
Attention Mechanism and Transformers
Video 1: Introduction to seq2seq Models
Analytics Vidhya

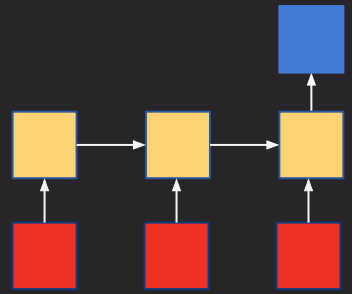
Analytics Vidhya

In Air

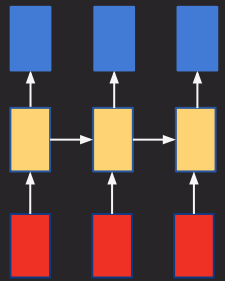
Types of RNN

Analytics Vidhya

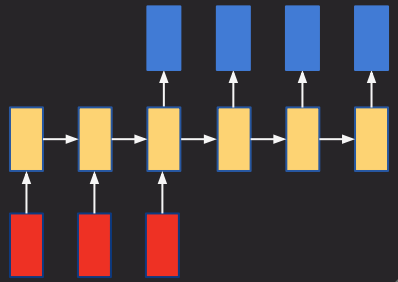
Many to One



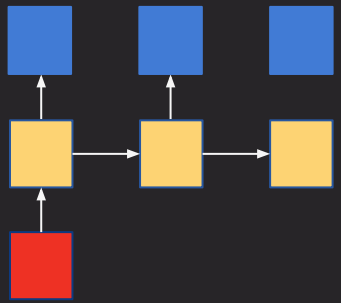
Many to Many



Many to Many



One to Many



One to One

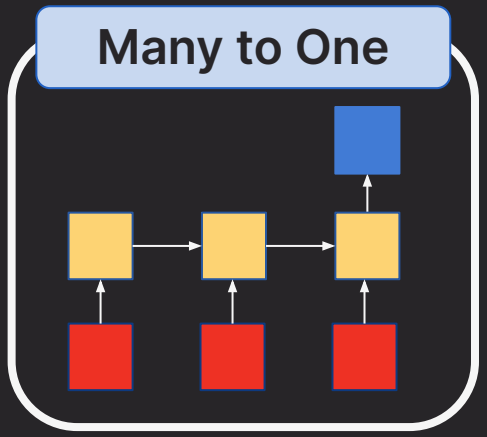
Output at each time step

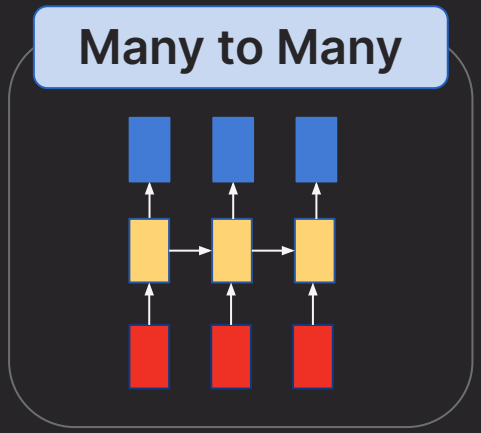
Hidden Layers

