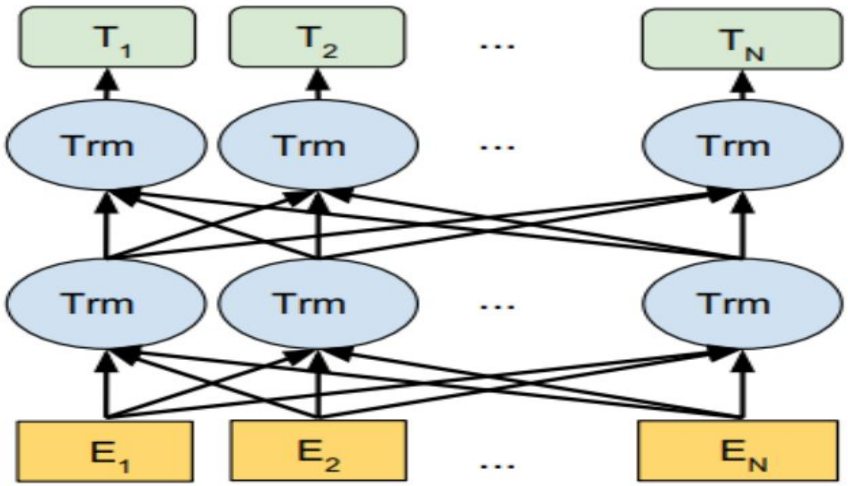
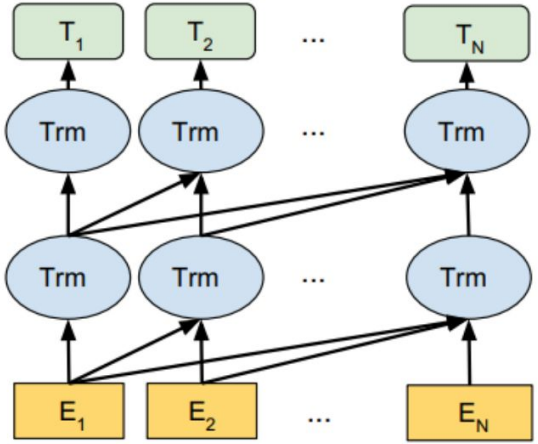
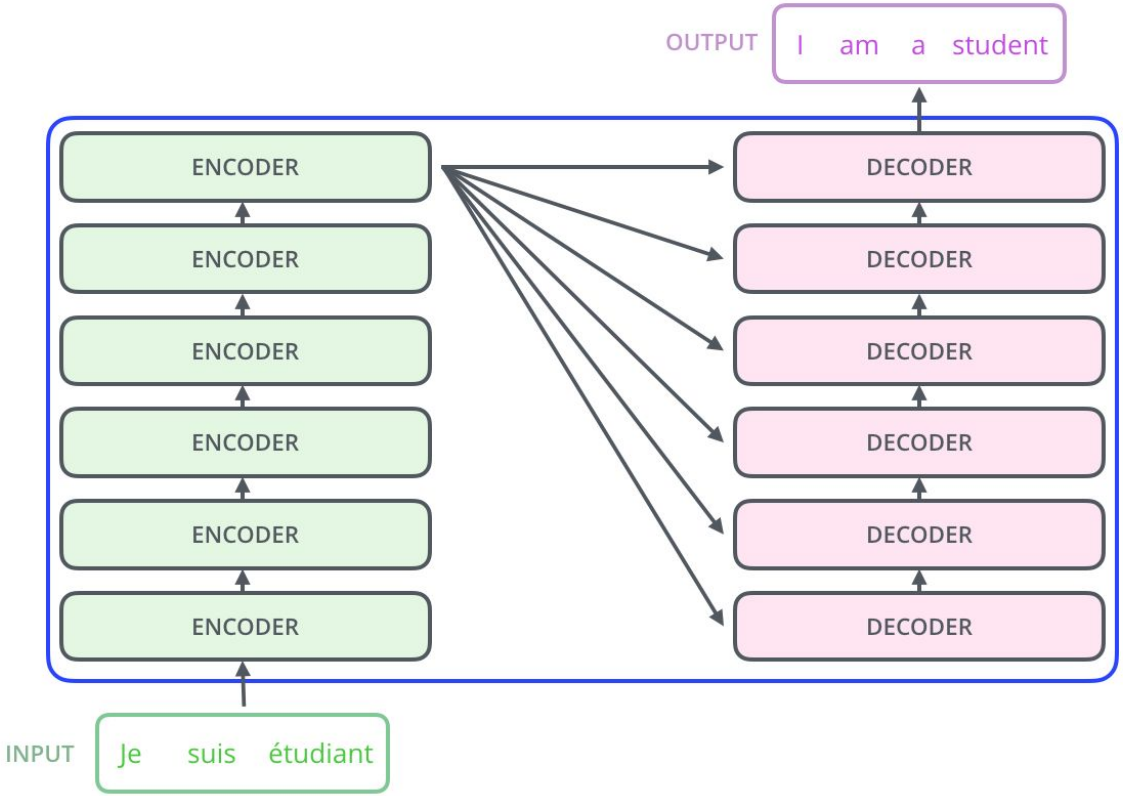
Transformers - usages

