A Research on Grammar Correction Tool with Casual to Formal Sentence Conversion

[1] **Ms. Pallavi Shejwal**, Assistant Professor, Information Technology Engineering Department,
Progressive Education Society's Modern College of Engineering, Pune - 411 005 (Maharashtra)(India)

[2]Student VIII Semester B.E(Information Technology),
Gauri Ade[1], Shubham Barge[2], Om Bothe[3], Prajwal Deokar[4]

Progressive Education Society's Modern College of Engineering, Pune - 411 005 (Maharashtra)(India)

Abstract - Most of the times, we as human tend to mistakes in different fields. Talking about the correction we need a tool which can help the people to correct it accordingly. GramFormer basically deals with correction of words by suggesting the next words which are grammatically correct and are constructed properly. Apart from that it also consists of a kind of conversion of sentence from Casual to formal. In daily life, we tend to write sentences which are correct but are in informal way. But in corporate word we cannot use that or some might feel bothered to write it correctly. Overall, GramFormer provides a system to the user to grammatically correct with proper sentence construction. The fundamental of this model is based on neural network and Representation Learning.
The words a person chooses to use are a key indicator of their sense of style. Although everyone expresses themselves in a unique way, people adapt their speech and writing to the social setting, the audience, the interlocutor, or the level of formality of the situation. Text style transfer is the process of adapting and/or changing the stylistic approach used to write a sentence while maintaining the original sentence's meaning.

Key Words*: *Natural Language Processing, Deep Learning, Sentence Conversion, Neural Networks.

1. INTRODUCTION

The key to fostering positive relationships with others and conducting organised communication is effective communication. It may result in a productive workplace that is healthy, which will eventually increase productivity. Therefore, it's important to convey our demands clearly to others in order for them to respond appropriately. People often converse casually with professionals due to the change in people's lifestyles.

People frequently struggle to compose proper phrases. There may be a number of causes behind that. Usually, a lot of people write precisely correct sentences without any mistakes. Writing with clarity, economy, and elegance does take effort. A statement that is inaccurate can give off an unexpected impression and result in unanticipated outcomes. People may therefore require a tool to address this, and natural language processing can be of assistance.
We can utilise NLP to make this issue less complicated. It is a subset of Artificial Intelligence, also known as AI, and stands for Natural Processing Language. It offers computers the ability to comprehend text or spoken words in the same way that people can.
Around the world, sentences are written in a variety of styles depending on the area, age, and gender. We, therefore, require a tool that can break down the linguistic barrier between them in order to successfully communicate.

2. Problem Statement: Building a framework for detecting, highlighting, and correcting grammatical errors in natural language text. It is going to deal with sentence conversion from casual to formal. It is a web app that we have built for creating an interface to detect grammar mistake and correct it.

3. Objective: The main objective of the project is to handle frequent simple grammatical errors and identify wrong spelling which will help students to correct low-level errors in their written work. Also the objective of casual to formal conversion is to increase understating level of language and make it professional in talks.

4. Datasets :

1. GYAFC dataset of informal and formal sentence - GYAFC has been proposed in for the research of formality style transmission for the English language (Rao and Tetreault, 2018). The training dataset consists of 53 000 pairs of informal utterances and their translated formal equivalents. The same procedure, but with stricter quality control, was utilised to produce tuning and testing.

2. C4 200M Grammar Error Correction dataset - Grammar Error Correction synthetic dataset consisting of 185 million sentence pairs. Grammatical Error Correction (GEC) is the task of correcting different kinds of errors in text such as spelling, punctuation, grammatical, and word choice errors.

5. Working: The foundation of GramFormer is to provide a platform for the conversion of informal sentences into formal ones using Natural Language Processing and to fix English in the proper manner using neural networks. To improve the model and obtain more accurate results, experts in the field of machine learning are working on this grammatical correction. Additionally, this interface will change informal sentences into professional ones.At first, it used statistical phrase-based machine translation (SMT) techniques that employed conventional statistical approaches.

Transformer model

A neural network called a transformer model follows relationships in sequential input, such as words in this sentence, to learn context and subsequently meaning.

