1. **What are Sequence-to-sequence models?**

Sequence to Sequence (often abbreviated to seq2seq) models is a special class of Recurrent Neural Network architectures that we typically use (but not restricted) to solve complex Language problems like Machine Translation, Question Answering, creating Chatbots, Text Summarization, etc

1. **What are the Problem with Vanilla RNNs?**

A major issue with the vanilla RNN is that they suffers from vanishing/exploding gradients similarly to issues with deep feedforward networks. At each timestep, the hidden state ht is multiplied by W, at the last timestep, the value of ht is effectively multiplied by W.

1. **What is Gradient clipping?**

Gradient clipping is a technique to prevent exploding gradients in very deep networks, usually in RNN.

1. **Explain Attention mechanism**

The attention mechanism is a part of a neural architecture that enables to dynamically highlight relevant features of the input data, which, in NLP, is typically a sequence of textual elements. It can be applied directly to the raw input or to its higher level representation.

1. **Explain Conditional random fields (CRFs)**

Conditional random fields (CRFs) are a class of statistical modeling methods often applied in pattern recognition and machine learning and used for structured prediction. Whereas a classifier predicts a label for a single sample without considering "neighbouring" samples, a CRF can take context into account .

1. **Explain self-attention**

Self-attention, also known as intra-attention, is an attention mechanism relating different positions of a single sequence in order to compute a representation of the same sequence. It has been shown to be very useful in machine reading, abstractive summarization, or image description generation.

1. **What is Bahdanau Attention?**

Bahdanau attention mechanism proposed an attention mechanism that learns to align and translate jointly. It is also known as Additive attention as it performs a linear combination of encoder states and the decoder states.

1. **What is a Language Model?**

A Language Model ( LM ) captures the probability of a sequence of words in the language. Equivalently, it tells us how likely a given word will follow a sequence of words. Traditionally, N-gram models and their variants were used as language models.

1. **What is Multi-Head Attention?**

Multi-head Attention is a module for attention mechanisms which runs through an attention mechanism several times in parallel. The independent attention outputs are then concatenated and linearly transformed into the expected dimension.

1. **What is Bilingual Evaluation Understudy (BLEU)**

BLEU (BiLingual Evaluation Understudy) is a metric for automatically evaluating machine-translated text. The BLEU score is a number between zero and one that measures the similarity of the machine-translated text to a set of high quality reference translations.