Reasoning in Language Models

core: CoT... are not new model architecture, but a prompt/generation strategy

① 2 taxonomy

- deductive reasoning
- Inductive reasoning
- abductive (溯因) reasoning
- Formal reasoning
- Informal reasoning

→ most we use in LMs

② Can LLMs really reason?

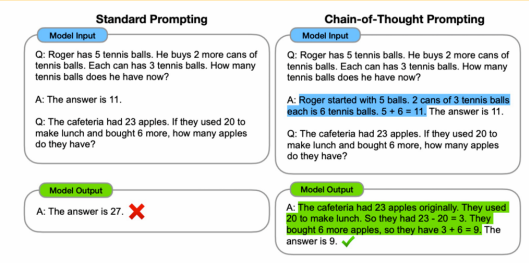
Large Language models are **REALLY GOOD** at predicting **plausible continuations of text (Lecture-9)**, that respect **constraints in the input (Lecture 10,11)**, and align **well with human preferences (Lecture-10, 11)**.

↳ core question: are these "reasoning steps" genuine reasoning or merely a fit of linguistic patterns? — idk

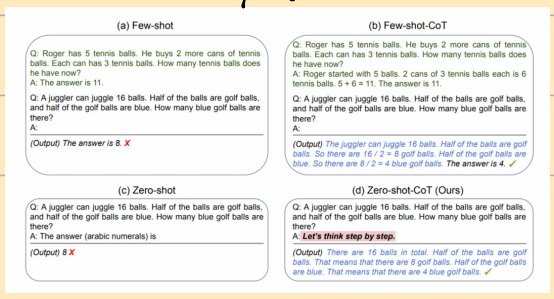
③ Main Approaches to improve reasoning

△ prompting-based methods

1. Chain-of-Thought (CoT): explicitly ask the model to show intermediate steps



2. zero-shot / few-shot CoT

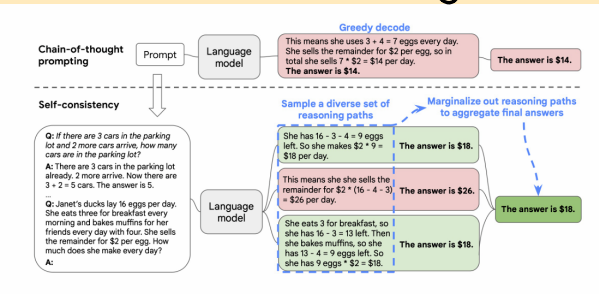


3. CoT with "self-consistency":

Replace greedy decoding with an ensemble of samples

Basic idea: if an answer keeps appearing for multiple rationales, the majority of the rationales agree on, then it's more likely to be correct

↳ That sounds like ensembling, but in practice, self-consistency is doing more than simple ensembling



4. Least-to-most prompting: solve simpler subproblems first

Q: Elsa has 5 apples. Anna has 2 more apples than Elsa. How many apples do they have together?
A: Anna has 2 more apples than Elsa, so Anna has 2 + 5 = 7 apples. Elsa and Anna have 5 + 7 = 12 apples together. The answer is 12.

Standard CoT prompting

Q: Elsa has 5 apples. Anna has 2 more apples than Elsa. How many apples do they have together?
A: Let's break down this problem: 1. How many apples does Anna have? 2. How many apples do Elsa and Anna have together?
1. Anna has 2 more apples than Elsa. So Anna has 2 + 5 = 7 apples.
2. Elsa and Anna have 5 + 7 = 12 apples together.

Q: {question}
A: Let's break down this problem:
—
The answer is:

Least-to-most prompting

Accuracy by Steps (GSM8K)	All	2 Steps	3 Steps	4 steps	≥ 5 steps
Least-to-Most	**62.39**	74.53	**68.91**	**59.73**	**45.23**
Chain-of-Thought	60.87	**76.68**	67.29	59.39	39.07

Generalizes to more #steps than in-context example!

