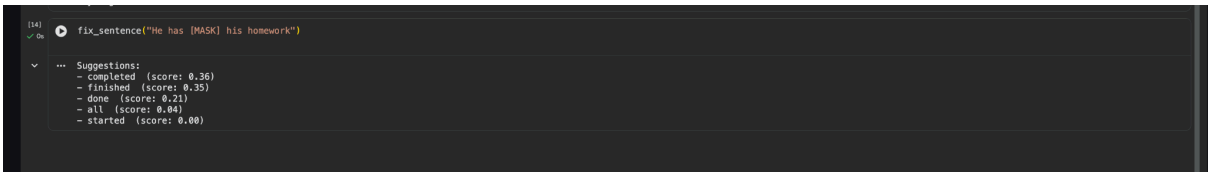
What I understood

Masked Language Models (MLMs) such as BERT are trained to predict missing words in a sentence based on context. By replacing an incorrect or unknown word with a special [MASK] token, the model can suggest grammatically and semantically appropriate alternatives. Since the output is restricted to ranked token predictions, the model avoids hallucination and produces stable results.

What I built

- Implemented a Grammar & Spell Fixer using Hugging Face's fill-mask pipeline.
- Used the bert-base-uncased model for accurate word prediction.
- Created functions to:
    - Display top replacement suggestions with confidence scores.
    - Automatically correct a sentence by replacing [MASK] with the best prediction.
- Executed the entire project in Google Colab for ease of setup and reproducibility.





CONCLUSION

The project successfully demonstrates how masked language models can be applied to basic grammar and spell correction tasks. Due to its simplicity, deterministic output, and absence of training requirements, this project serves as a strong introduction to transformer-based NLP systems.