

StatsQuest - Encoder-only Transformer

example of encoder only = BERT

Chat GPT = Decoder only

Decoder only is much more popular

Focus on the word embeddings part

instead of predicting the next word from
1 word / token

↳ we can predict the next word from the
preceding 4

issue is that in its most basic form, Supplying
4 word does not retain any info on the
order of ~~the~~ the 4 words

↳ need positional encoding

encoder-only needs: Embed → Positional → Self-ATT

together these ~~are~~ final values are known as:

- Context aware embedding
- Contextualised Embedding

these embeds ~~are~~ allow us to organise
& cluster similar words or documents
or sentences

= BERT

↳ Docs & sentences = basis
for RAG

RAG = Break Document down into blocks of text & then use encode-only transformer create context embeds

Prompt = What is pizza?

↳ embed & find similar

Common thing to do ~~into~~ is to use context embeds as inputs to a neural network