Transformer models use an expanding collection of mathematical approaches known as attention or self-attention to find minute relationships between even far-flung data pieces in a series.

Transformers are one of the newest and most potent kinds of models created to date. They were first introduced in a 2017 study from Google. They are at the forefront of a wave of machine learning innovations that some have called transformer AI

words in this sentence, to learn context and subsequently meaning.

Transformer models use an expanding collection of mathematical approaches known as attention or self-attention to find minute relationships between even far-flung data pieces in a series.

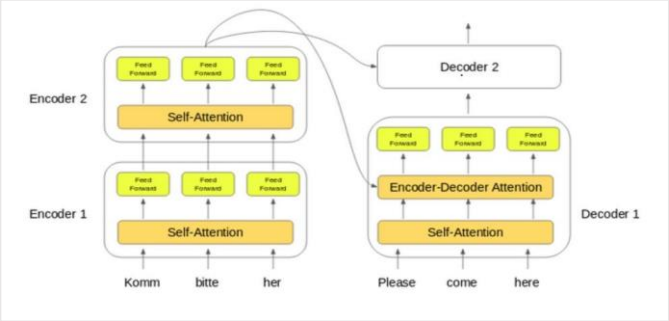
Transformers are one of the newest and most potent kinds of models created to date. They were first introduced in a 2017 study from Google. They are at the forefront of a wave of machine learning innovations that some have called transformer AI.

Autoencoders

The input and output of feedforward neural networks that use autoencoders are identical. They reduce the input's dimension before using this representation to recreate the output. The code, also known as the latent-space representation, is an efficient "summary" or "compression" of the input. Encoder, code, and decoder are the three parts of an autoencoder. The input is compressed by the encoder, which also creates a code. The decoder then reconstructs the input exclusively using the code.

T5 Model

Text-to-Text-Transfer-Transformer model proposes reframing all NLP tasks into a unified text to-text-format where the input and output are always text strings. This model generates a revised version of inputted text with the goal of containing fewer grammatical errors. It was trained with Happy Transformer using a dataset called JFLEG.



Training a T5 Model:

The initial step in transfer learning is pre-training, which involves training a model on a self-supervised task using a vast amount of unlabelled text data. In comparison to merely training on the small, labelled datasets without pre-training, the model is then fine-tuned (trained) on smaller labelled datasets customised to specific tasks.

All NLP tasks are presented in text-to-text format in the T5 model, which is a significant difference. A single class label or a span of text from the input is output by BERT-like models, which, on the other hand, use a text sequence

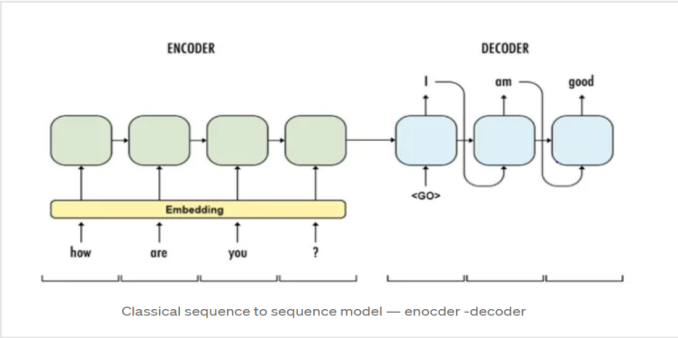
as input. In order to adapt a BERT model for a specific purpose, an appropriate output layer is added on top of the transformer model. For classification jobs, for instance, a straightforward linear classification layer is added. T5, on the other hand, rejects this strategy and reframes any NLP problem so that the input and output are both text sequences. As a result, all NLP tasks may be performed using the same T5 model without the need for any additional architecture modifications.