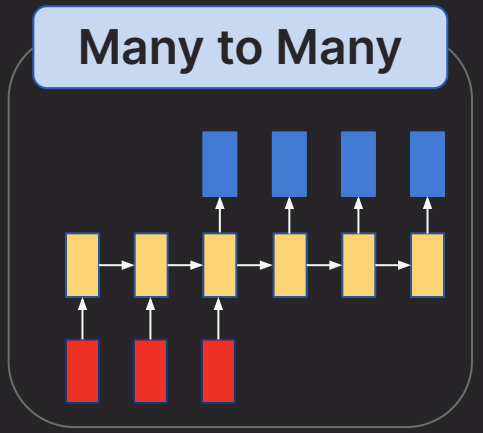
Vectorized Input

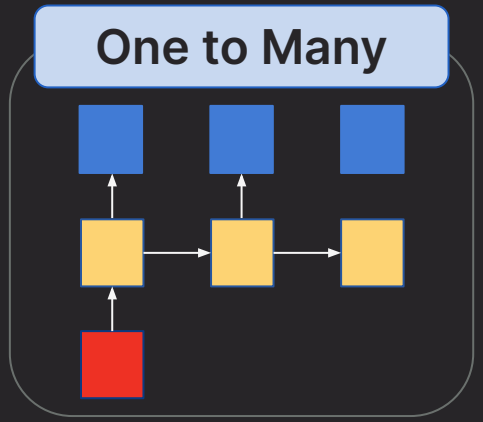
Types of RNN

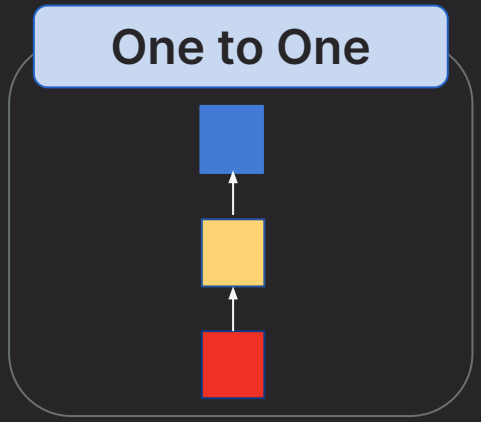
Analytics Vidhya









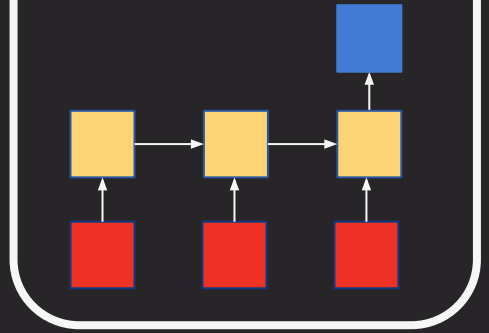


Output at each time step

Hidden Layers

Vectorized Input

Many to One



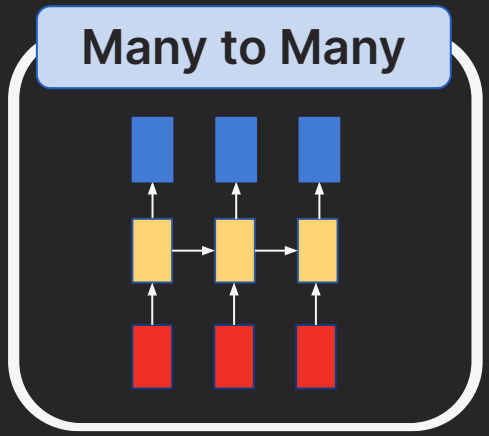
Input & Output : Sequential Data

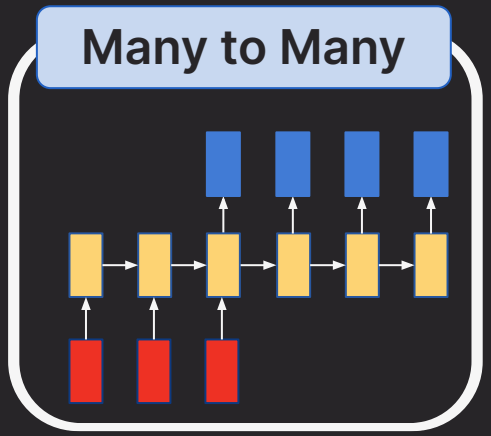
- ~~Machine Translation~~
- ~~Question Answering~~
- ~~Code Generation~~
- ~~Text Summarisation~~