Prompting method	**Accuracy**
Zero-Shot	16.38
Standard prompting	17.06[3]
Chain-of-Thought (original)	61.18
Chain-of-Thought (1-shot)	60.88
Least-to-Most (1-shot)	62.39
Chain-of-Thought (best)	**68.61**[3]
Least-to-Most (best)	68.01

But with enough prompt engineering, CoT ≈ Least-to-Most

⚠2 Distillation of reasoning

In ⚠1, we might think: instead of trying to get really large language models to do reasoning, maybe we want to somehow get this kind of reasoning behavior in a smaller LM

Idea:
- Use strong models (e.g., GPT-4) to generate reasoning traces
  ↓
- Fine-Tune smaller models on these traces

Representative work : Orca

1. Collect a wide variety of instructions from the FLAN-v2 collection.
2. Prompt GPT4 or ChatGPT with these instructions along with a system message
3. Finetune Llama-13b on outputs generated via ChatGPT + GPT4

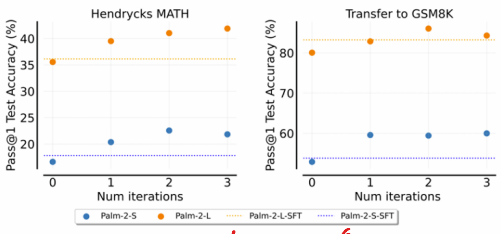
③ Self-Generated Reasoning Data

BC: in ② we might think: why not just finetune the big LM on its own rationales

one of the methods: Reinforced self-training (ReST)

ReST^{EM} alternates between the following two steps:

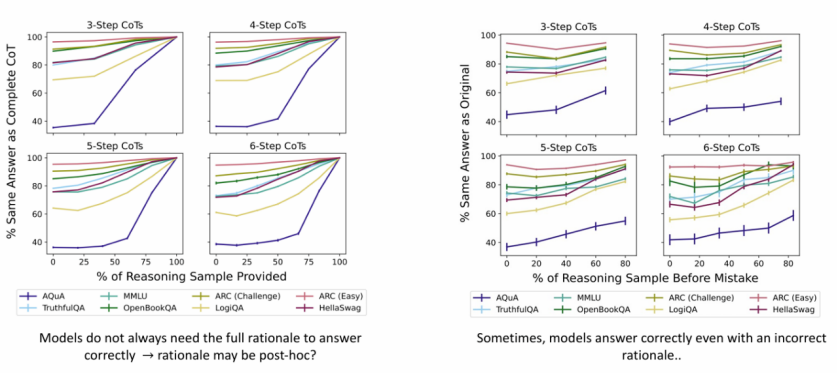
1. `Generate (E-Step)`: Given reasoning problem, sample multiple solutions from language model. Filter based on some (problem specific) function [answer correctness for math problems]
2. `Improve (M-Step)`: Update the language model to maximize probability of filtered solutions, using supervised finetuning



↳ we did supervised fine tuning on human provided rationales

④ Challenges in Reasoning Evaluation

[1] CoT Rationales are often not faithful



Models do not always need the full rationale to answer correctly → rationale may be post-hoc?

Sometimes, models answer correctly even with an incorrect rationale..

[2] Models may rely on memorization rather than reasoning

↳ So how to distinguish ?

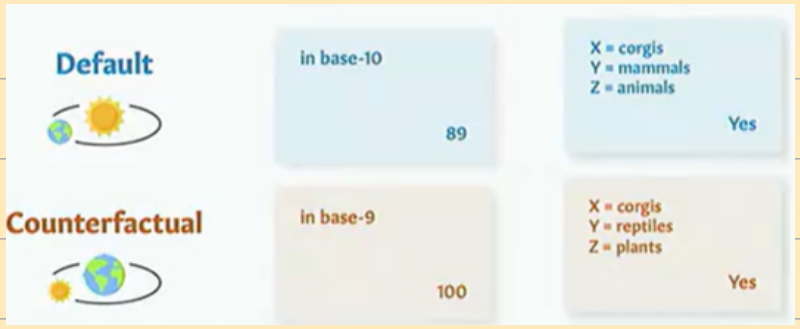
↳ Using Counterfactuals (反事实)

What is Counterfactuals?

