

as word prediction, and sentence topic prediction. [47] In order to observe the word arrangement in forward and backward direction, bi-directional LSTM is explored by researchers [59]. In case of machine translation, encoder-decoder architecture is used where dimensionality of input and output vector is not known. Neural networks can be used to anticipate a state that has not yet been seen, such as future states for which predictors exist whereas HMM predicts hidden states.

#### (d) BERT

Bi-directional Encoder Representations from Transformers (BERT) is a pre-trained model with unlabeled text available on BookCorpus and English Wikipedia. This can be fine-tuned to capture context for various NLP tasks such as question answering, sentiment analysis, text classification, sentence embedding, interpreting ambiguity in the text etc. [25, 33, 90, 148]. Earlier language-based models examine the text in either of one direction which is used for sentence generation by predicting the next word whereas the BERT model examines the text in both directions simultaneously for better language understanding. BERT provides contextual embedding for each word present in the text unlike context-free models (word2vec and GloVe). For example, in the sentences “he is going to the riverbank for a walk” and “he is going to the bank to withdraw some money”, word2vec will have one vector representation for “bank” in both the sentences whereas BERT will have different vector representation for “bank”. Muller et al. [90] used the BERT model to analyze the tweets on covid-19 content. The use of the BERT model in the legal domain was explored by Chalkidis et al. [20].

Since BERT considers up to 512 tokens, this is the reason if there is a long text sequence that must be divided into multiple short text sequences of 512 tokens. This is the limitation of BERT as it lacks in handling large text sequences.

## 5 Evaluation metrics and challenges

The objective of this section is to discuss evaluation metrics used to evaluate the model’s performance and involved challenges.

### 5.1 Evaluation metrics

Since the number of labels in most classification problems is fixed, it is easy to determine the score for each class and, as a result, the loss from the ground truth. In image generation problems, the output resolution and ground truth are both fixed. As a result, we can calculate the loss at the pixel level using ground truth. But in NLP, though output format is predetermined in the case of NLP, dimensions cannot be specified. It is because a single statement can be expressed in multiple ways without changing the intent and meaning of that statement. Evaluation metrics are important to evaluate the model’s performance if we were trying to solve two problems with one model.

- a) BLEU (BiLingual Evaluation Understudy) Score: Each word in the output sentence is scored 1 if it appears in either of the reference sentences and a 0 if it does not. Further the number of words that appeared in one of the reference translations is divided by the total number of words in the output sentence to normalize the count so that it is always between