Tafseer Ahmed

Encoder vs Decoder

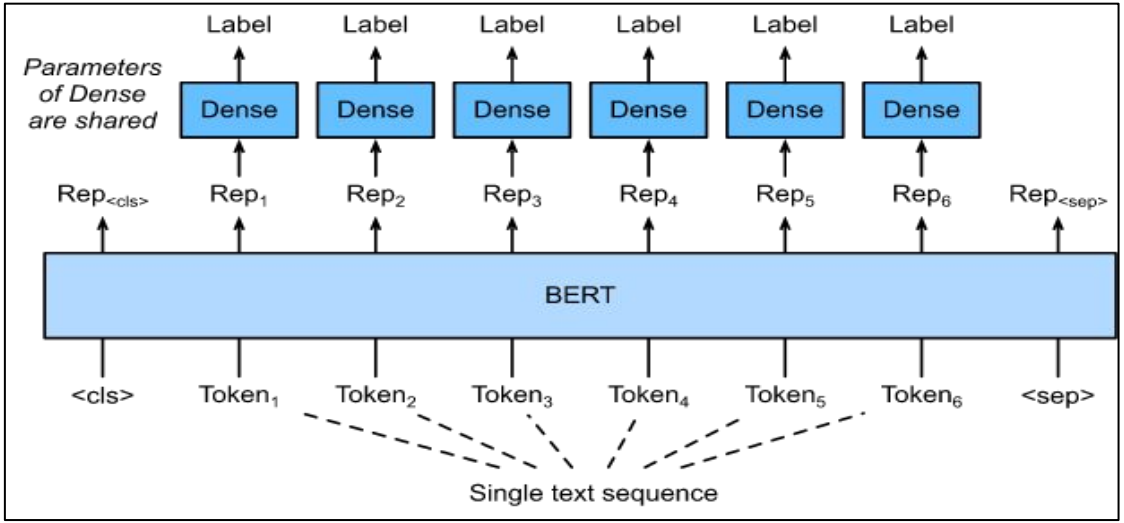
Bidirectional **E**ncoder **R**epresentation for ***Transformers***	**G**enerative **P**re-trained ***Transformer***
	

Encoder-Decoder



Encoder - Word based Task

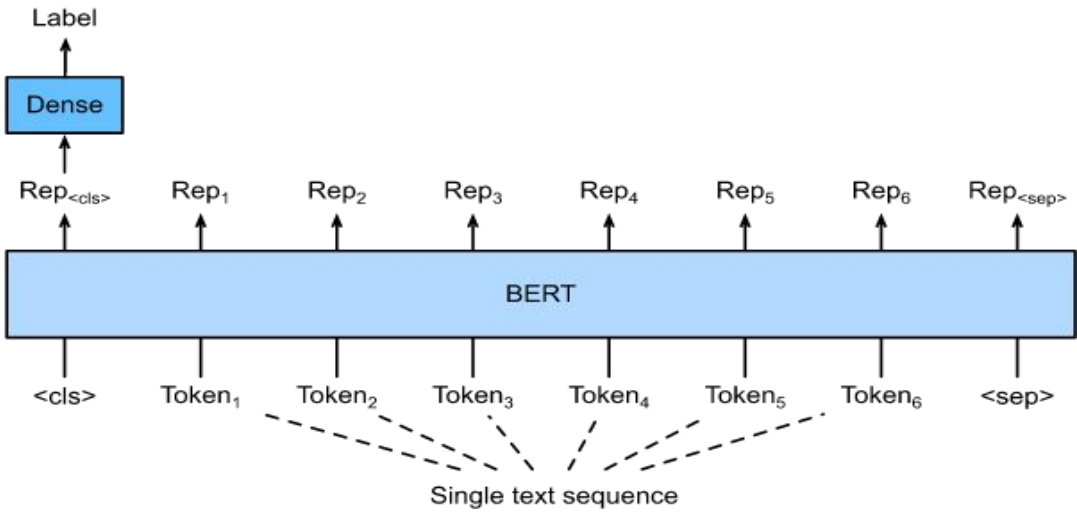
- Word Embedding Extraction
- Part-of-Speech Tagging
- Named Entity Recognition (NER)
- Contextual Word Disambiguation
- Semantic Relation Classification
- Synonym Generation
- Word Sense Induction
- Masked Word Prediction
- Contextual Similarity Matching
- Coreference Resolution



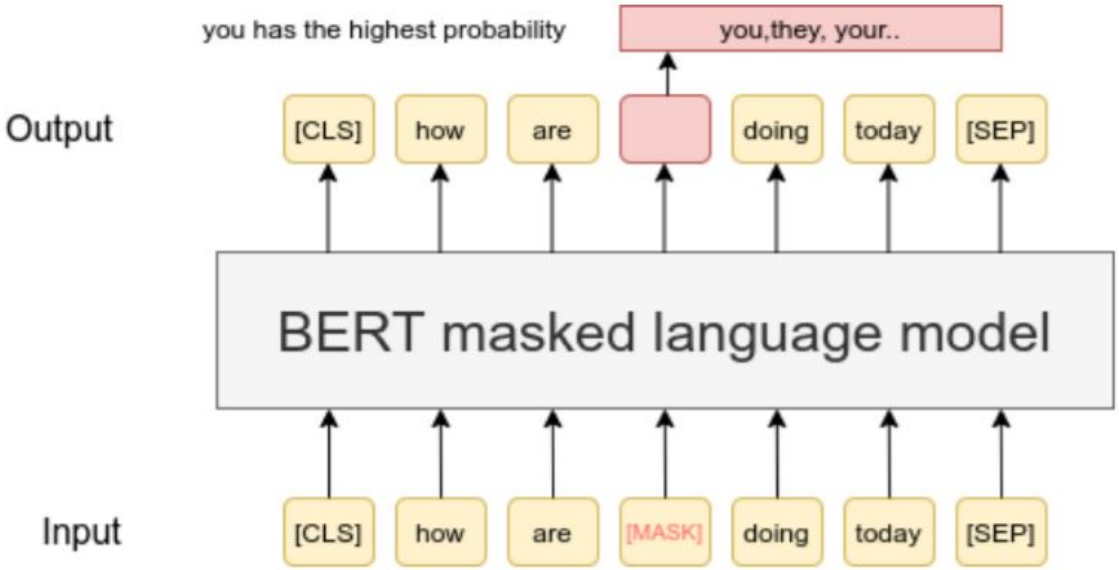
- *Summarization*
- *Question Answering*