Many to Many RNN

NER

POS Tagging





Language translation

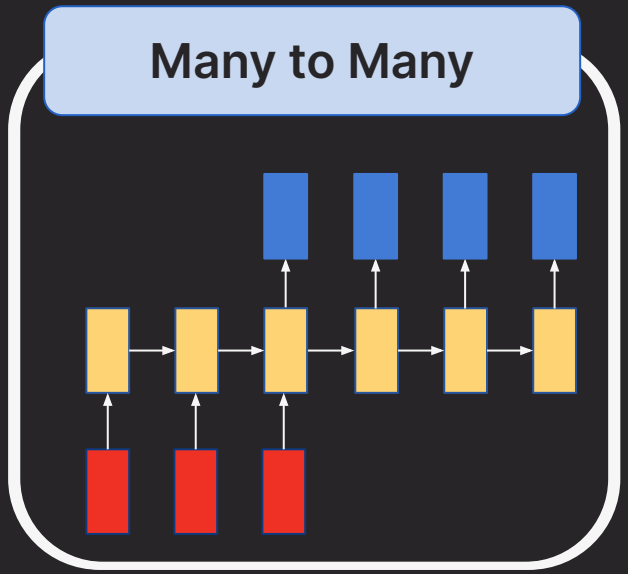
Summary generation

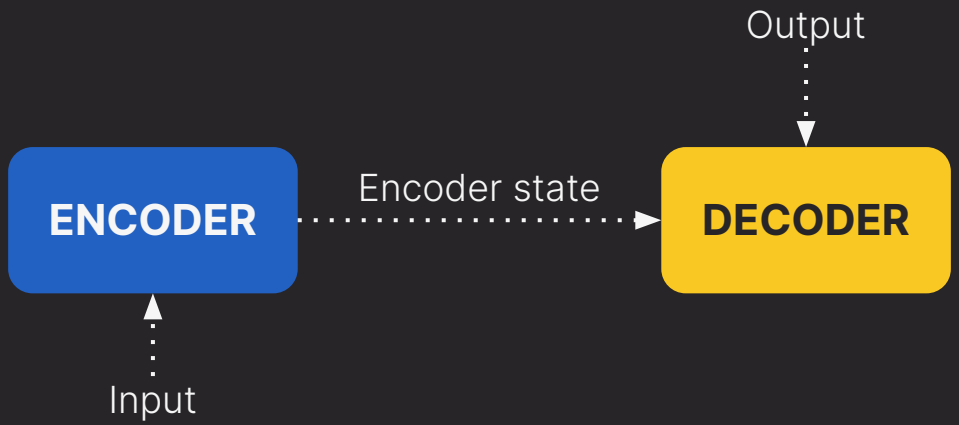
Question & Answering

Text Generation

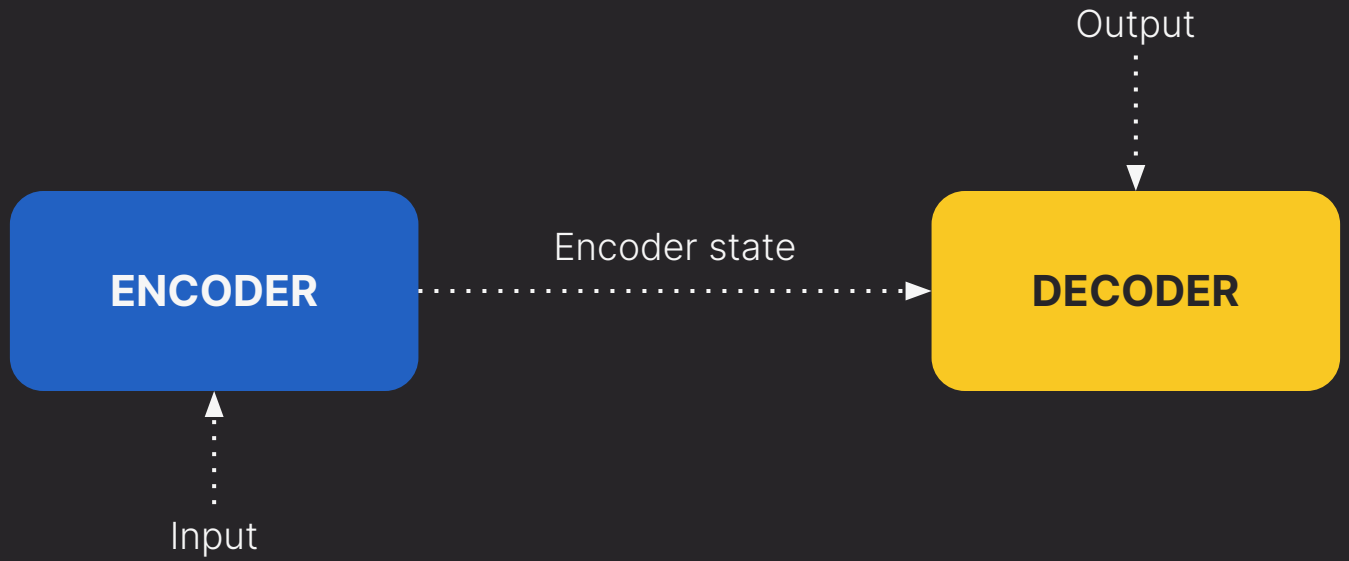
Seq2Seq model

