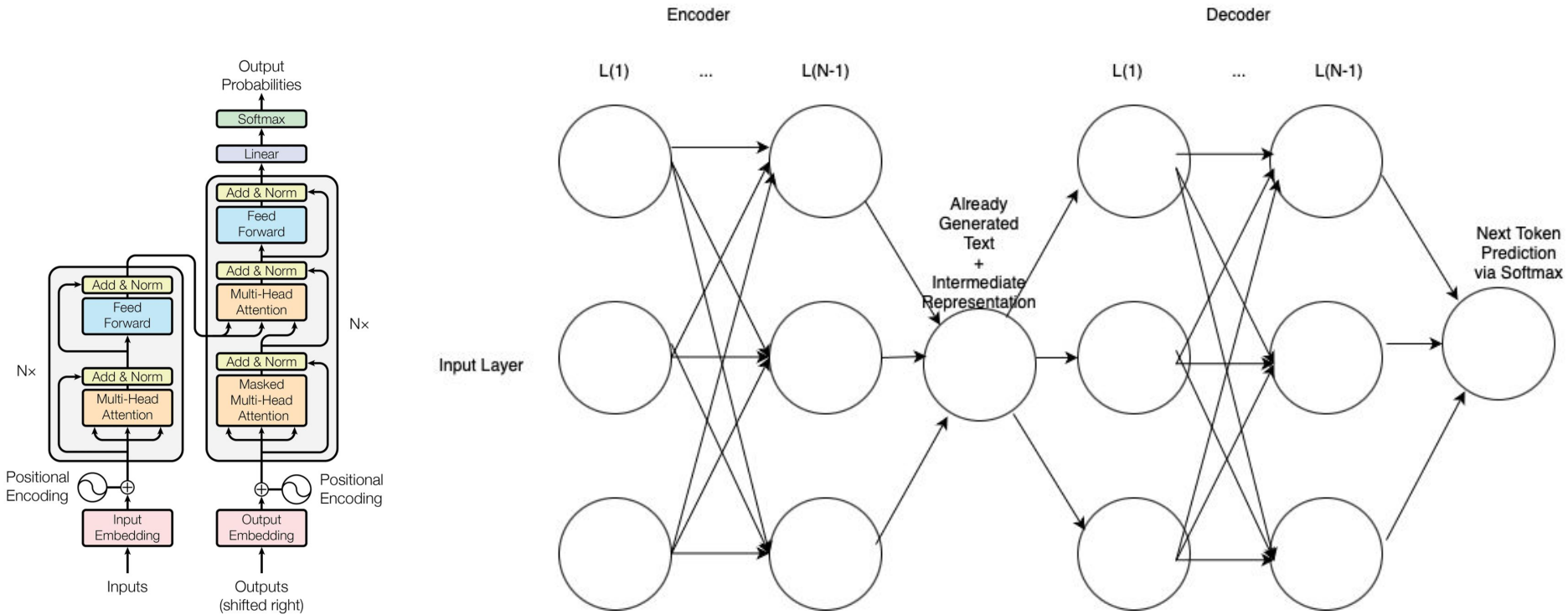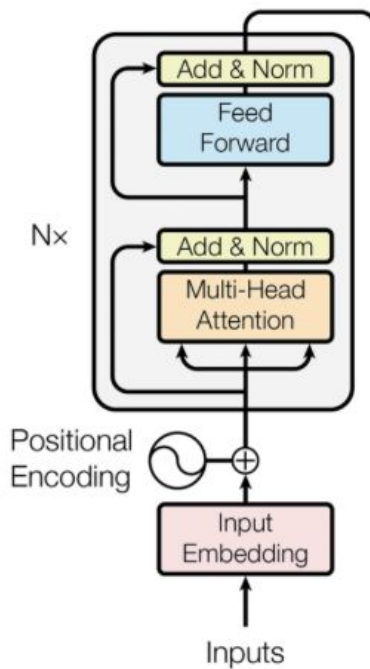# Tutorial 11/11/2024 Transformers in Action ML4NLP Tutorial

