d) There are around 160,000 sentence pairings in the IWSLT 14. The dataset includes descriptions in English-German (En-De) and German-English (De-En) languages. There are around 200 K training sentence sets in the IWSLT 13 dataset.
e) The IIT Bombay English-Hindi corpus comprises parallel corpora for English-Hindi as well as monolingual Hindi corpora gathered from several existing sources and corpora generated over time at IIT Bombay's Centre for Indian Language Technology.

4. Question Answering System: Question answering systems provide real-time responses which are widely used in customer care services. The datasets used for dialogue system/ question answering system are as follows:

a) Stanford Question Answering Dataset (SQuAD): it is a reading comprehension dataset made up of questions posed by crowd workers on a collection of Wikipedia articles.
b) Natural Questions: It is a large-scale corpus presented by Google used for training and assessing open-domain question answering systems. It includes 300,000 naturally occurring queries as well as human-annotated responses from Wikipedia pages for use in QA system training.
c) Question Answering in Context (QuAC): This dataset is used to describe, comprehend, and participate in information seeking conversation. In this dataset, instances are made up of an interactive discussion between two crowd workers: a student who asks a series of open-ended questions about an unknown Wikipedia text, and a teacher who responds by offering brief extracts from the text.

The neural learning models are overtaking traditional models for NLP [64, 127]. In [64], authors used CNN (Convolutional Neural Network) model for sentiment analysis of movie reviews and achieved 81.5% accuracy. The results illustrate that using CNN was an appropriate replacement for state-of-the-art methods. Authors [127] have combined SST and Recursive Neural Tensor Network for sentiment analysis of the single sentence. This model amplifies the accuracy by 5.4% for sentence classification compared to traditional NLP models. Authors [135] proposed a combined Recurrent Neural Network and Transformer model for sentiment analysis. This hybrid model was tested on three different datasets: Twitter US Airline Sentiment, IMDB, and Sentiment 140: and achieved F1 scores of 91%, 93%, and 90%, respectively. This model's performance outshined the state-of-art methods.

Santoro et al. [118] introduced a rational recurrent neural network with the capacity to learn on classifying the information and perform complex reasoning based on the interactions between compartmentalized information. They used the relational memory core to handle such interactions. Finally, the model was tested for language modeling on three different datasets (GigaWord, Project Gutenberg, and WikiText-103). Further, they mapped the performance of their model to traditional approaches for dealing with relational reasoning on compartmentalized information. The results achieved with RMC show improved performance.

Merity et al. [86] extended conventional word-level language models based on Quasi-Recurrent Neural Network and LSTM to handle the granularity at character and word level. They tuned the parameters for character-level modeling using Penn Treebank dataset and word-level modeling using WikiText-103. In both cases, their model outshined the state-of-art methods.

Luong et al. [70] used neural machine translation on the WMT14 dataset and performed translation of English text to French text. The model demonstrated a significant improvement