Many to Many RNN



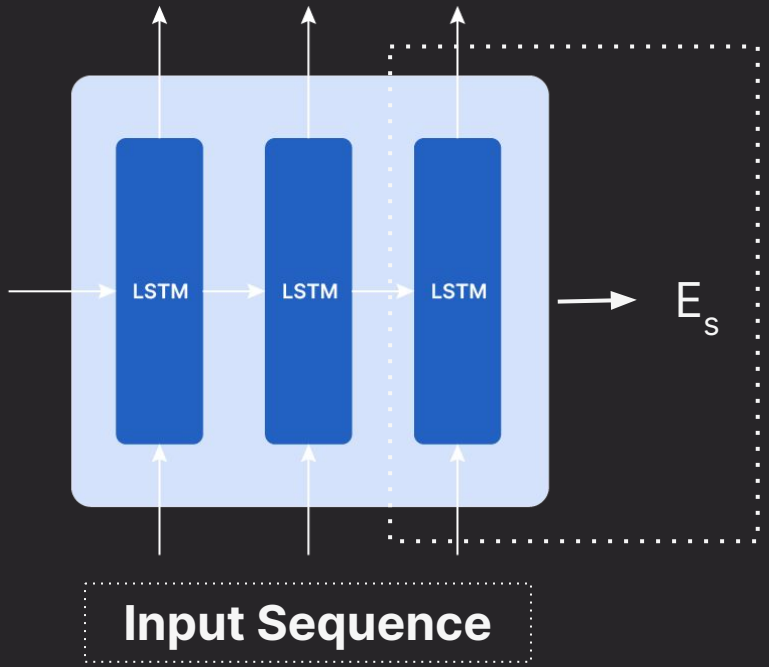


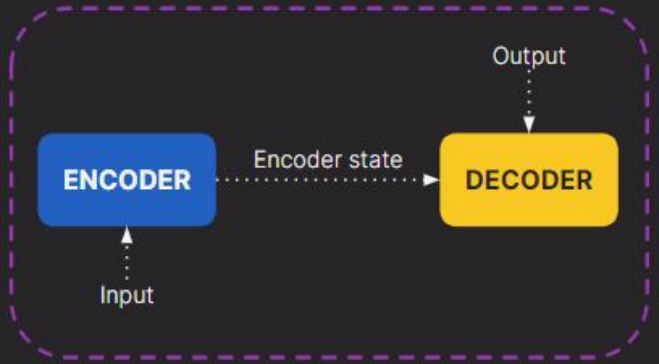
Encoder-Decoder Models



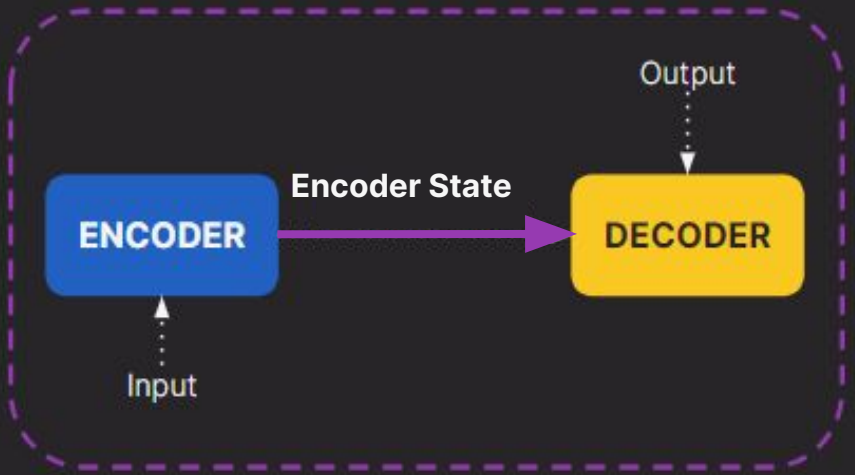
Encoder

Discards the output at every step and preserves only the encoder state.





