# Project Report: Grammar Detection and Correction

## Abstract

This project focuses on the implementation of a Grammar Detection and Correction system utilizing Natural Language Processing (NLP) and deep learning techniques. Two distinct models are employed: a BERT-based classification model for grammar detection and a T5-based sequence-to-sequence model for grammar correction. The system demonstrates the capability to identify grammatical errors in sentences and provide corrected outputs. The project showcases a practical application of state-of-the-art NLP techniques in improving written communication.

## Introduction

Grammatical accuracy is a crucial aspect of effective communication. Manual proofreading is time-consuming and prone to errors, making automated grammar detection and correction an essential tool. This project integrates pre-trained NLP models like BERT and T5 to detect and correct grammar in text, leveraging their advanced language understanding capabilities.

## Objectives

1. Develop a grammar detection model using BERT to classify sentences as grammatically correct or incorrect.

2. Fine-tune a T5 model for correcting ungrammatical sentences.

3. Demonstrate the system's effectiveness with a dataset of ungrammatical and corrected sentences.

## Methodology

The methodology involves the following steps:

1. \*\*Data Preparation\*\*: The dataset contains ungrammatical sentences, their corrected counterparts, and error types. The data was preprocessed to handle missing values and convert categorical labels for grammar detection.

2. \*\*Model Development\*\*:  
 - Grammar Detection: A BERT-based model was fine-tuned for binary classification of sentences.  
 - Grammar Correction: A T5 model was fine-tuned to generate corrected sentences.

3. \*\*Training\*\*: The models were trained and evaluated using appropriate metrics and validation strategies.

## Results

The models achieved satisfactory performance in both detection and correction tasks. For the grammar detection task, evaluation metrics like precision, recall, and F1-score were used. For the correction task, the T5 model was able to generate corrected sentences, demonstrating its ability to understand and rectify grammatical errors.

Sample Outputs:  
 - Input: 'She go to school'  
 Detected: Grammatical Error  
 Corrected: 'She goes to school'  
 - Input: 'I likes apples'  
 Detected: Grammatical Error  
 Corrected: 'I like apples'

## Tools and Technologies

1. \*\*Programming Language\*\*: Python  
2. \*\*Libraries\*\*: Transformers, PyTorch, Scikit-learn, Pandas, NLTK  
3. \*\*Models\*\*: BERT (bert-base-uncased) for classification, T5 (t5-small) for sequence-to-sequence tasks  
4. \*\*Framework\*\*: Hugging Face Transformers

## Challenges

1. Fine-tuning pre-trained models required careful selection of hyperparameters to avoid overfitting.  
2. Handling long or complex sentences while maintaining model efficiency.  
3. Dataset limitations and ensuring diverse grammatical error coverage.

## Conclusion

This project successfully demonstrates the application of BERT and T5 models for grammar detection and correction tasks. The results highlight the potential of pre-trained language models in improving text quality and aiding language learning.

## Future Work

1. Expanding the dataset to include more diverse grammatical errors.  
2. Optimizing models for faster inference.  
3. Developing a user-friendly interface for real-time grammar detection and correction.