Encoder - text based tasks

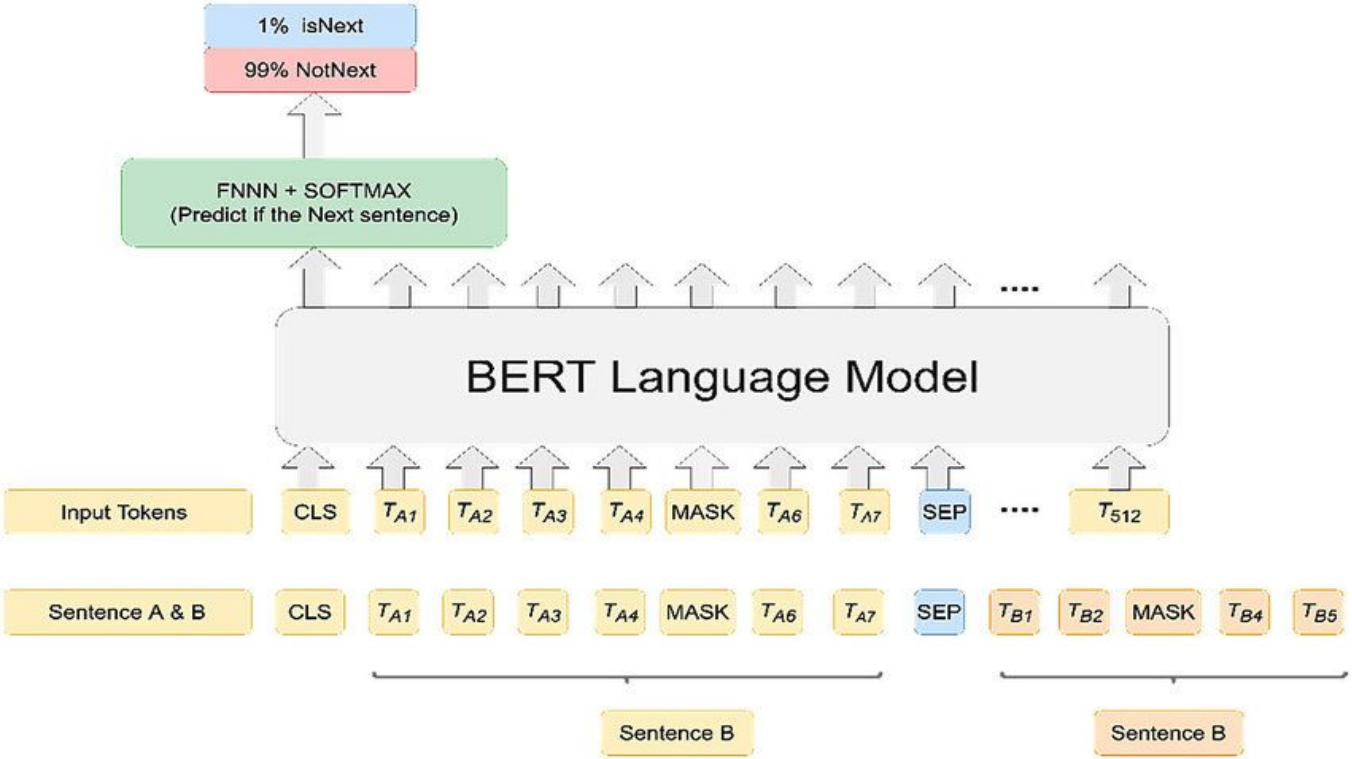
- Text Classification
- Sentence Pair Classification
- Question Answering
- Paraphrase Detection
- Sentiment Analysis
- Next Sentence Prediction
- Sentence Similarity
- Natural Language Inference (NLI)



