1)NLP stands for Natural Language Processing, which is a branch of artificial intelligence that focuses on the interaction between computers and humans through natural language.

Deep learning has revolutionised the field of natural language processing (NLP) by allowing for the creation of highly accurate and flexible models for tasks such as language translation, sentiment analysis, and language understanding.

2)Word embeddings are used to convert words or phrases into numerical representations in NLP by representing words as dense vectors in a high-dimensional space, capturing semantic and syntactic relationships between words.

It is an advancement in NLP that has improved

the ability of computers to understand text-based content in a better way.

3)RNNs can be harder to train due to issues like vanishing gradients but are optimized for sequence data. In essence, FNNs handle static mappings, while RNNs excel at sequences. Feedforward Neural Networks (FNNs) process information linearly from input to output, ideal for tasks like image classification.

If the task involves static or fixed data, such as images, feedforward neural networks are more suitable but task involves sequential or temporal data, such as text or speech, recurrent neural networks are more suitable.

4)The attention mechanism aligns the input and output sequences. It helps to pay attention to the most relevant information in the source sequence. Attention mechanisms improves performance and accuracy on long and complex sequences. This is achieved by reducing information loss and the burden on the hidden states.

Attention mechanisms enhance deep learning models by selectively focusing on important input elements, improving prediction accuracy and computational efficiency.

5)The Transformer architecture introduced the concept of attention mechanism in deep learning, which allows the model to focus on different parts of the input sequence when making predictions, leading to significant improvements in tasks like language translation and language understanding.

Using transformers can improve meaning clarity

The Transformer architecture was originally designed for translation. During training, the encoder receives inputs (sentences) in a certain language, while the decoder receives the same sentences in the desired target language.

6)The primary drawback of deep learning models is that they only learn from observations.

* They therefore only know the information included in the training data. NLP models must understand the broader context, including idiomatic expressions, cultural references, etc. This requires advanced algorithms and extensive, diverse training data.
* Biases are another significant concern with deep learning algorithms. When a model is trained on biased data, it will replicate similar biases in its predictions. This can lead to unfair or discriminatory outcomes, particularly in sensitive applications like hiring or law enforcement.
* Neural networks tend to overfit the data they are trained on, which means they perform well on the training data but poorly on unseen or novel data. Thus major challenge faced is how fluid and inconsistent language can be which can lead to inaccurate result.

7)Transfer Learning in NLP works by leveraging the knowledge and representations learned by a pre-trained language model.

Pre-trained layers that comprehend particular dialects or vocabulary or simultaneously training a model to recognize various linguistic components.

Models for translating between languages can also be modified via transfer learning. Due to the prevalence of digitized English language materials, models can be trained on a sizable dataset before having their components translated to a model for a different language.

Example: a model that has been trained on news articles to generate summaries of blog posts.

8)Tokenization is the process of breaking down text into words, phrases, symbols, or other meaningful elements, which is a fundamental step in processing textual data in NLP.

Stemming is a normalization technique where list of tokenized words are converted into shorten root words to remove redundancy.

1. In NLP, getting base word is important to know whether the word is positive or negative. Stemming is used to get that base word.
2. Before, stemming, tokenization is done so as to break text into chunks. In this case, paragraph to sentences for easy computation.
3. After that, stopwords (such as the, and, etc) are ignored and stemming is applied on all other words. Finally, stem words are joined to make a sentence.

9) Perplexity is an evaluation metric that measures the quality of language models by evaluating how well a probability model can predict a sample.

It is often used in the context of language modelling, where the goal is to predict the next word in a sequence of words. A lower perplexity indicates that the model is better at predicting the sample. It can be used to compare different language models or to tune the parameters of a language model.

10) Named Entity Recognition (NER) is a sub-task of information extraction in Natural Language Processing (NLP) that classifies named entities into predefined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, and more.

Recurrent Neural Networks (RNN) and transformers are ideal due to their ability to model long-term dependencies in text. They're ideal for large-scale tasks with abundant training data.

11) Word2Vec is technique in natural language processing (NLP) that is used to represent words as continuous vector spaces. The underlying idea is that words with similar meanings should have similar vector representations. In Word2Vec every word is assigned a vector. Word2Vec models can perform better with larger datasets.

GloVe word embedding is to use statistics to derive the link between the words. GloVe embeddings enable algorithms to compute semantic similarity between words or phrases and solve analogy-based tasks. It is more computationally efficient than other word embedding techniques, such as Word2Vec.

12) Common evaluation metrics for assessing the performance of NLP models

1. Accuracy: refers to the fraction of times the model makes a correct prediction as compared to the total predictions it makes.
2. Precision: evaluates the percent of true positives identified given all positive cases.
3. Area Under Curve: combines true positives vs false positives as threshold for prediction is varied. Used to measure the quality of a model independent of prediction threshold, and to find the optimal prediction threshold for a classification task.
4. Root mean squared error — used to capture a model’s performance in a real-value prediction task.

*13)* Backpropagation through time (BPTT) is a method used in recurrent neural networks (RNNs) to train the network by backpropagating errors through time.It is used to train recurrent neural networks (RNNs) and other models with sequential data like a time series.

The general algorithm is as follows:

1. Present a training input pattern and propagate it through the network to get an output.
2. Compare the predicted outputs to the expected outputs and calculate the error.
3. Calculate the derivatives of the error with respect to the network weights.
4. Adjust the weights to minimize the error.
5. Repeat.