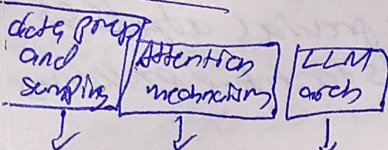


"Emergent Behavior" → ability of a model to perform tasks that the model wasn't explicitly trained to perform
 How to explain? STILL isn't known
 ○ Good research topic

Stages of Building LLM

Stage 1



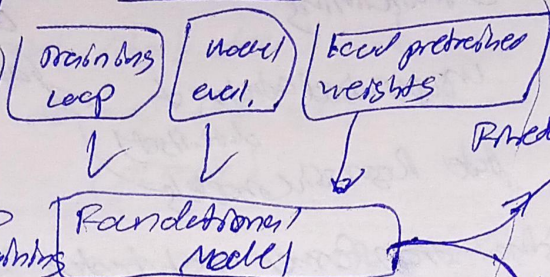
Building on LLM

Improve data sampling;
 understand basic mechanism

Tokenization: break text into tokens
vector embedding: every word/token translated to high dimensional space; capture semantic meaning
context vectors of data: context; How to feed

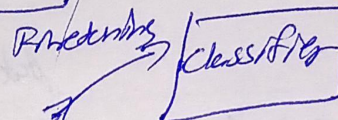
- 1 data on GPUs so computation becomes
- 2 more efficient

Stage 2



pretrain on LLM on unlabeled data

Stage 3



personal assistant

fine-tune pretrained LLM

Recap

① LLMs have transformed the field of NLP: best achievement in generating, understanding and translating human language

② Modern LLMs are trained in 2 main steps

~~costly~~ Pretrained on unlabeled data

fine-tuning on a smaller, labeled dataset

~~from a dp~~

can adapt to perform on specific tasks

- ③ LLMs are based on transformer architecture
 - ↳ Attention mechanism solves LM selective access to whole input sequence when generating output one word at a time
- ④ original transformer: Encoder and Decoder
- ⑤ GPT: only Decoder, no Encoder
- ⑥ While LLMs are only trained for predicting next word they show **EMERGENT PROPERTIES** (ability to classify, translate and summarize texts)