# TRANSFORMERS

By Muhammad Tufail

# Neural Networks

ANN —----------> Tabular data

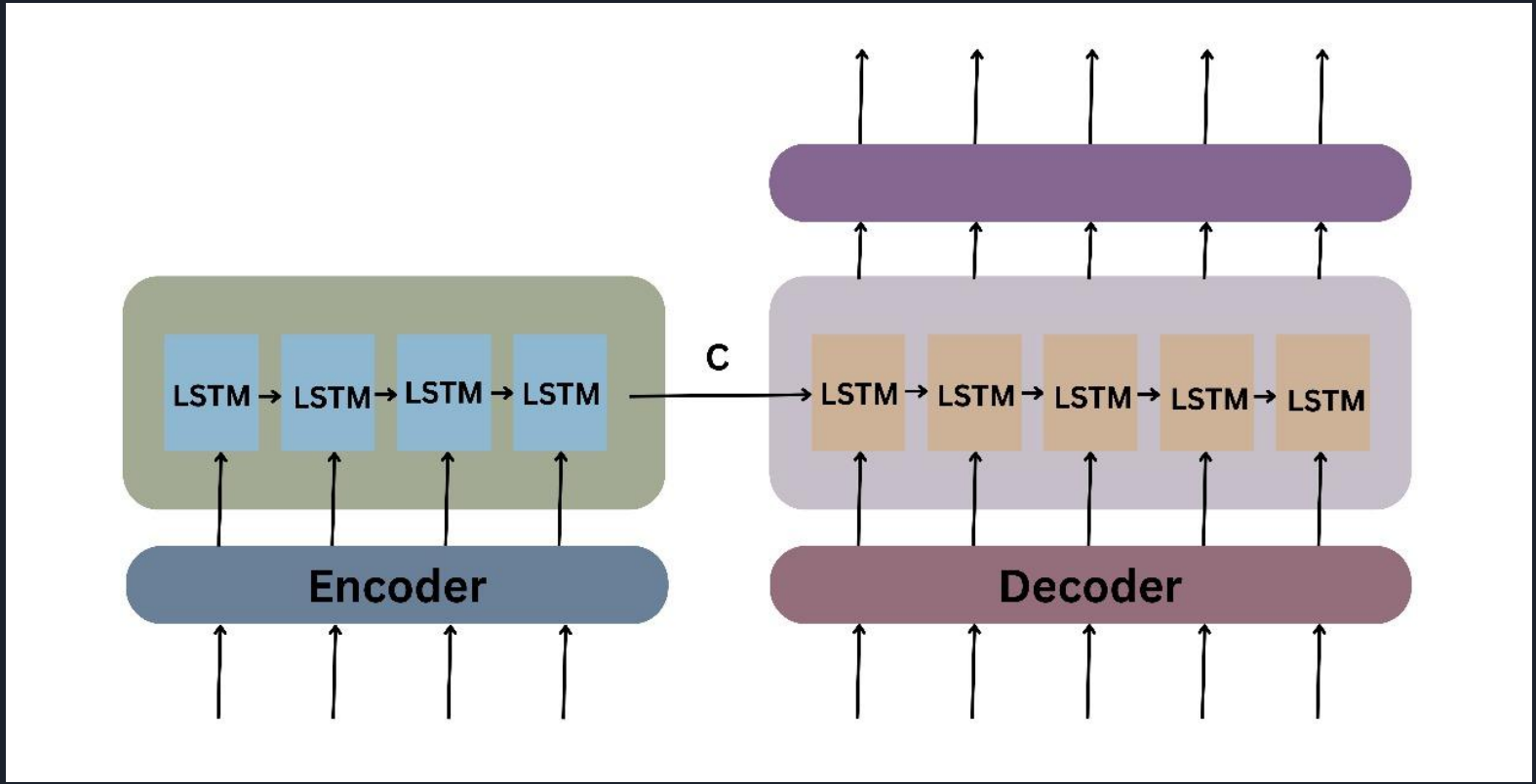
CNN —----------> Images

RNN —----------> Sequential Data (Text)

LSTM , GRU

(Machine Translation, Text Summarization, Q/A Systems)