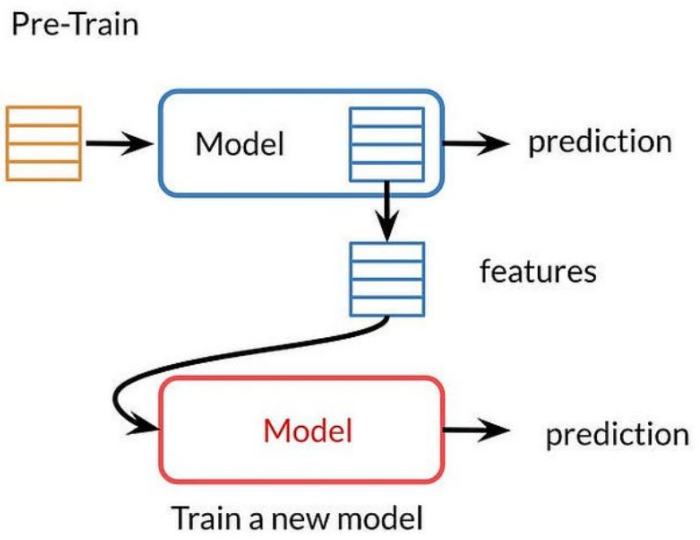
Encoder - pre training

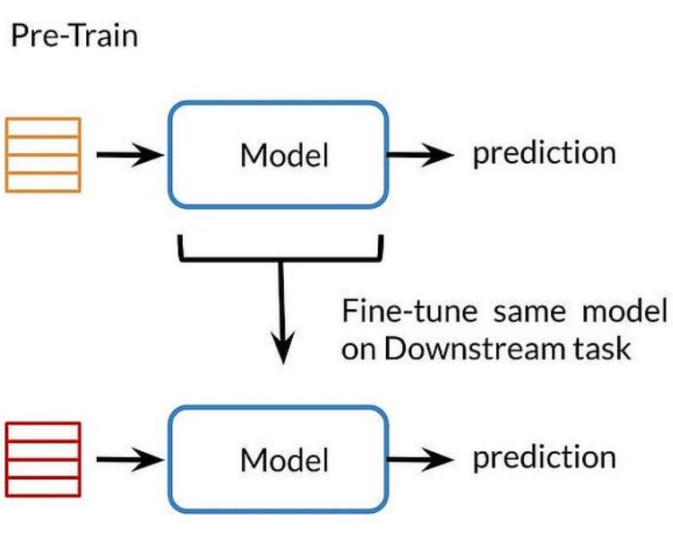


Encoder - pre-training

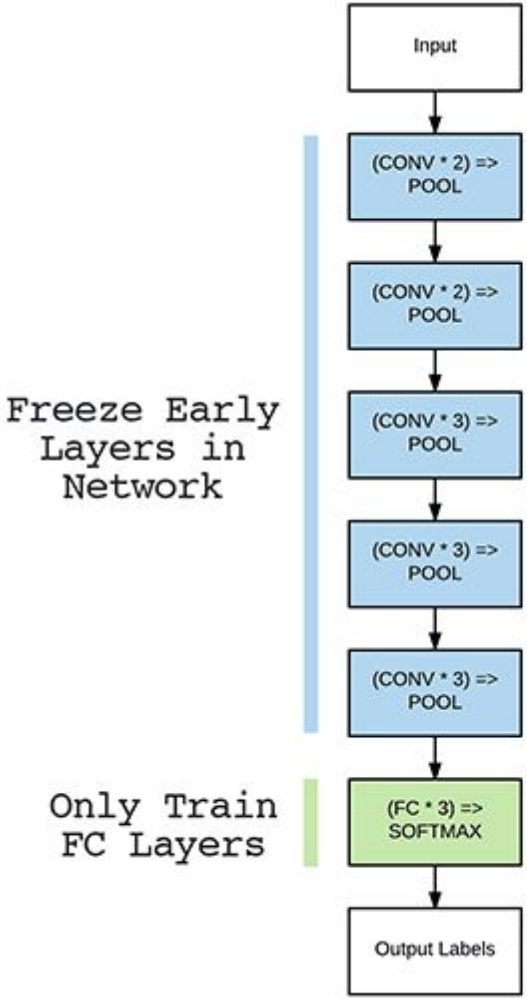


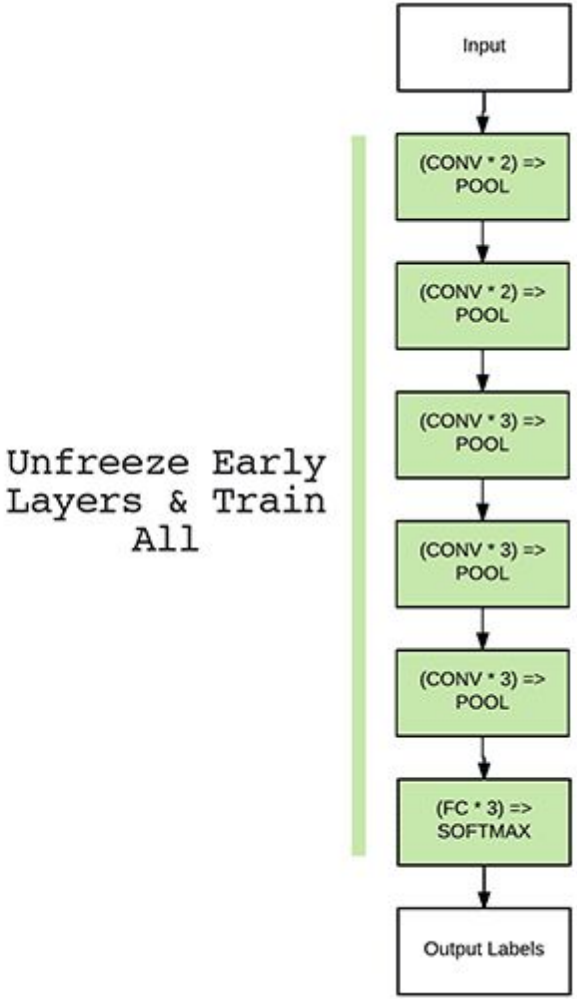
Fine Tuning vs Transfer Learning





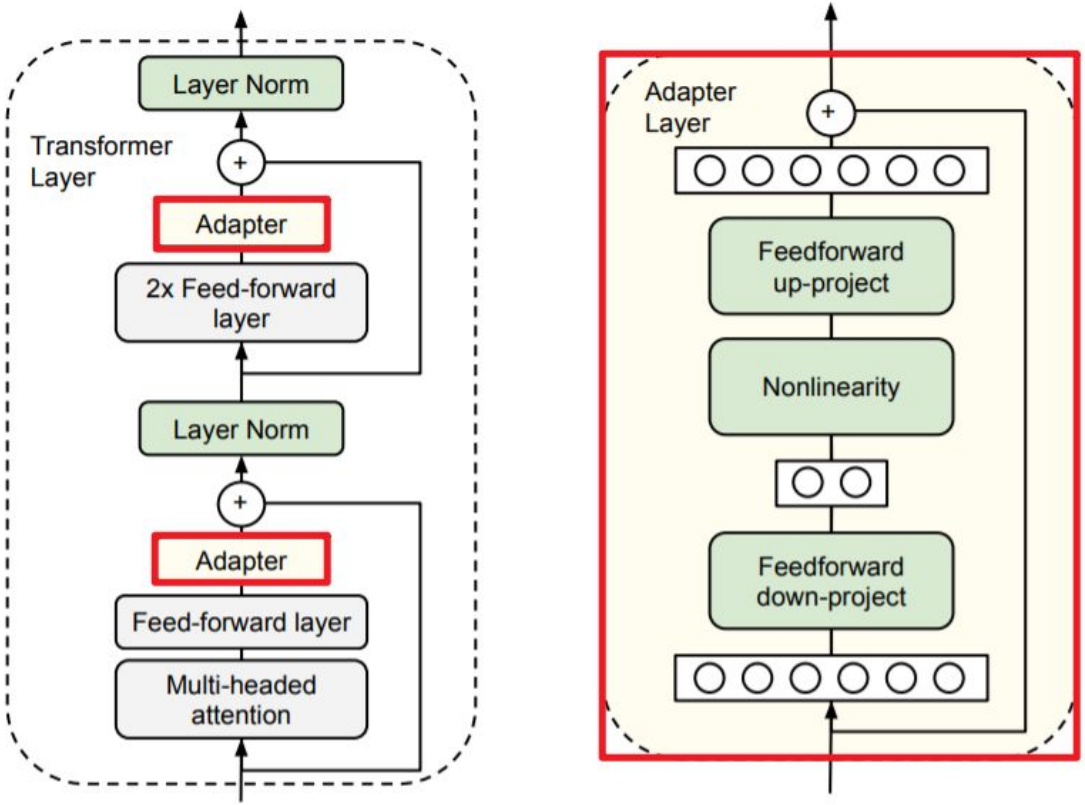
Fine Tuning


Freeze Early Layers in Network
Only Train FC Layers
Input
(CONV * 2) => POOL
(CONV * 2) => POOL
(CONV * 3) => POOL
(CONV * 3) => POOL
(CONV * 3) => POOL
(FC * 3) => SOFTMAX
Output Labels


Unfreeze Early Layers & Train All
Input
(CONV * 2) => POOL
(CONV * 2) => POOL
(CONV * 3) => POOL
(CONV * 3) => POOL
(CONV * 3) => POOL
(FC * 3) => SOFTMAX
Output Labels