Expect to not be present that frequently in the training data

or say become rightly unnatural

eg:



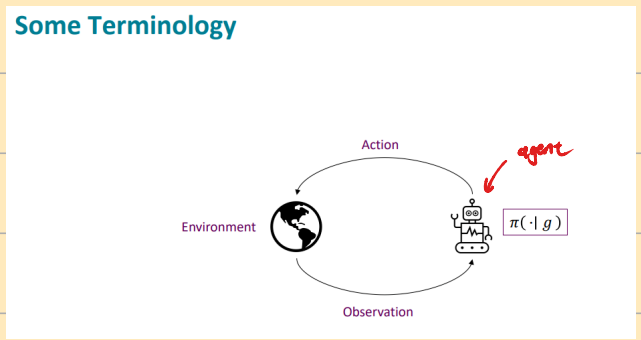
By using it, we found that performance drops significantly under counterfactual tests

⇓

This suggest current reasoning abilities are fragile and non-systematic

Language Model Agents

① Concept



this setting also called
"Instructed-following agent"
"Language conditioned policy"
"digital agent"

agent receives an observation from its environment

⇓

based on the observation, agent issues an action

along with that, agent relives second variable g

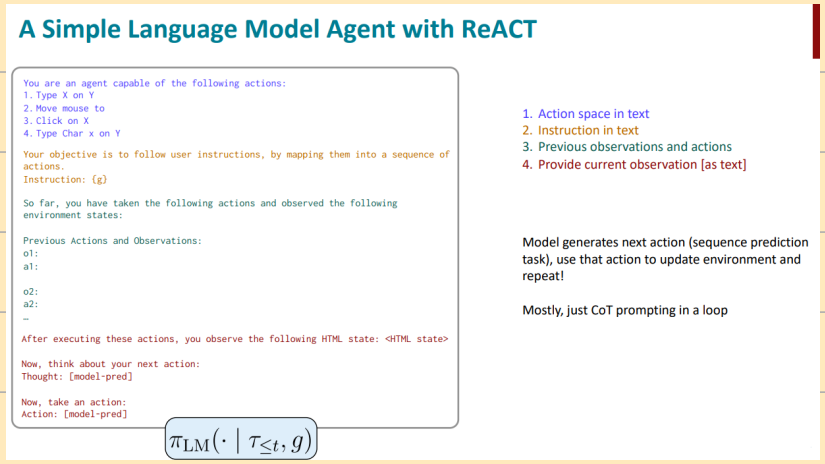
g represents a language instruction

2 Instruction following agents [Pre LLMs]

1. training semantic parsers
2. Infering plans from instruction-trajectory pairs
3. modeling plans directly and then have an execution model that can execute plans
4. doing reinforcement learning with reward signal

3 Instruction following agents [in 2024]

main idea: Generative trajetory modeling with causal transformers



Some popular benchmarks for LM agents

MiniWoB++, WebArena, WebLINX

1 training data for Languages model Agents

Standard practice: In-context learning with few-shot demonstrations of humans performing following similar instructions