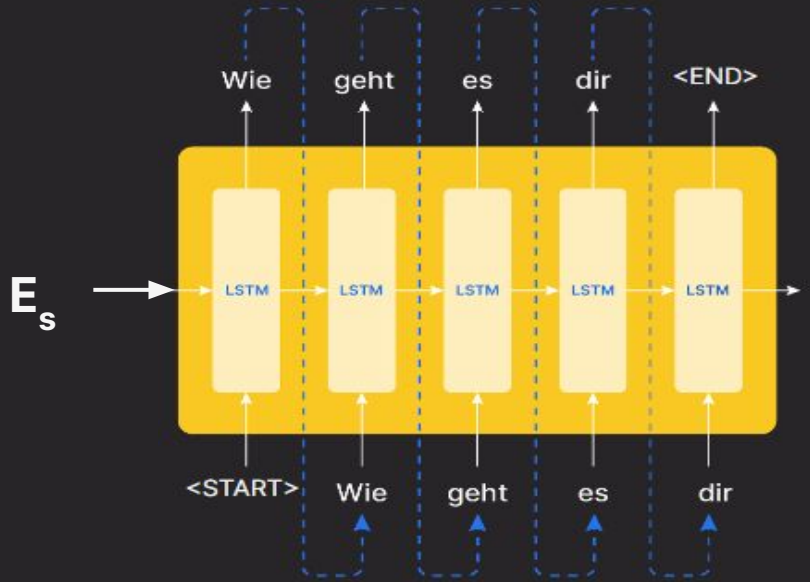
Encoder State

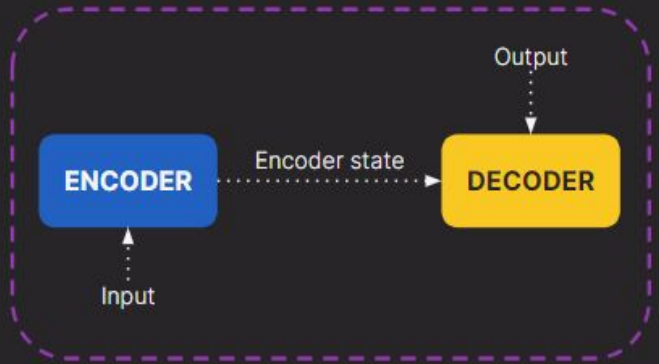


Encapsulates the full meaning of the input sequence

Guides decoder in making accurate output predictions.