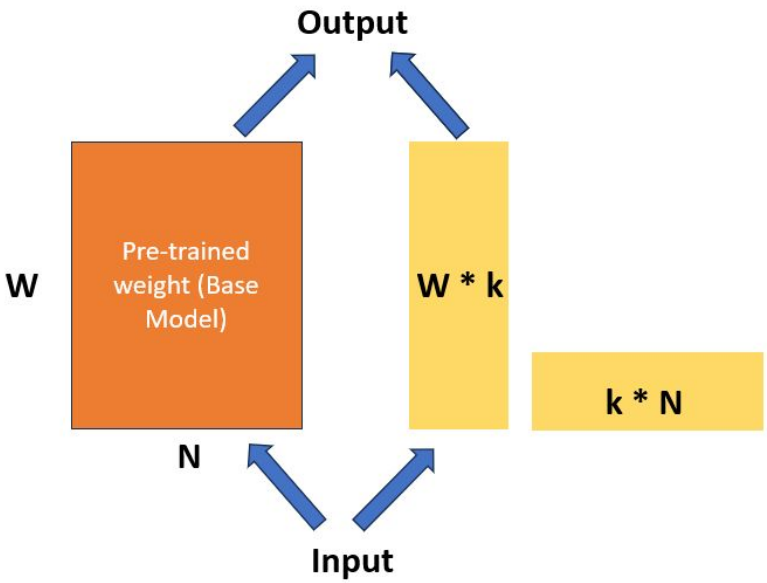
Parameter Efficient Fine Tuning (PFET)

Adapter



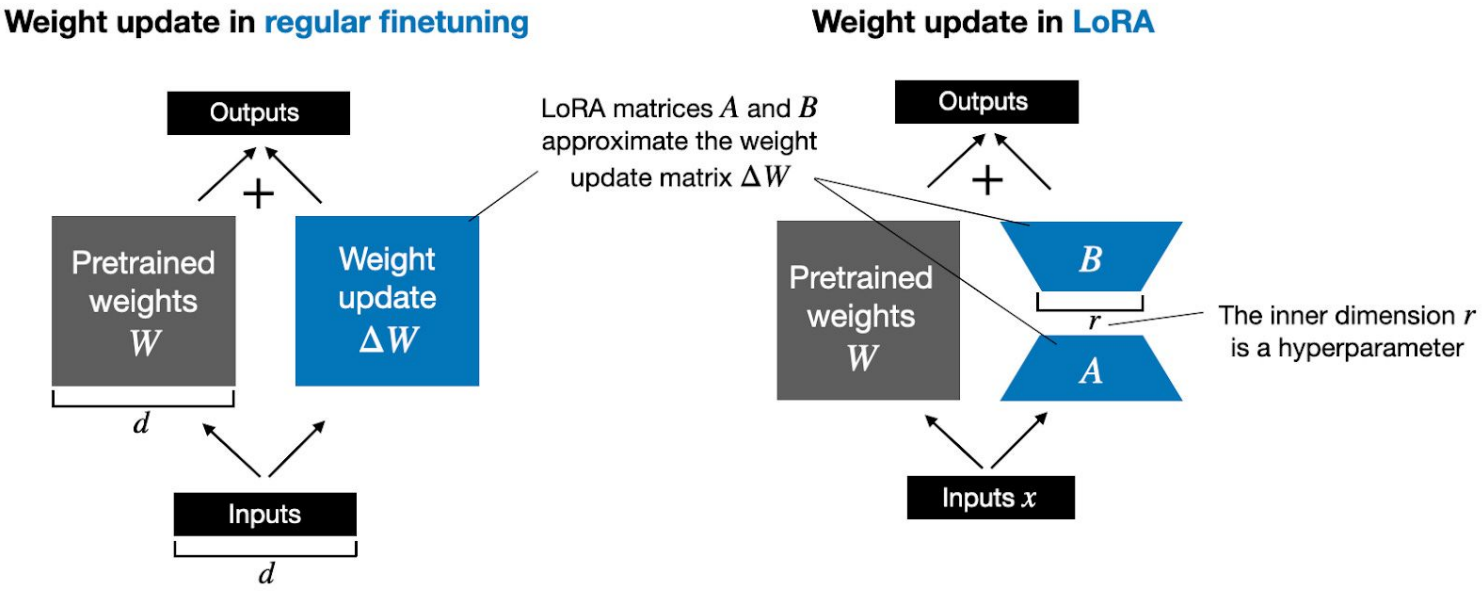
Parameter Efficient Fine Tuning (PFET)

Low Rank Adaptation (LoRA)



Parameter Efficient Fine Tuning (PFET)

LoRA



Datasets

https://huggingface.co/datasets

Natural Language Processing

- Text Classification
- Token Classification
- Table Question Answering
- Question Answering
- Zero-Shot Classification
- Translation
- Summarization
- Feature Extraction
- Text Generation
- Text2Text Generation
- Fill-Mask
- Sentence Similarity
- Table to Text
- Multiple Choice
- Text Retrieval