A-non accurate Encoder - Decoder Representation
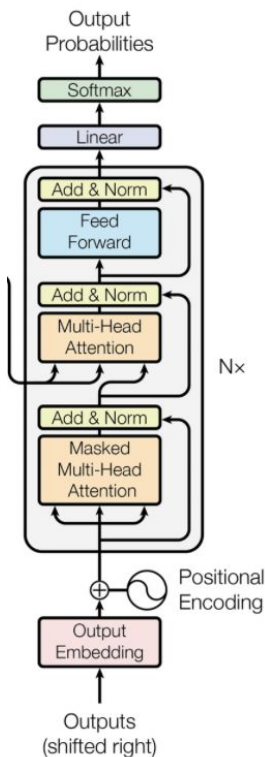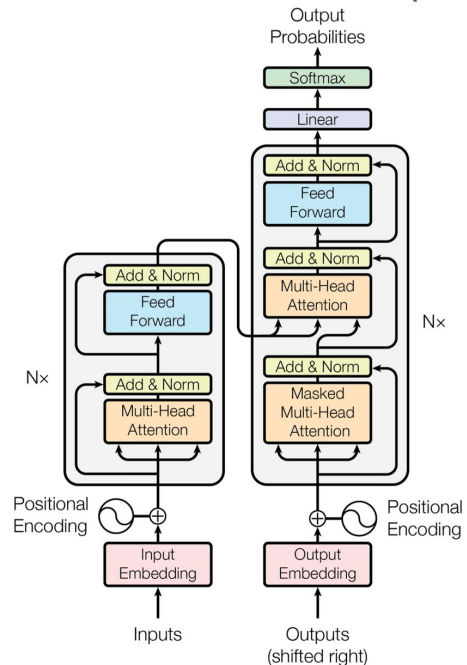

(Vaswani et al., 2017)

Encoder
(ex. BERT)

Decoder
(ex. GPT)

Encoder-Decoder
(ex. Text to Text
Transformer (T5)

7. Transformers in Action (6 Points)

3 (a) Drawing on the architecture

(i) On the original transformer diagram, draw a labeled circle (use the labels 'BERT' and 'GPT') around the part relevant to BERT and the part relevant to GPT. (1 Point)

Left is BERT, Right is GPT. Extra note: Correct remotely correct circles (even if they exclude the bottom format)

(ii) On the original transformer diagram, mark one position with a **labelled arrow** (A→) where an **adapter layer** can be added. (1 Point)

Many positions for adapter possible, accepting all of them.

(iii) Label the boxes marked with a question mark (**?**) with their intended name! (1 Point)

Half point for each type, the Feedforward layer accept anything remotely close to Feedforward (MLP, FFNN or anything)
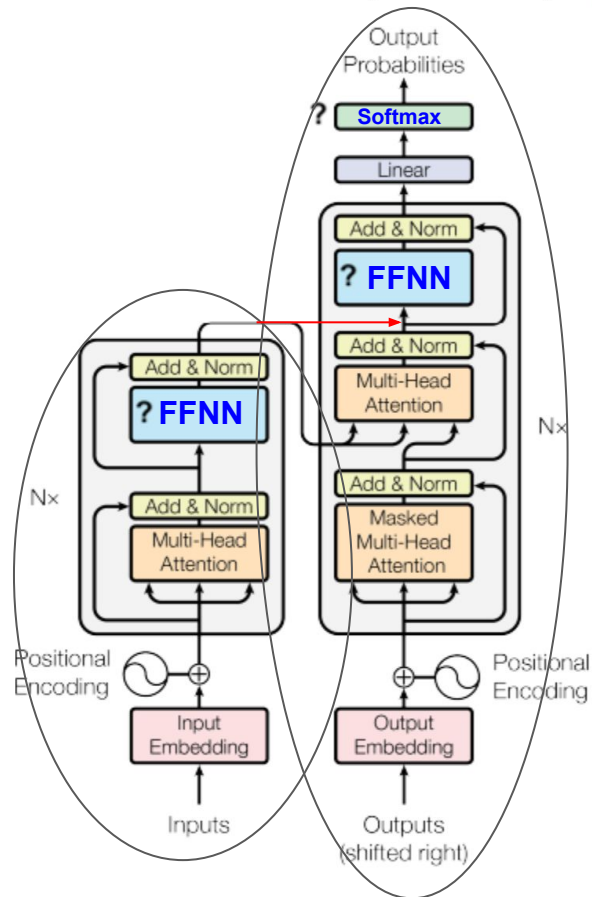
Another Question from last year's exam:

Using a GPT-2 Language Model for Political Speech Analysis
Given 10,000 short parliamentary speeches labeled as 'Left' or 'Right',
suggest a detailed strategy how to use a pre-trained GPT-2 model for
detecting whether an unseen speech is 'Left' or 'Right'.

1. Which architecture is this?   It's a Decoder   Note: That was before the emergent abilities of LLM's, hence zero/in-context learning and prompting wouldn't work

2. Detailed strategy denotes a need for an operationalizable plan.

Solutions:
1. (Full Points) Take intermediate representation of the input text through GPT-2 and use it externally as features to any classification method
2. (Full Points) "Text" :/specialToken (ST) "Label" (Label can also be text or another Special Token for more consistency) and finetune the original model autoregressively. During inference "Text" : (ST) 'Prediction' and use the autoregressive prediction to select the higher of the two labels.
3. (Full Points) One model finetuned on "Left" text, One model finetuned on "Right" Text and perform prediction on the new text by classifying through choosing the orientation of the model that has the lowest loss of the tw

Live Demo Task Solving