Instruction Tuning and RLHF

- What was the biggest step in going from GPT-3 to 3.5?
- Why do we even do instruction tuning after pre-training?
- Distribution mismatch!

More SFT Recap

- The language humans use on the internet isn't reflective of how they talk to models for everyday purposes
- Collecting instructions lets model efficiently respond
- Basically, models aren't great at OOD generalization still

Difference from pre-training

- No loss generated on prompts
- Generally a few orders of magnitude less data
- Misconception that you can't learn new info during FT, you can – its just harder because there is less data

Why does prompting matter?

- Attention (transformer self-attention w/ position embeddings) is Turing Complete (Perez et al 2021)
- Autoregressive LLMs can simulate a universal Turing Machine (Schurmanns et al 2024)
- One view is that models learn “programs” and prompts are a way of indexing
- There exists a prompt for any given program

What does changing a prompt do?

- In practice finding this prompt is very difficult… Infinite Monkey Theorem
- Really dumb sounding prompts work
  - “believe in yourself”
  - “I am an incredible 10x expert”
- Why?

Sources of Model Error

- Models are very sensitive to prompts and even the format of the prompts
- Generation sampling parameters also can cause much error
- It is a UX issue: much of the knowledge required for a task exists from pre-training but no one wants to poke a model many times

Terminology

- What do zero / one / few shot mean? What is the “standard” way of doing it?
- Whatever the companies tried that works best on a benchmark.
- Zero shot = just prompt, one shot = prompt + one example in context, etc
- Some evidence that examples are teaching only formatting and not new knowledge (Min et al 2022)

Chain of Thought

"Let's think step by step"

- Breaking the problem down into smaller steps works!
- Can do more attention computations
- Can more easily draw from traces of "human" reasoning
- Will talk more about this when we discuss test-time scaling

ReAct

- Similar concept, break down problem to first think before you act

RAG: Retrieval Augmented Generation

- Two steps:
  - Retrieval:
    - Take lots of documents, embed them
    - Create vector DB of all documents
    - For a given query find the n most relevant docs
  - Generation:
    - Shove all the retrieved docs into context + ask question
- E.g.
- Much of the bottleneck is in the Retriever. Current LLMs are good at generating well given the right set of docs.
- The docs are prompting view point
  - Will talk about the retrieval is memory view point later

RAG Usecases

- Grounding response in the docs reduces hallucinations
- Most enterprise usecases are just RAG systems
  - Talk to your internal docs
  - Talk to your database
  - Talk to your codebase

RAG vs Long Context

RAG

Pros:
- Cheaper
- Most tasks aren't hard enough to need long context

Cons
- Bottlenecked by retriever and human design
- Harder to prototype

Long Context

Pros:
- Less hassle, just shove it all into context and let model figure it out
- Learning is prob better than human knowledge hard coding for harder problems

Cons
- Computationally expensive
- Current methods don't work too well

Automatic Prompt Optimization - RL

- Fix a big LLM as the “environment”
- Train a smaller LLM to prompt the bigger one
- Reward is how well the bigger one does at a task
  - E.g. RLPrompt (Deng et al 2022)
- Prompt “rewriting” also works well for
  - Diffusion (LLM as input to translate into diffusion prompts)
  - Text-games (LLM as the input to translate into parser commands)
  - Small to big LLM

Automatic Prompt Optimization - Bayesian

- DsPY, I do not understand how this works. Read Omar Khattab's papers.

Prithvi's Pet Peeves of Prompting Papers

- Many NLP papers between GPT-3/4 found very bespoke methods of prompting that work on a specific closed source API models in some cases
- These are all **useless**
- Prompting/in-context is only interesting scientifically insofar it gives you repeatable results and tells you something about
  - How to train a model
  - How to change data that is used to train the model