Magpie-Align/MagpieLM-SFT-Data-v0.1
Viewer • Updated 7 days ago • 550k • 84 • 12

jondurbin/gutenberg-dpo-v0.1
Viewer • Updated Jan 12 • 918 • 465 • 109

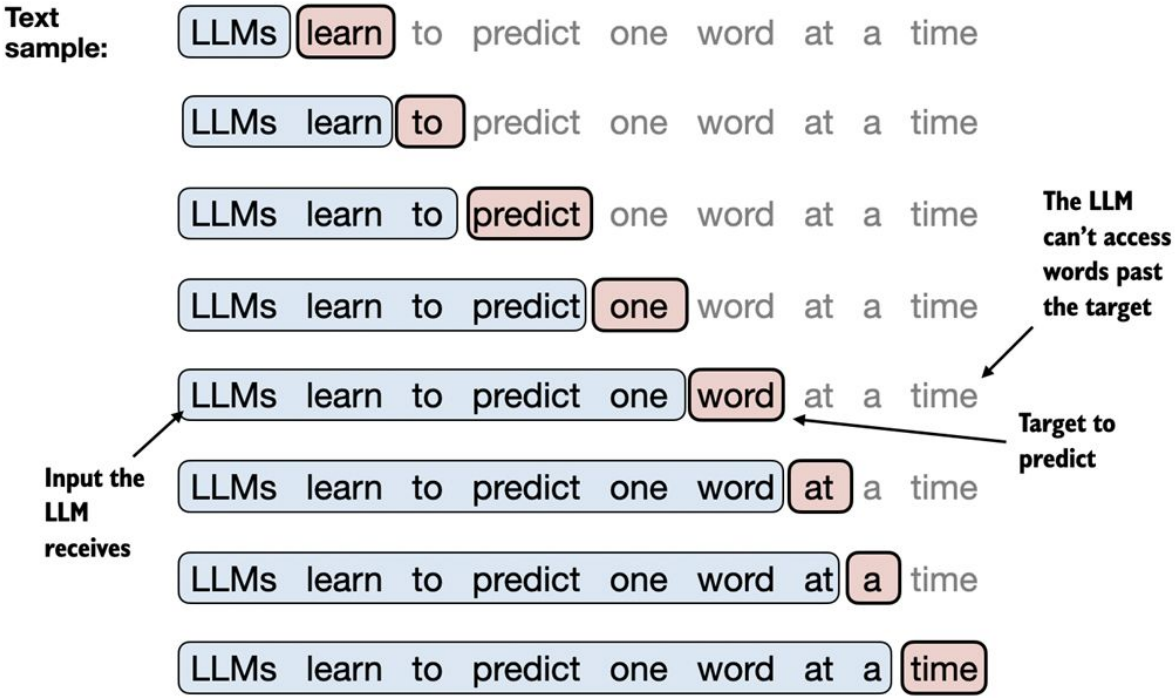
yahma/alpaca-cleaned
Viewer • Updated Apr 11, 2023 • 51.8k • 35k • 553

CaraJ/MMSearch
Viewer • Updated 1 day ago • 900 • 198 • 10

Decoder - PreTraining

Language Model: P(word | LLM learn to predict one)

Next Word Prediction:



Large Language Models (LLMs)

- **Language Model**

  P (example | It is an)

- **Next Word Prediction**

- **Billions of Parameters**

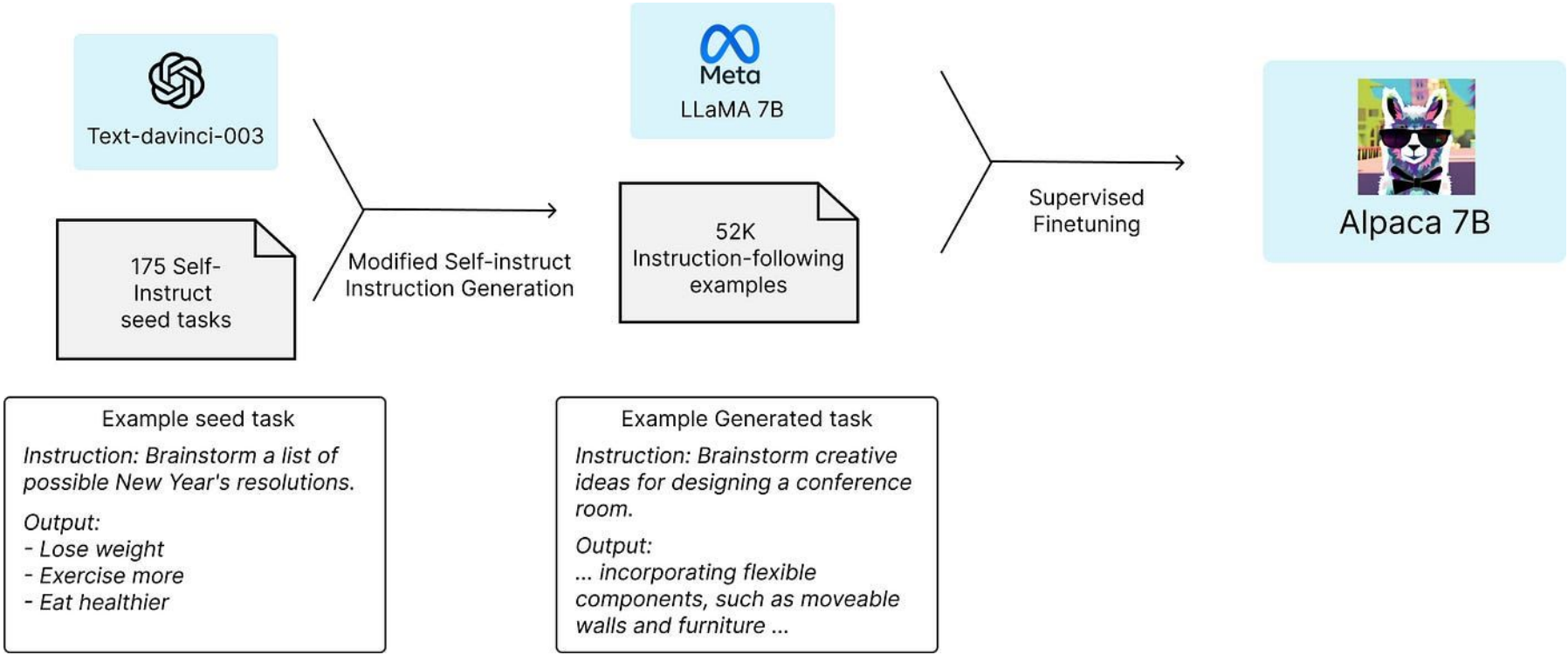
LARGE LANGUAGE MODEL	Parameters
Phi-1.5	1.3B
Phi-2	2.7B
Llama2	7B, 13B, or 70B
BloombergGPT	50B
Claude2	130B
GPT-3	175B
GPT-4 "32k"	1.76T