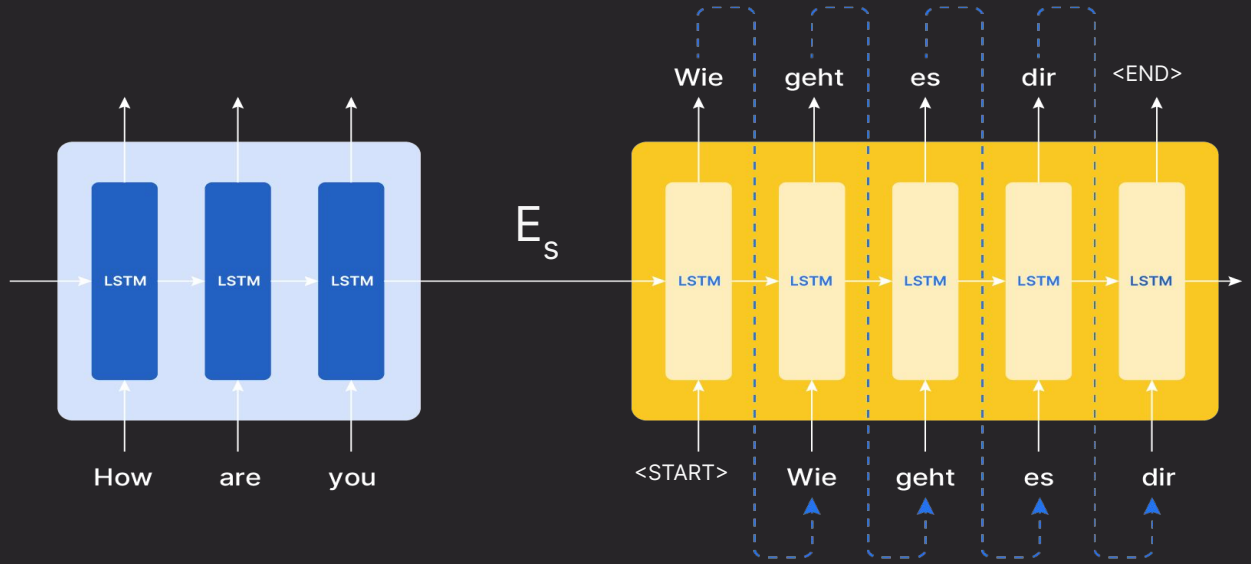
Decoder



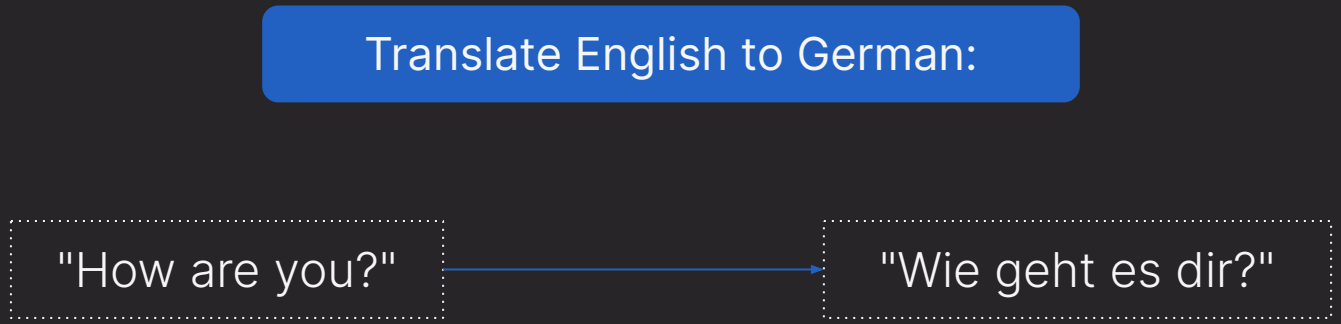


Utilizes encoder state for Sequential Token Prediction

Encoder-Decoder Model

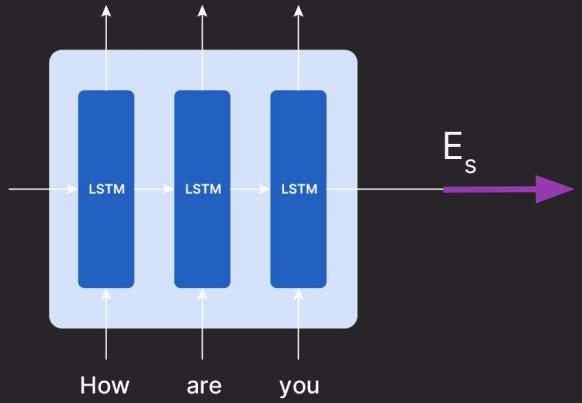


Encoder-Decoder Model



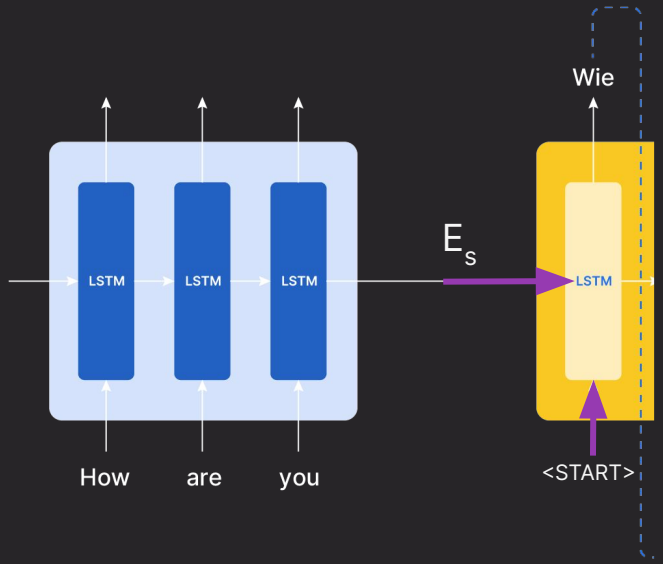
Analytics Vidhya

Working of Encoder-Decoder Model



Working of Encoder-Decoder Model

Timestep 1 : Initialization of Decoding



3 inputs -

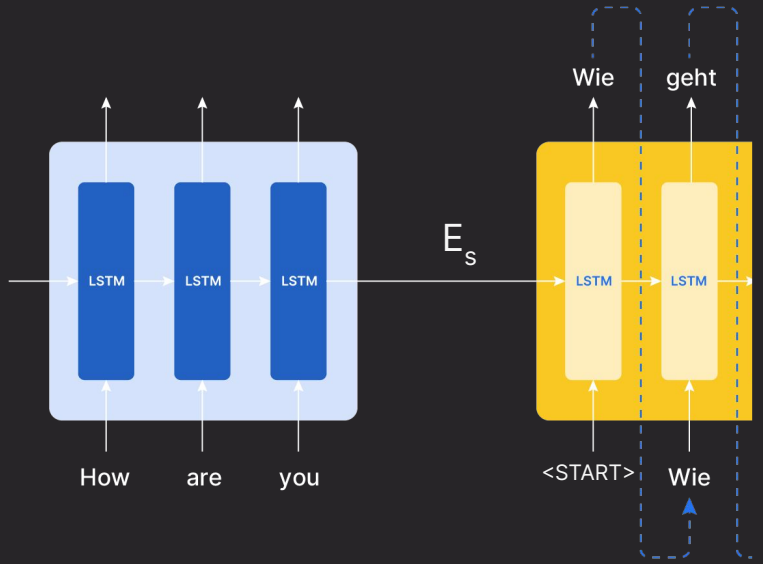
1. Context vector, h_3
2. Cell state, c_3
3. <START> token

Decoder combines context vector $h3$, cell state c3, and <START> token to predict initial output word “wie”.

Analytics Vidhya

Working of Encoder-Decoder Model

