Model Alignment, Prompting, and Learning

Phạm Quang Nhật Minh

minhpham0902@gmail.com

Recap

- We learned how to pre-train a language model by self-supervised learning method
- In this lecture, we will learn how to fine-tune a pre-trained model so that the model can follow instructions

Prompt: Sentimen Analysis

We would like to classify a hotel review into positive or negative

Did not like the service that I was provided, when I entered the hotel. I also did not like the area, in which the hotel was located. Too much noise and events going on for me to feel relax.

How to do sentiment analysis

Write a prompt with the review plus an incomplete statement

Did not like the service that I was provided, when I entered the hotel. I also did not like the area, in which the hotel was located. Too much noise and events going on for me to feel relax. In short, our stay was

Generated by LLM

...disappointing and I would not recommend this hotel to anyone looking for a peaceful stay.

Prompts

- Instructions for language models
- Get models to perform tasks
- Simply by describing in natural language
- Creates context for useful output

Prompts templates

- **Summarization**: {input} ; tldr;
- **Translation**: {input} ; translate to French:
- **Sentiment**: {input}; Overall, it was
- **Fine-Grained-Sentiment**: {input}; What aspects were important in this review?

Prompt Design

- We can constrain generation with prompt design
- Specify potential answers in prompt

Human: Do you think that “input” has negative or positive sentiment?
Choices:
(P) Positive
(N) Negative
Assistant: I believe the best answer is: (

More advanced prompting

- Specify roles, restrict output length
- Generate structured output (JSON, code)
- Break down complex tasks (chain-of-thought)

Prompting process

- Develop task-specific template
- Instantiate filled prompt with input
- Generate token sequence with LLM
- Extract or use output directly

Do you think that the review “Did not like the service that I was provided, when I entered the hotel. I also did not like the area, in which the hotel was located. Too much noise and events going on for me to feel relax.” has negative or positive sentiment
Choices:
(P) Positive
(N) Negative
Assistant: I believe the best answer is: (

N) Negative

Few-shot prompting example

Definition: This task is about writing a correct answer for the reading comprehension task. Based on the information provided in a given passage, you should identify the shortest continuous text span from the passage that serves as an answer to the given question. Avoid answers that are incorrect or provides incomplete justification for the question.

Passage: Beyoncé Giselle Knowles-Carter (born September 4, 1981) is an American singer, songwriter, record producer and actress. Born and raised in Houston, Texas, she performed in various singing and dancing competitions as a child, and rose to fame in the late 1990s as lead singer of R&B girl-group Destiny's Child. Managed by her father, Mathew Knowles, the group became one of the world's best-selling girl groups of all time. Their hiatus saw the release of Beyoncé's debut album, Dangerously in Love (2003), which established her as a solo artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100 number-one singles "Crazy in Love" and "Baby Boy".

Examples:

Q: In what city and state did Beyoncé grow up?
A: Houston, Texas

Q: What areas did Beyoncé compete in when she was growing up?
A: singing and dancing

Q: When did Beyoncé release Dangerously in Love?
A: 2003

Q: When did Beyoncé start becoming popular?
A:

Few-shot prompting

- Include labeled examples (demonstrations)
- Demonstrations improve prompt effectiveness
- Contrasted with zero-shot (no examples)

Number of demonstrations

- Small number often sufficient
- Largest gains from first example
- Why fewer demonstrations work:
  - Demonstrate task and format
  - Not provide relevant answer information
  - Incorrect answers can still help

Selecting demonstrations

- Format examples from labeled training set
- Use similar examples to current input
- Dynamically retrieve based on similarity

In-context learning

- Learning from prompts
- Improves token prediction as processing continues
- Can teach model new tasks

Base model limit

Prompt: Explain the moon landing to a six year old in a few sentences.
Output: Explain the theory of gravity to a 6 year old.

Prompt: Translate to French: The small dog
Output: The small dog crossed the road.

Problems with base LLMs

- Ignore instructions, continue text instead
- Generate false or toxic content
- Pretraining objective misaligned with human needs

Post-training and model alignment

- Address limits of pretrained LLMs
- Make models more helpful, less harmful

Two techniques for model alignment

- Instruction tuning (supervised finetuning)
- Preference alignment (e.g. RLHF)

Instruction tuning

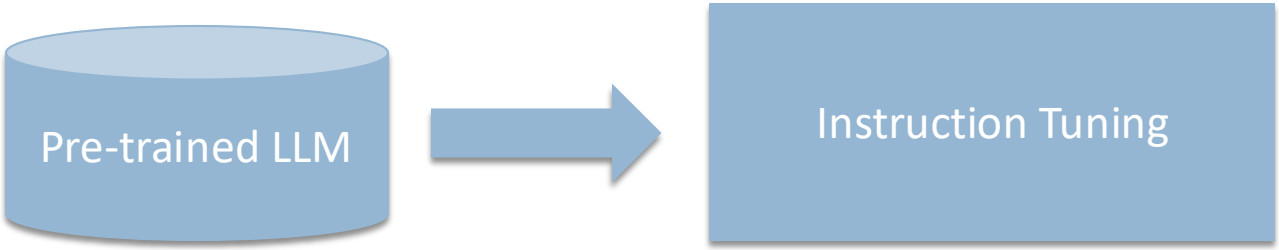
- Finetune on instructions
  - With corresponding correct responses
- Improves ability to follow instructions

Preference alignment

- Train separate model
  - Decide if response aligns with preferences
  - Use to finetune base model

Instruction tuning process

- Take pretrained LLM
- Train on instruction-response corpus
- Improves instruction-following, enables meta-learning



