Analytics Vidhya

Model Architecture

Instructor

Sourab Mangrulkar

Machine Learning Engineer at

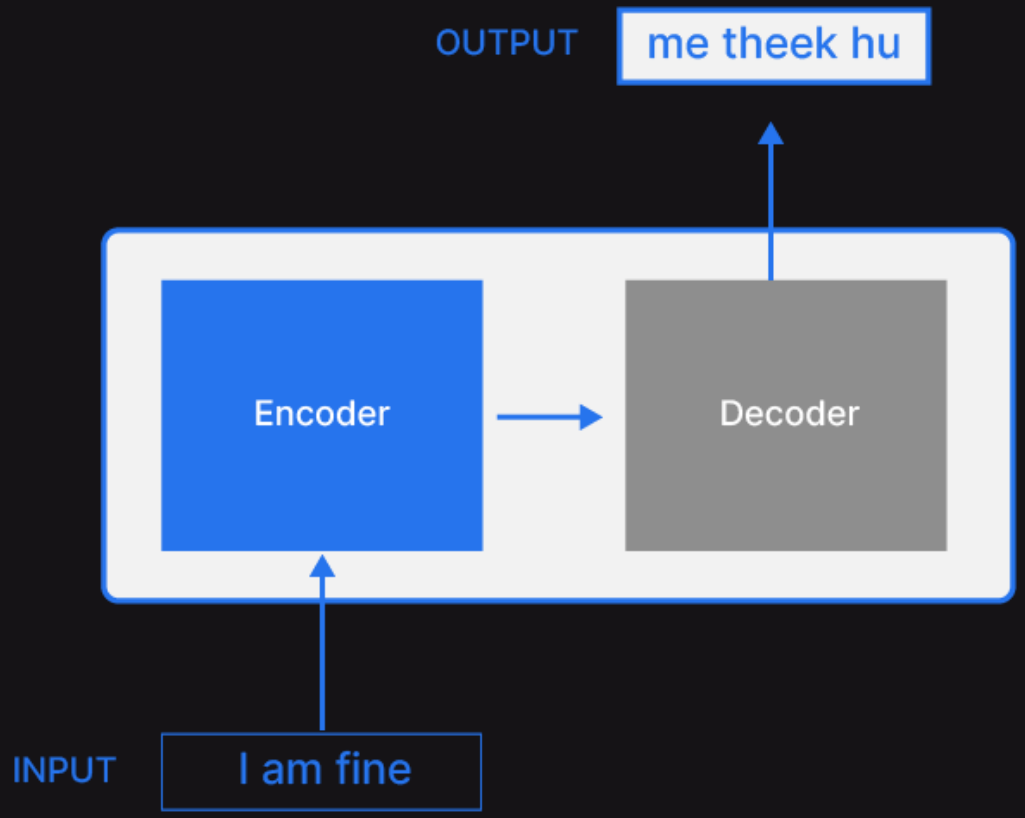
Creator of PEFT

Recap

- Training data curation
- Data Preprocessing
- Tokenization
- Model Architecture