Implementation Of Casual to Formal Using Attention Layer

- Importing the T5 modules from hugging face/transformers:

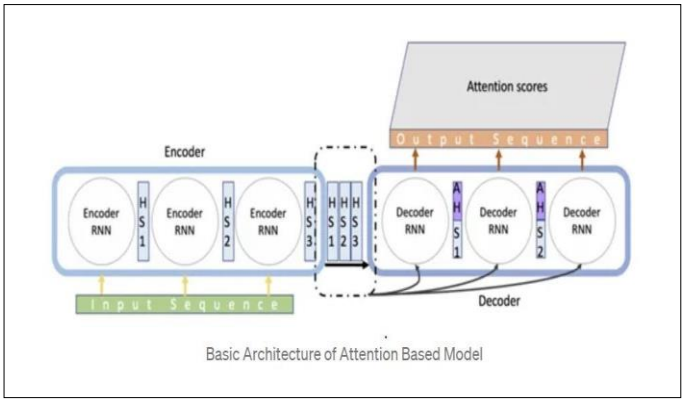
Hugging Face – Hugging Face, the AI community building the future, is a large open-source community that builds tools to enable users to build, train, and deploy machine learning models based on open-source code and technologies. One of the major advantages of using Hugging Face's tools is that you can reduce training time, resources and environmental impact of creating and training a model from scratch.

Encoder Decoder Algorithm:



One of the main weaknesses of this network is its inability to extract strong contextual relations from long semantic sentences. In other words, if a piece of long text contains context or relations within its substrings, a basic seq2seq model [short for sequence to sequence] will not be able to identify those contexts, which will eventually lead to a decline in model performance and accuracy.
In light of this, it became clear that this network needed to be upgraded further so that significant contextual relations could be examined and our model could produce and offer more accurate predictions.

Encoder-Decoder Algorithm With Attention Based Model



The encoder-decoder model is a way of organizing recurrent neural networks for sequence-to-sequence prediction problems or challenging sequence-based inputs like texts [sequence of words], images [sequence of images or images within images] to provide many detailed predictions.
The network of sequence-to-sequence models that currently addresses this restriction has been upgraded to include attention. Its ability to find significance in

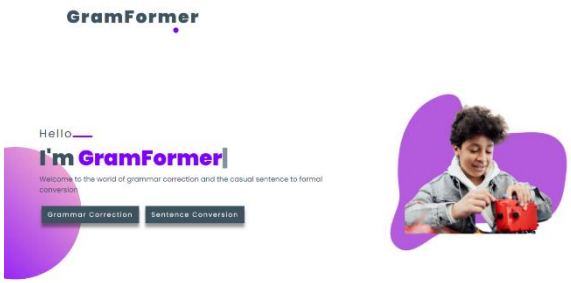
sequences is the main reason attention is termed what it is.

The first way it functions is by giving the decoder a learning mechanism whereby the decoder can comprehend where to really give the future encoding network more "attention" when forecasting outputs at each time step in the output sequence.

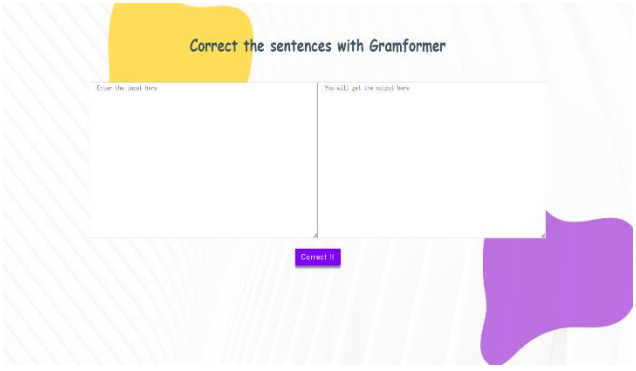
We can think of the idea of liberating the current encoder-decoder architecture from the fixed-short-length internal representation of text by applying the attention mechanism. This is accomplished by keeping the intermediate outputs from the encoder LSTM network from each step of the input sequence, which correspond to a certain level of significance, and simultaneously training the model to learn and pay attention only to these intermediate elements before relating them to elements in the output sequence.