Timestep 2, t=2 in Decoding Process



3 inputs -

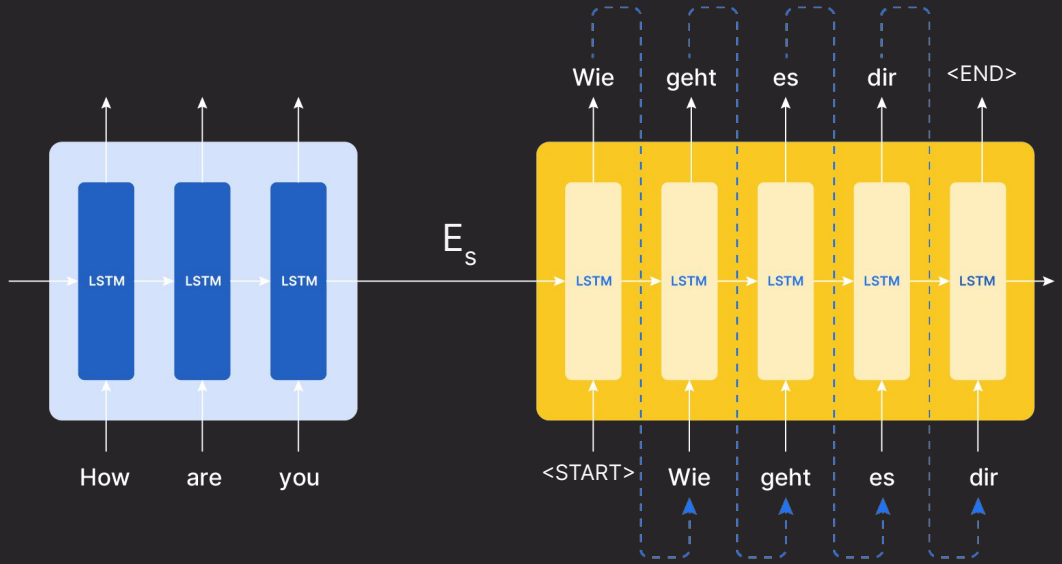
1. Previous hidden state
2. Previous cell state
3. Previous output “wei” as the current input

Uses previous hidden state, cell state, and output “wei” to predict next word “geht”.

Analytics Vidhya

Working of Encoder-Decoder Model

Continuation and Termination of Decoding



Each time step's output is fed into the next until the "<END>" symbol indicates sequence end; final states discarded

- "<START>" and "<END>" can be replaced with any unique strings like “<SOS>” AND “<EOS>”.

- Ensure symbols are not present in the data corpus to prevent model confusion

Differences in training and testing phases