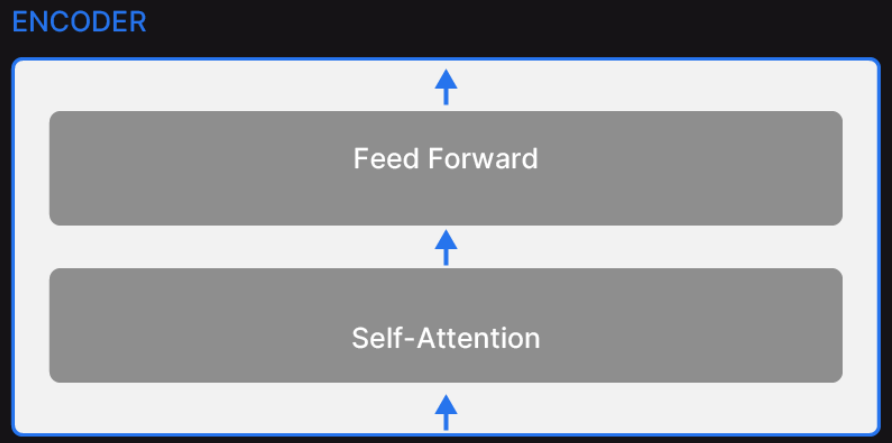
Transformers are the backbone of LLMs

Model Architecture



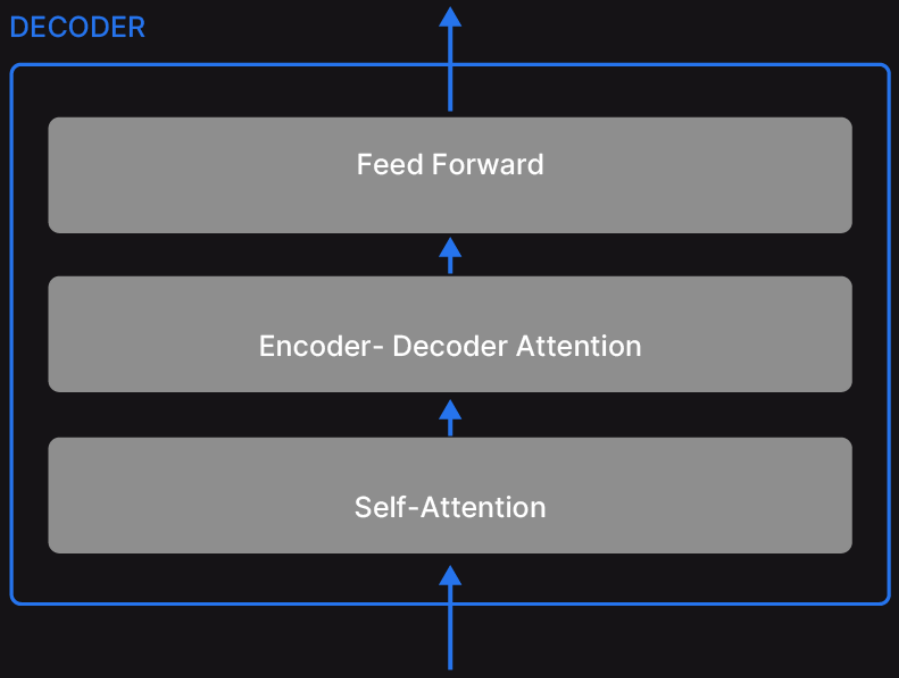
High Level Design of Transformers

Model Architecture



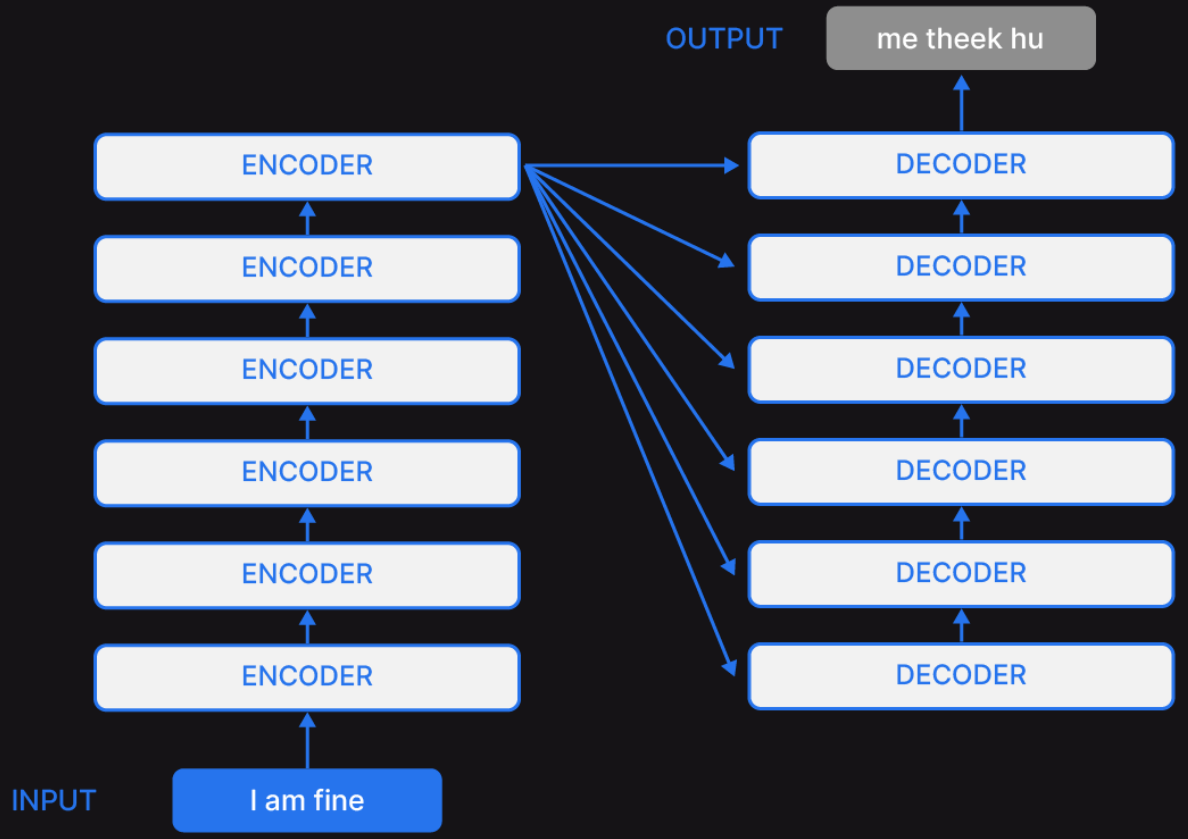
Encoder

Model Architecture



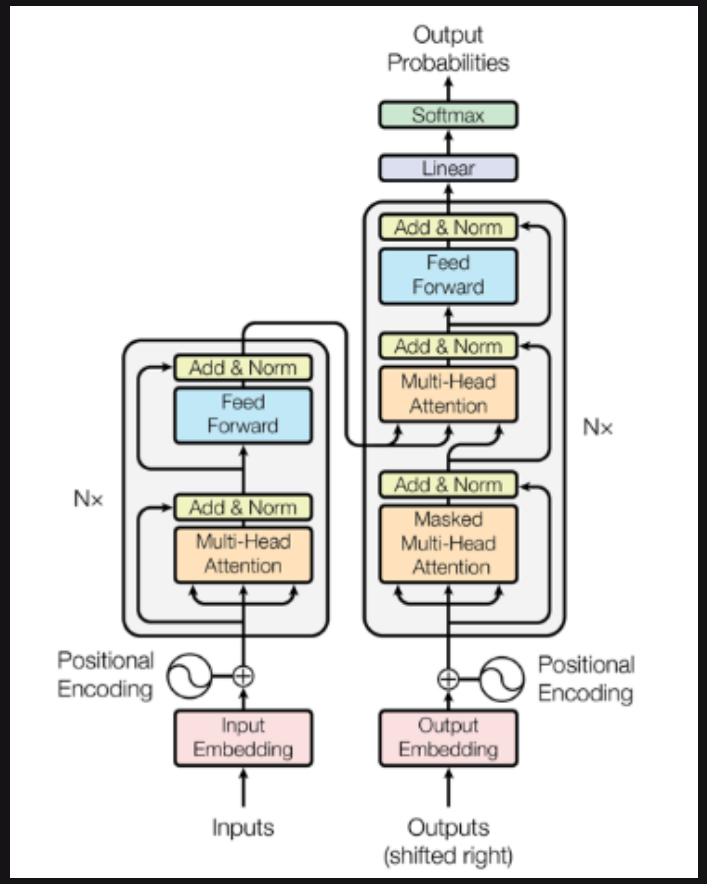
