1)What are the pros and cons of using a stateful RNN versus stateless RNN ?

Ans : Stateless: In the stateless LSTM configuration, internal state is reset after each training batch or each batch when making predictions.

Stateful: In the stateful LSTM configuration, internal state is only reset when the reset\_state() function is called.

Setting an RNN to be stateful means that it can build a state across its training sequence and even maintain that state when doing predictions. The benefits of using stateful RNNs are smaller network sizes or lower training times.

2)why do people use encoder decoder RNN rather than plain sequence to sequence RNNs for automatic translation ?

Ans : seq-2-seq RNNs translate one word at a time

encoder-decoder RNNs read & translate a sentence at a time

3)How can you deal with Variable length input sequence? What about variable length output sequence ?

Ans : The first and simplest way of handling variable length input is to set a special mask value in the dataset, and pad out the length of each input to the standard length with this mask value set for all additional entries created. Then create a Masking layer in the model, placed ahead of all downstream layers.

4) what is beam search and why would you use it ?what tool can you use to implement it ?

Ans : beam search is a heuristic search algorithm that explores a graph by expanding the most promising node in a limited set. Beam search is an optimization of best-first search that reduces its memory requirements.

Greedy search & beam search algorithm we used to implement & encoding decoding.

5)what is an Attention mechanism ?how does it help

Ans : The Attention mechanism in Deep Learning is based off this concept of directing your focus, and it pays greater attention to certain factors when processing the data.

6)what is the most important layer in the transformer architecture ?what is its purpose ?

Ans : The most important part here is the “Residual Connections” around the layers. This is very important in retaining the position related information which we are adding to the input representation embedding across the network. The network displayed catastrophic results on removing the Residual Connections

7)when would you need to use sampled softmax ?

Ans : Sampled Softmax is a drop-in replacement for softmax cross entropy which improves scalability e.g. when there are millions of classes.It is very similar to Noise Contrastive Estimation (NCE) and Negative Sampling, both of which are popular in natural language processing, where the vocabulary size can be very large.so we need to used sampled softmax