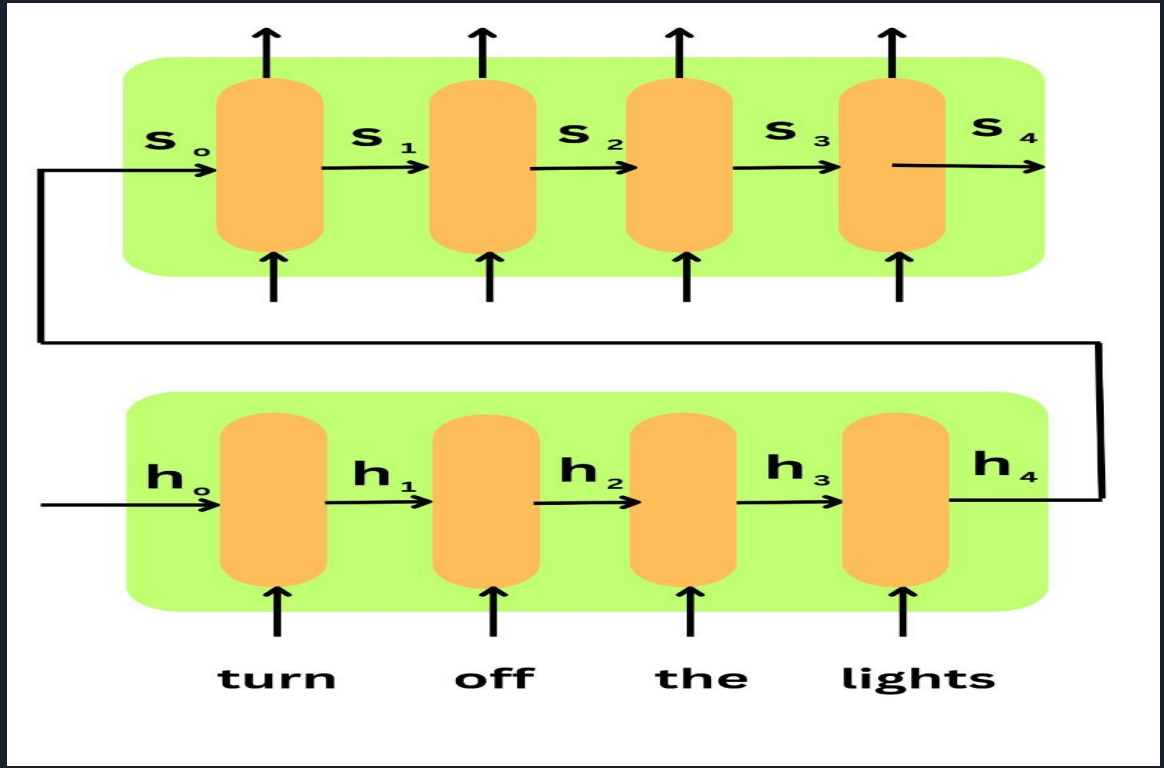
# Encoder- Decoder Mechanism:



# Challenges of Before Transformer Architecture:

- Training on Large Dataset was not possible
- Fine Tuning + Transfer Learning
- Sequential processing (Vanishing Gradient)

# Attention Mechanism:



- Context Vector is dynamically calculated
- Decoder print word influence by which encoder word
- Context vector is weighted sum of states of encoder (these weights are called attention weights)
- With attention mechanism quality of translation was much improved (specially in sentences containing words greater than 30)
- Still there was big challenge of “How to train model on Large Datasets”

# Importance of Transformers:

- Transformers were basically used to solve NLP problems.
- Transformers has speedily boost NLP (Revolutionized Deep Learning)
- Scalable training on large dataset (GPT, BERT available for use)
- You can use on your small dataset and fine Tuning
- Hugging face available for fine tuning
- Multimodal capability (Very Flexible, you can use it with image /video, input text output image/video)

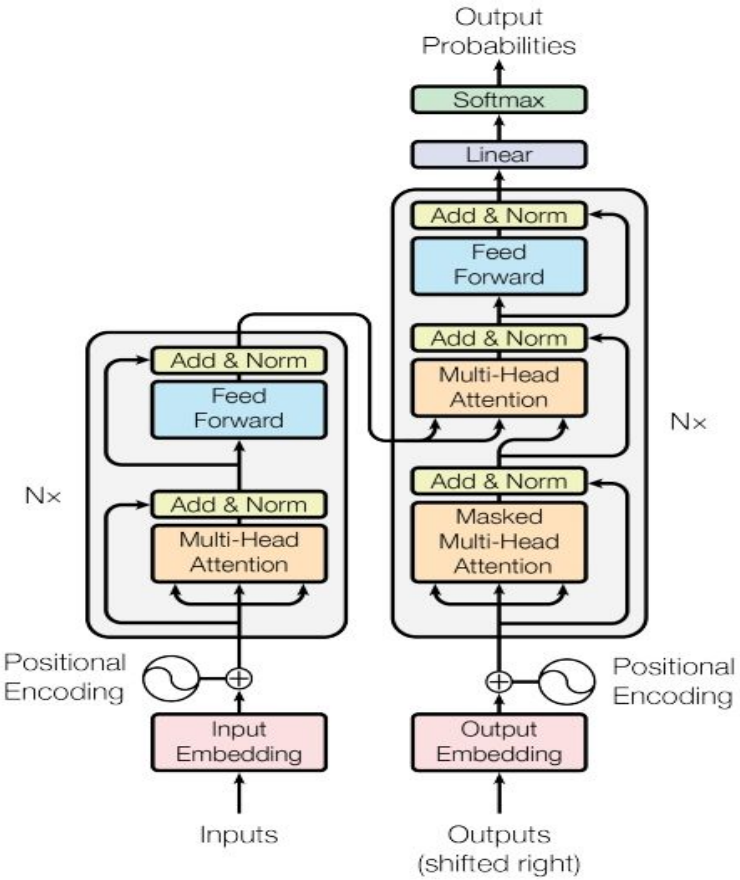



Figure 1: The Transformer - model architecture.

# Transformer Architecture

- Uniqueness of Transformer:
- Residual Connection
- Attention Mechanism (Replaced RNN, LSTM)
- Layer Normalization
- Parallel Processing
- With lot of components, it is highly stable & flexible