Context Window Length

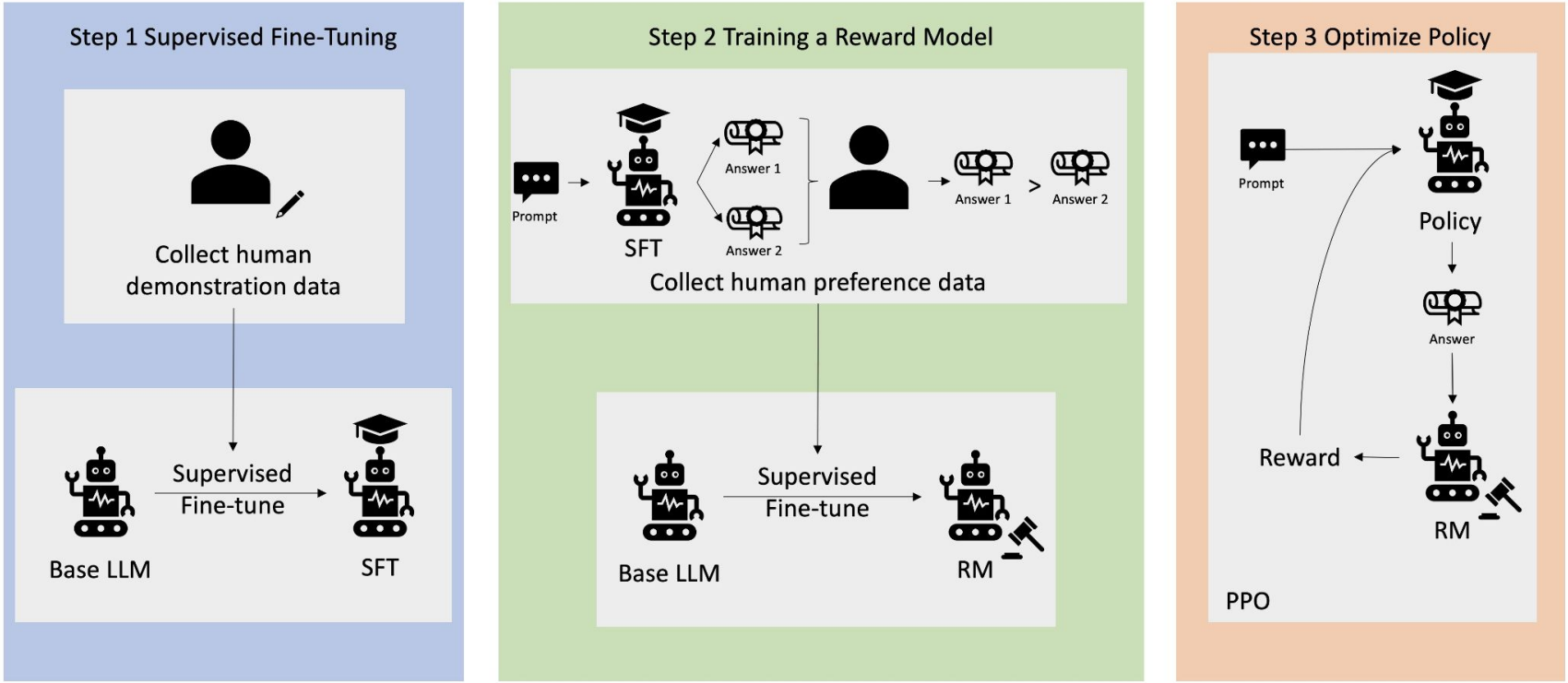
https://www.linkedin.com/pulse/why-does-context-length-matter-sam-shamsan-i11pc

Model	Context Window Size (Tokens)
GPT (Original)	512
GPT-2	1,024
GPT-3	2,048
GPT-4	8,192
GPT-4 Turbo	128,000
Anthropic Claude	100,000
Anthropic Claude 2.0	200,000
LLaMA 1	2,048
LLaMA2	4,096

Prompt Training



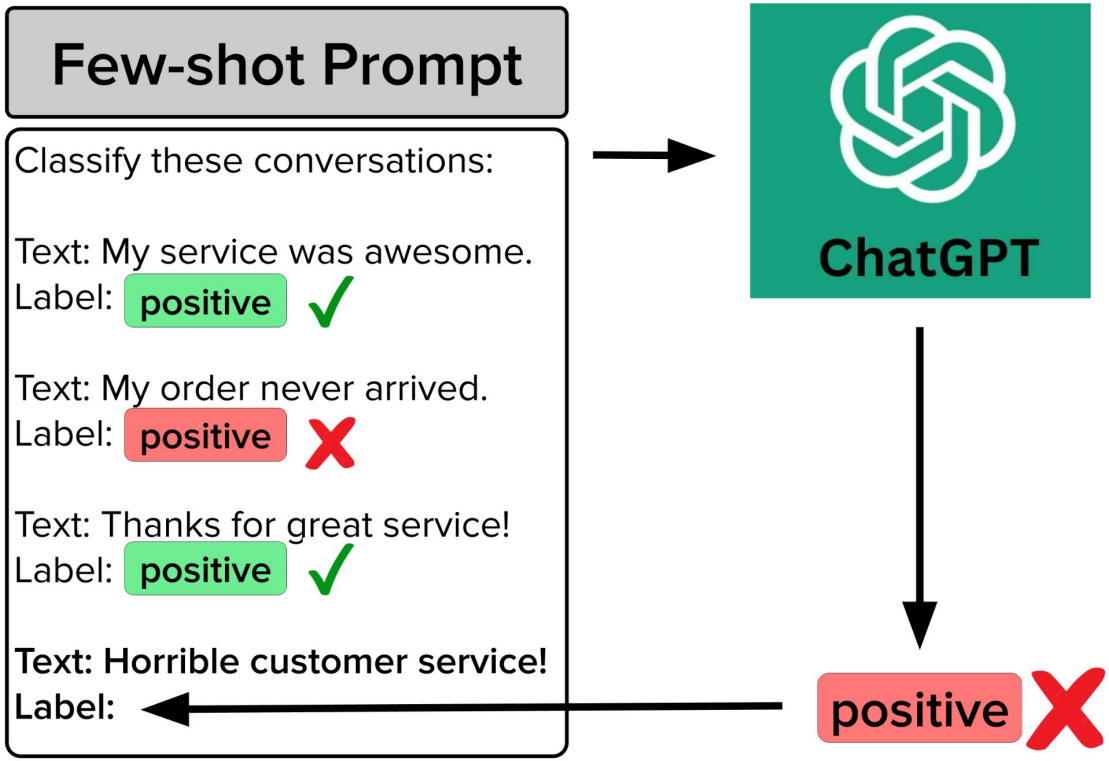
Reinforcement Learning with Human Feedback (RLHF)



Zero Shot, One shot and Few Shot Learning

Zero-Shot Learning	The Model cannot look at any examples from the target class during training
One-Shot Learning	The Model observes one example from the target class during training
k-Shot Learning	The Model observes k-examples from the target class during training