Instruction tuning vs other finetuning

- Continues training on instructions
- Uses same language modeling objective
- Called "supervised" due to correct responses

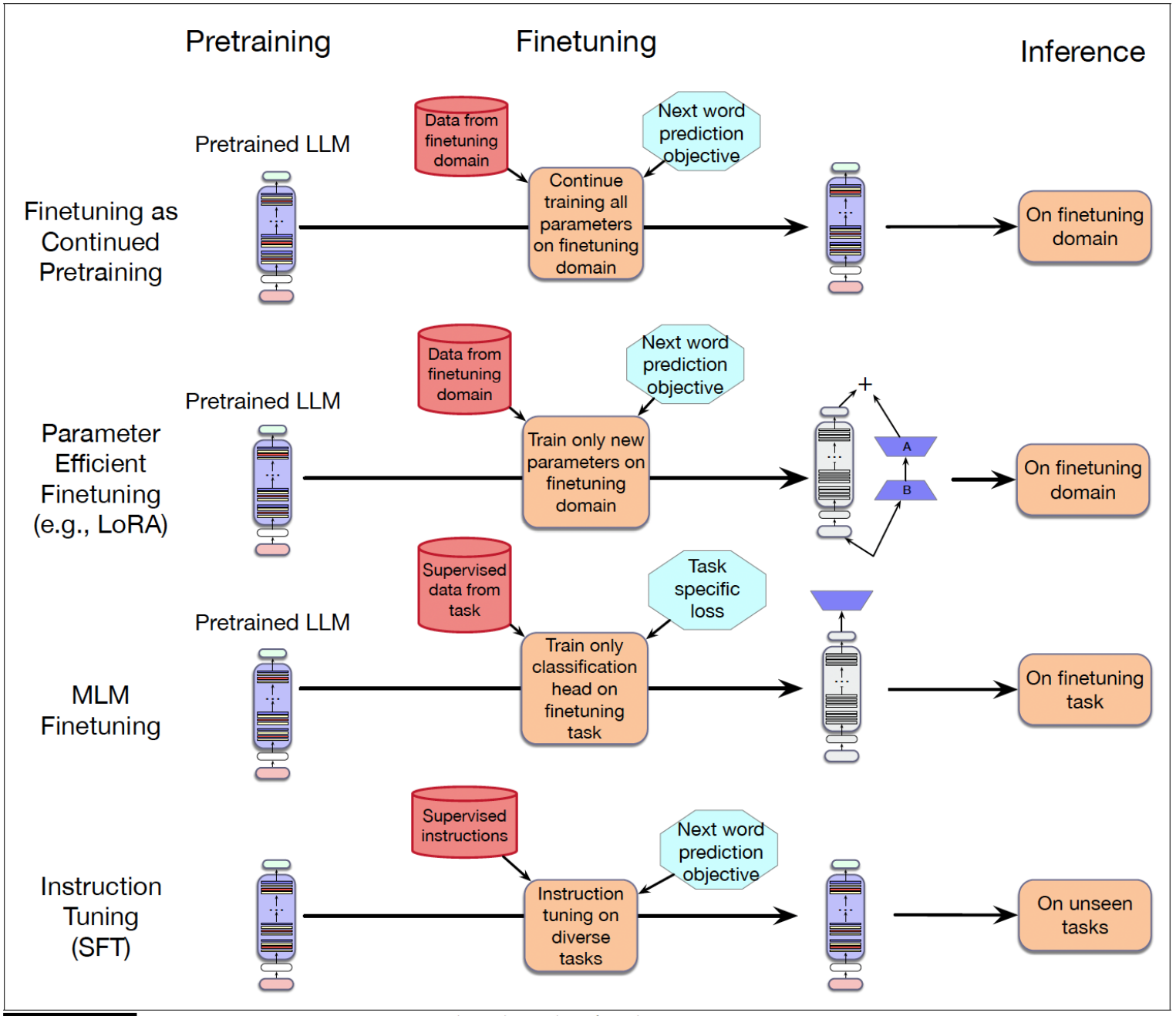


Figure 12.4 Instruction tuning compared to the other kinds of finetuning.

Instructions as training data

- Natural language task descriptions
- Combined with labeled task demonstrations
- Can include constraints, personas, examples

Example:
https://huggingface.co/datasets/Muennighoff/natural-instructions?row=0

Large instruction tuning datasets

- Aya: 503M instructions, 114 languages
- SuperNatural Instructions: 12M examples, 1600 tasks
- Flan 2022: 15M examples, 1836 tasks

Creating instruction-tuning datasets

- Direct writing by people
- Convert existing supervised datasets
- Use annotation guidelines as prompts
- Generate with language models

Evaluation of instruction-tuned models

- Assess performance on novel tasks
- Use leave-one-out approach on task clusters
- Allows uniform task-appropriate metric

Chain-of-thought prompting

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. ✅

@akshay_pachaar

source: https://arxiv.org/pdf/2201.11903.pdf

Chain-of-thought process

- Add reasoning text to demonstrations
- Encourages model to output similar steps
- Leads to more correct answers

Summary

- **Simple Prompting**: Maps real-world applications to LLMs without modifying the model.
- **Few-Shot Learning**: Uses labeled examples to guide the model's response.
- **Chain-of-Thought**: Designs prompts to improve complex reasoning.
- **Model Alignment**: Adjusts pretrained models to desired behaviors.
- **Instruction Tuning**: Finetunes models using instruction-response datasets, often derived from NLP tasks.