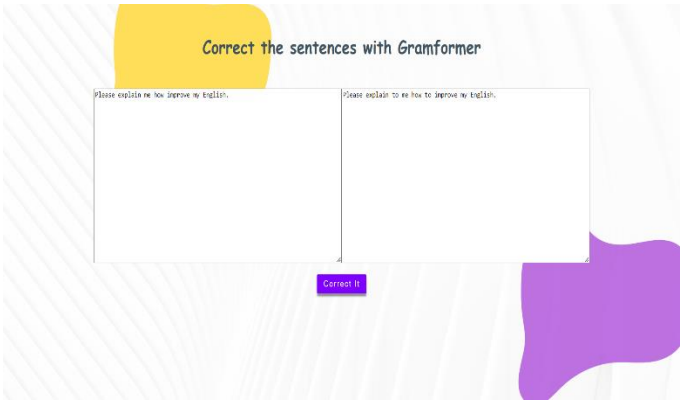
Result:

1. Home Page:

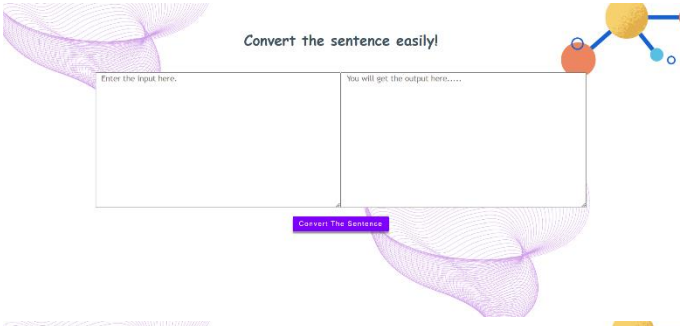


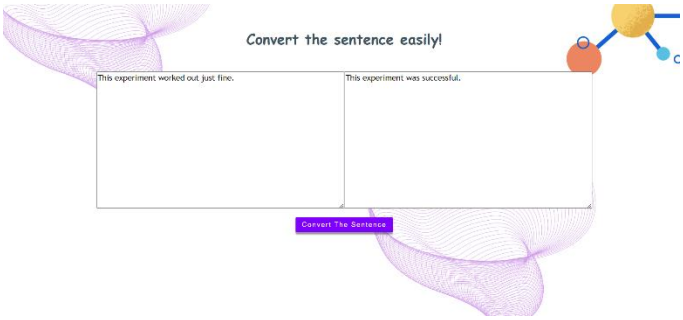
2. Grammar Correction:





3. Casual to Formal Conversion:





Conclusion: In recent trends, due to advanced technology it has led us to study about this particular topic with enthusiasm. During survey we got to know about different problems these models faced. By performing certain algorithms, these mistakes can be removed slightly which is definitely going to increase the efficiency of the system. Accurate use of grammar and writing is an essential aspect of any writing. Today various online grammar spelling checker tools are available for improving writing quality. Talking about the models which are used is extension of Deep Learning. It is basically a subset of Machine Learning which contains layers of neural networks. We immensely hope that this paper would help who want to study it further and want to excel in it. Transformer Model and Autoencoders are the fields that can make progress together and can provide

good platform the user. Neural network is vast to learn but apart from that it helps World to improvise various things and to get accurate results in less time.

References:

[1] A Research on Online Grammar Checker System Based on Neural Network .
Senyue Hao and Gang Hao 2020 J. Phys.: Conf. Ser. 1651 012135

[2] A Review of Text Style Transfer using Deep Learning. Faculty of Computer Science and Engineering, Ss.
Cyril and Methodius University - Skopje, North Macedonia

[3] Usage of Grammarly – Online Grammar and Spelling Checker Tool

Gain, Anita; Rao, Mahabaleshwara; and Bhat, Shivananda, "Usage of Grammarly – Online Grammar and Spelling Checker Tool at the Health Sciences Library, Manipal Academy of Higher Education, Manipal: A Study" (2019). Library Philosophy and Practice.

[4] Detecting Text Formality: A Study of Text Classification Approaches

Daryna Dementieva1 , Ivan Trifinov2 , Andrey Likhachev3 , and Alexander Panchenko1

[5] Deep Learning for Text Style Transfer: A Survey-2021

Di Jin∗, Zhijing Jin, Zhiting Hu, Olga Vechtomova, Rada Mihalcea