Zero Shot, One shot and Few Shot prompting



Prompt Engineering

AI Prompt Engineering Best Practices

- ☑ Start with the instructions
- ☑ Use “ “ or ### to separate instructions from context
- ☑ Be descriptive and provide multiple output indicators
- ☑ Provide examples for formatting or tone
- ☑ Get rid of unspecific language (ex. somewhat, a few)
- ☑ Focus more on asking what to do vs. what not to do
- ☑ Fine-tune your results by asking follow-up questions

In Context Learning

Chain of Thought Prompting

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ✗

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

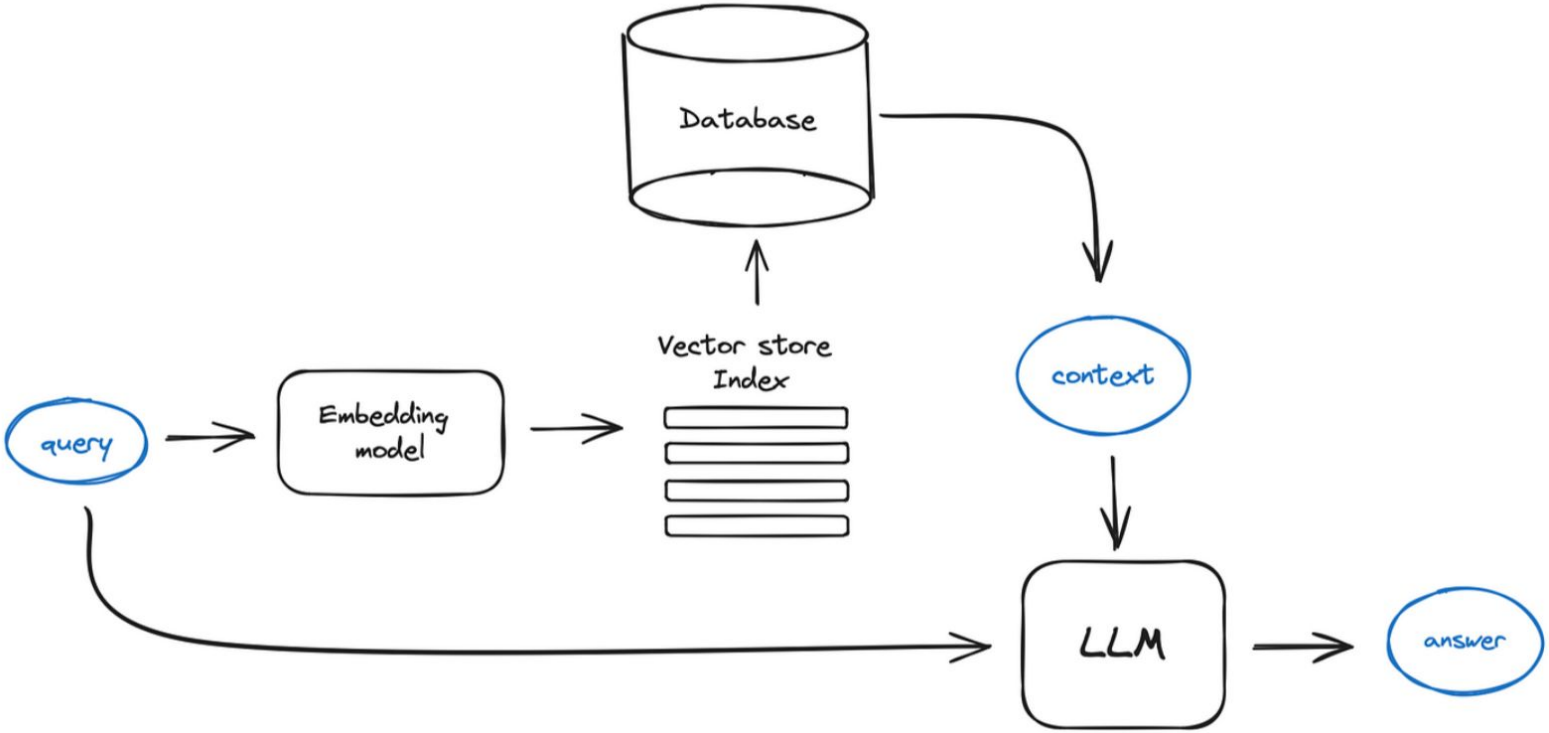
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had 23 - 20 = 3. They bought 6 more apples, so they have 3 + 6 = 9. The answer is 9. ✓

Retrieval Augmented Generation

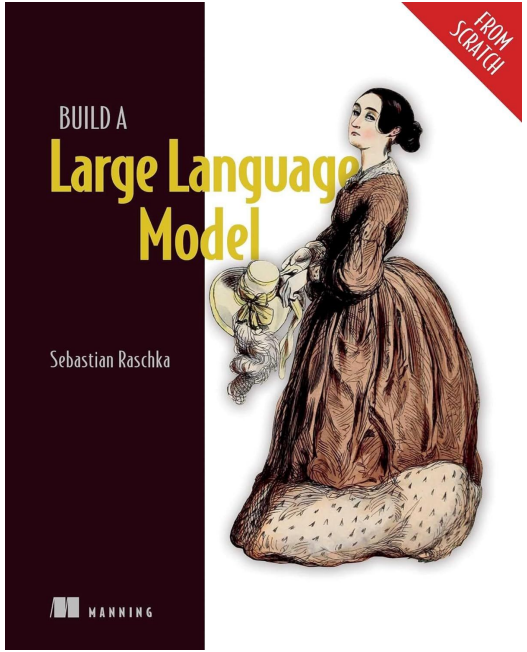


Quantization

	sign	exponent					mantissa										
FP16	0	0	1	1	0	1	1	0	0	1	0	1	0	0	1	1	= 0.395264
BF16	0	0	1	1	1	1	1	0	1	1	0	0	1	0	1	0	= 0.394531
FP8 E4M3	0	0	1	0	1	1	0	1									= 0.40625
FP8 E5M2	0	0	1	1	0	1	1	0									= 0.375

Further Study

https://github.com/rasbt/LLMs-from-scratch



https://github.com/HandsOnLLM/Hands-On-Large-Language-Models