Decoder

Model Architecture

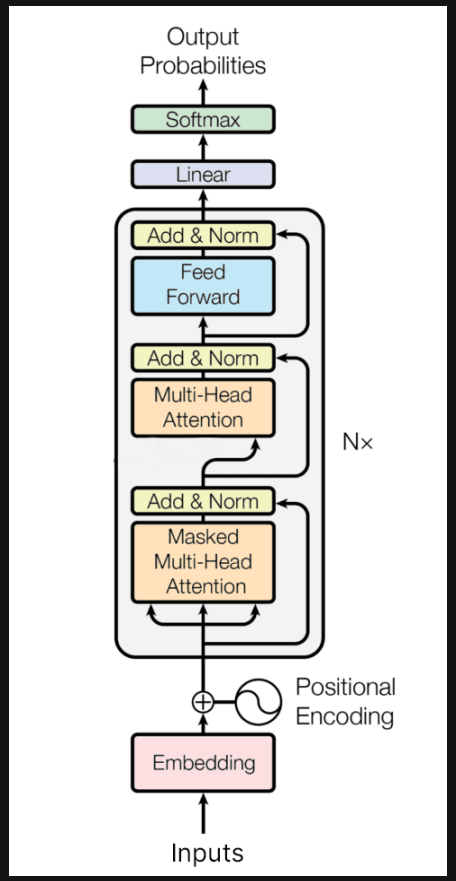


Transformers

Model Architecture

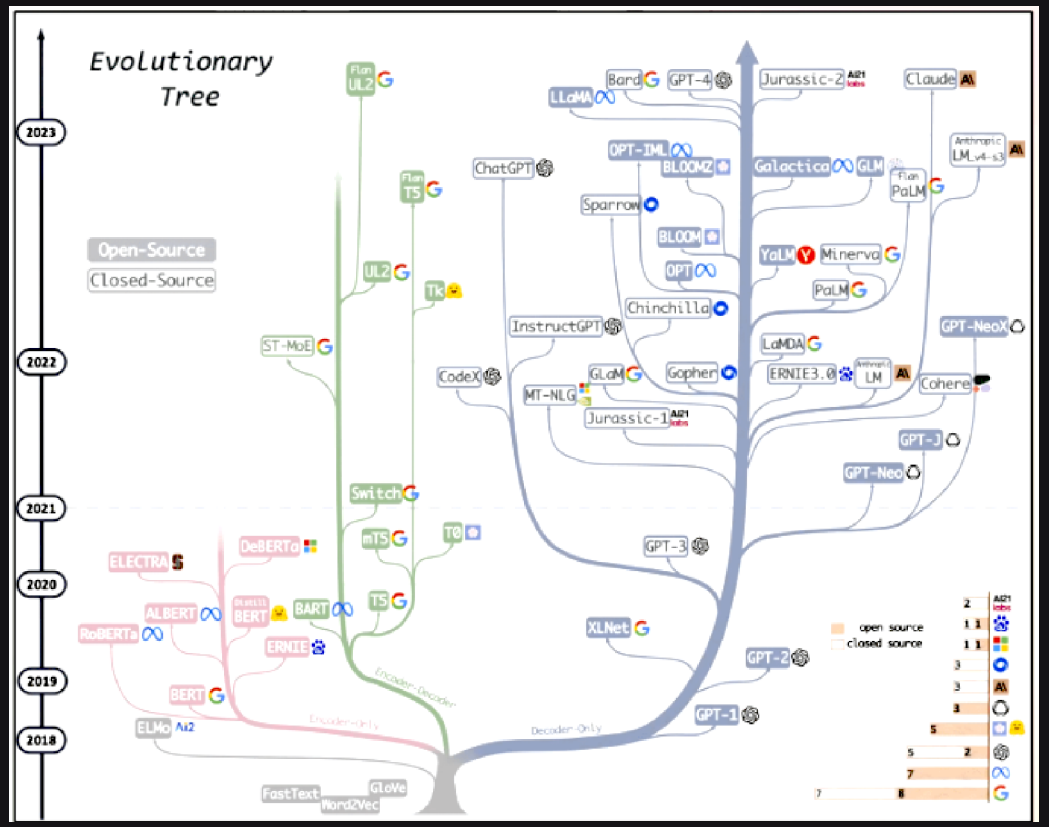


Decoder Architecture



Decoder only LLMs

- February 2023:
  - **LLaMa**
- March:
  - **Alpaca, Vicuna**
- April:
  - **Koala**
- May:
  - **StarCoder, StarChat, MPT-7B, Guanaco**
- June:
  - **Falcon, MPT-30B, Phi-1**
- July:
  - **LLaMa-2**
- September:
  - **Falcon 180B, Mistral-7b**
- November:
  - **Yi-34B, Zephyr-7b**
- December:
  - **Mixtral-8x7b, Phi-2**

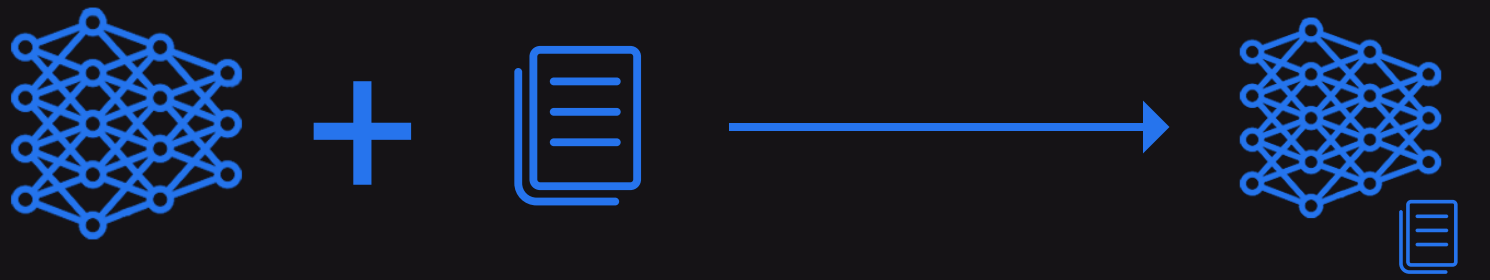


Analytics Vidhya

How to define model architecture?

