

What do you mean by Training LLMs from Scratch?

Instructor

Sourab Mangrulkar

Machine Learning Engineer at

Creator of PEFT



What is Training LLMs from Scratch?

Pretraining LLMs



Training Objective

Predict the next token

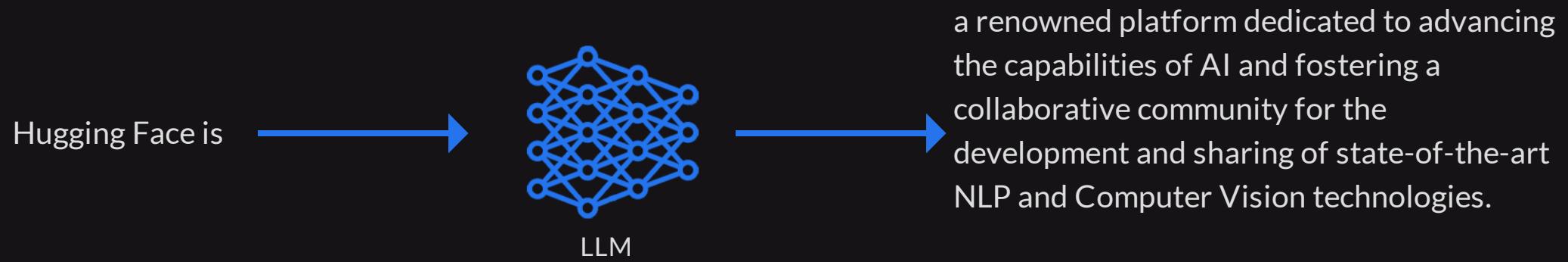
Training Objective: Example

StarCoder and StarCoderBase are Large Language Models for Code

```
194     peft_config = None
195     if args.use_peft_lora:
196         peft_config = LoraConfig(
197             r=args.peft_r,
198             lora_alpha=args.peft_lora_alpha,
199             target_modules=["q_proj", "v_proj"],
200             lora_dropout=args.peft_lora_dropout,
```

Training LLMs from Scratch

Good at continuing the text



Training LLMs from Scratch

- Parameters: Billions and Trillions
- Training dataset: Billions and Trillions of tokens

Challenges

Massive GPU Computing Power

Challenge: Massive GPU computing power



GPT-4

GPT-4	25000 GPUs for 100 days
-------	-------------------------

Challenge: Massive GPU computing power



Llama 2 70B

6000 GPUs for 12 days

Challenges

- Massive GPU computing power
- Cost: Millions of \$\$\$

Small Language Models

Phi 1.5

32 A100 40GB GPUs for 8 days

Thank You