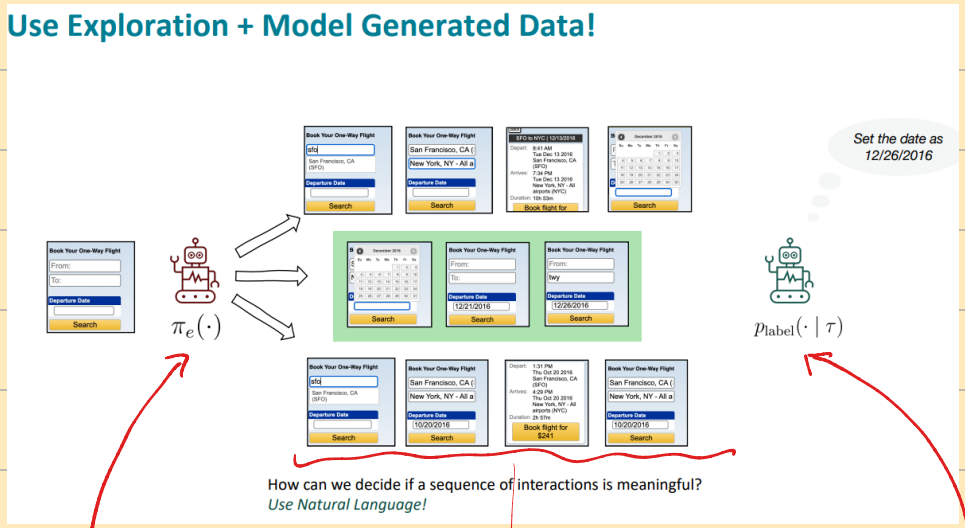
↳ But this's not scalable/reliable (cause 1000s of environments, many kinds of interactions possible)

Is there somethings better that we can do than just sort of getting humans to provide demonstrations for every new use case?

Using LM to generate rationales and then fine tune on that!

In here, we don't have rationales, but we could produce action trajectories. And than we're going to use that as supervision.

↳ example:



a agent that can randomly explore the environment

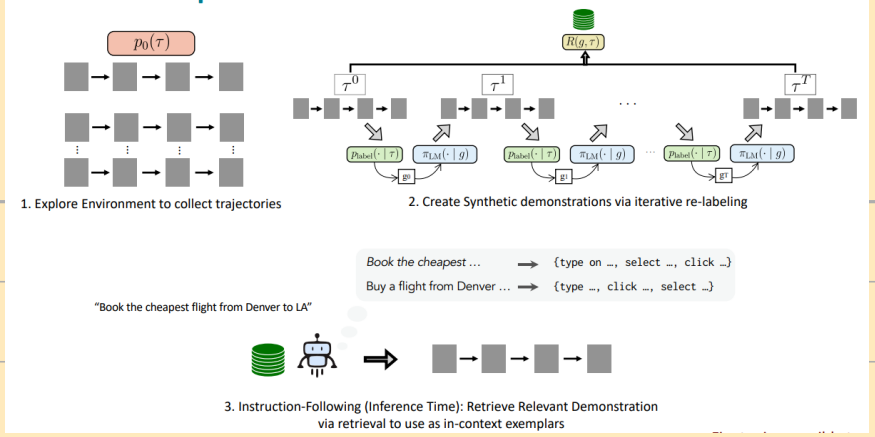
agent! produces some trajectories

produce a description of these trajectories

- if the trajectory seems pretty good for completing tasks
  - we add it to a set of examples
- if not

Instead of throwing away this interaction (Cause interactions are pretty costly) we invoke the re-labeler to take the trajectory and assign it a new label

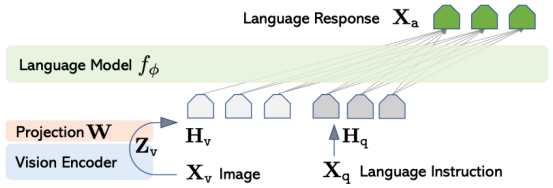
BAGEL (BootStrapping Agents by Guiding exploration with Languages)



(2) Multimodality

- So far, we've looked at using text-only language models for agents
- This is intractable for real-world UIs with very long HTML
- Can we instead operate directly over pixel space?

LLaVA



Prompt GPT-4 to generate instructions and responses given textual descriptions of images

Finetune a CLIP encoder jointly with a Vicuna-13B decoder on this data

Pix2Struct

Finetune a ViT encoder and a transformer decoder on a new HTML screenshot parsing task