1.Existing Model Architecture

Adapt the existing model architecture



1.Existing Model Architecture

Llama 2	Code Llama
GPT-4	Github copilot

Pros

- Very little effort
- High performance

2.Modify Existing Model Architecture

- Changing few settings of existing model architecture

2.Modify Existing Model Architecture

- Changing few settings of existing model architecture
  - Few Layers
  - Changing hyperparameters

2.Modifying Existing Model Architecture

- Sliding window architecture
- Rolling buffer cache
- Prefill and chunking

2.Modifying Existing Model Architecture

- It uses Flash Attention
- It uses ALiBi (Attention with Linear Biases) and does not use positional embedding
- It does not use biases

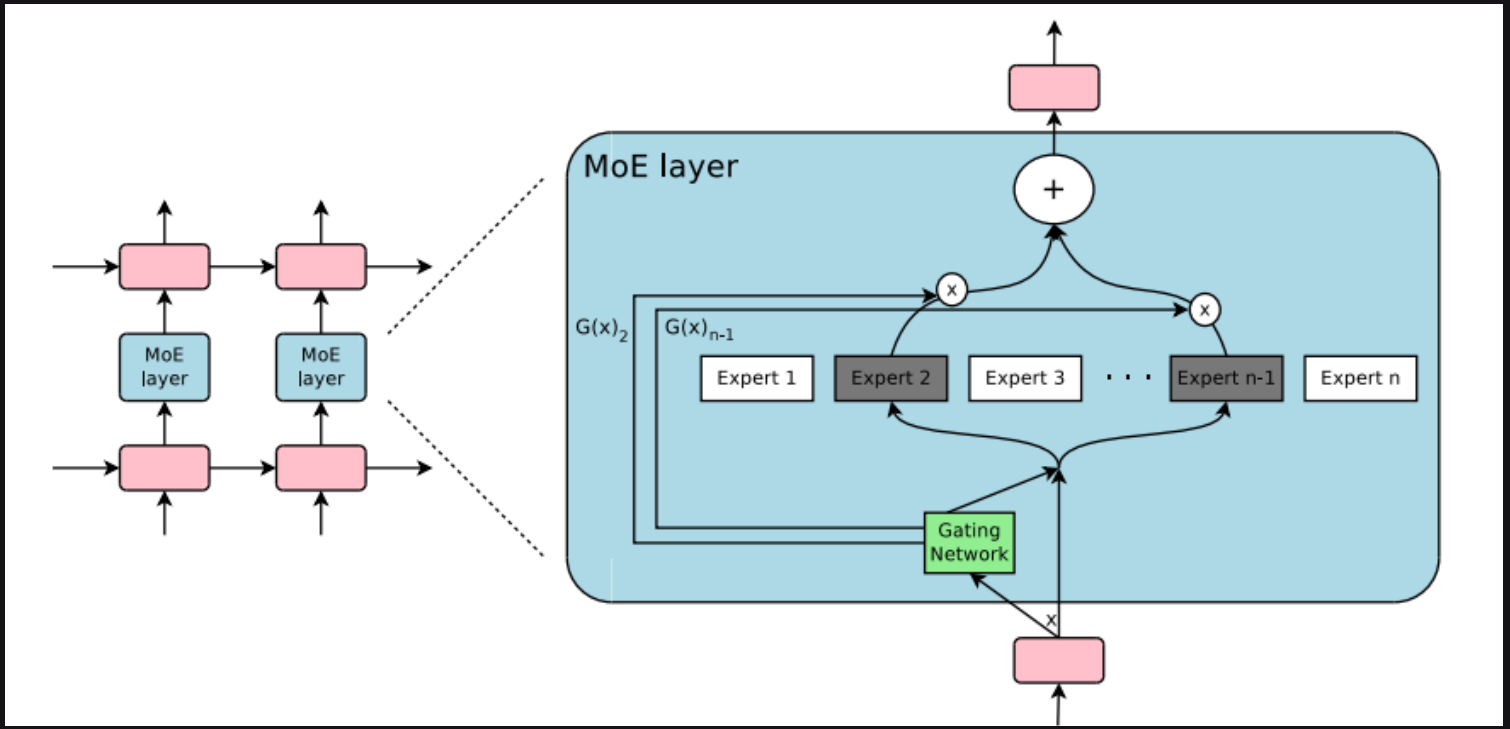
Pros

- State of the art model

3.Design New Model Architecture

- Done by well funded organisations
- Run multiple experiments

3.Design New Model Architecture



Which is the ideal one?

Adapt/Modify existing model architecture

Model Training

- Define the hyperparameters

Model Training

- Define the hyperparameters
- Set up the parallel and distributed strategy
- Train the model

Model Training

- Define the hyperparameters
- Set up the parallel and distributed strategy
- Train the model
- Save the training environment

Thank You

Analytics Vidhya