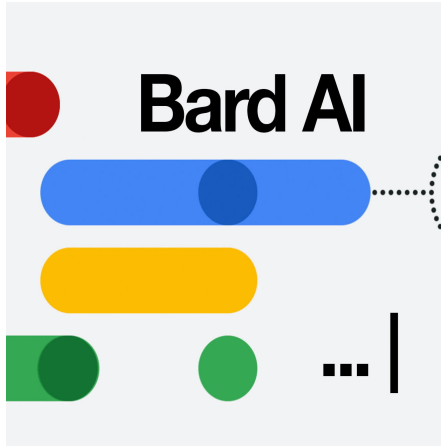
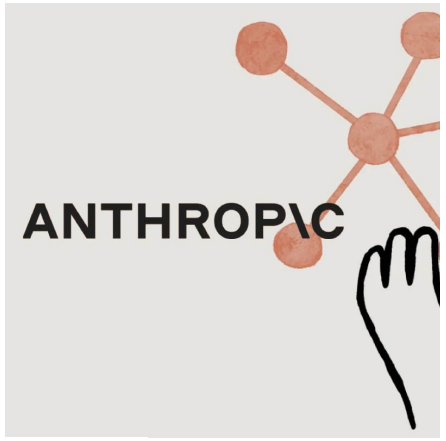
Growing Chatbots Out of Thin-Air:

Opportunities and Limits of Language Models for Guiding Themselves

Daniel Khashabi

The success we dreamed of





Language models that are remarkably capable at solving many important NLP benchmarks.

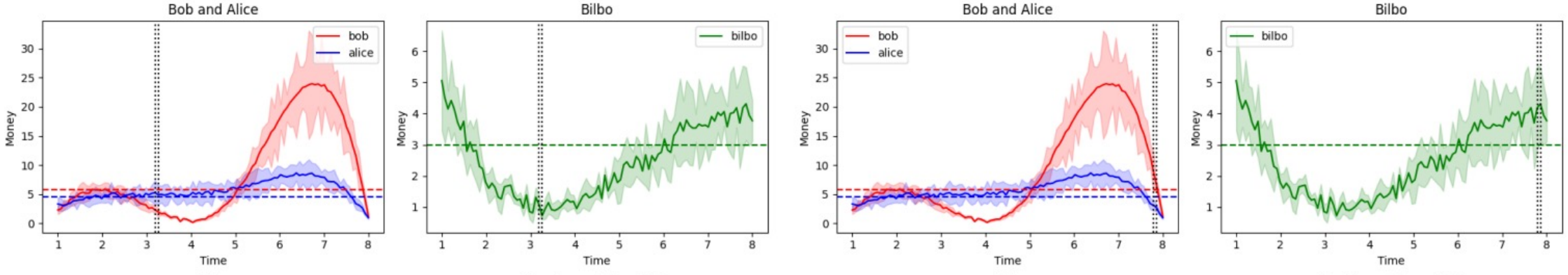
[Bubeck et al. 2023]

The Success We Dreamed Of

Prompt: Can you generate a pyplot for the following data: $x = [1, 3, 5, 6, 8], y1 = [2, 3, 5, 18, 1], y2 = [3, 5, 6, 8, 1], y3 = [5, 1, 2, 3, 4], y4 = [9, 7, 2, 3, 1]$. I want $y1, y2$ to be in the same plot, but $y3$ is in another plot next to that plot, $y4$ is in below. I want the legend of $y1$ to be "bob", $y2$ to be "alice", $y3$ to be "bilbo", $y4$ to be "allie". I want the x-axis to be labeled with "time" and y axis to be labeled with "money". I want to add a 10 %-40% random error bar to each curve, through all times (including non-integers). I want smoothed curves to show the plot, and smoothed error bar. Do not use linear interpolation, use smooth interpolation! I want to also add some small zig-zag to the smoothed curve to make it look more real. I want to put a baseline as the mean of each line. I want to put a pie chart below indicating the fraction of the four people in each time step. I also want an animation to show how the fractions are changing in the pie chart in **continuous time**. Interpolate the missing fractions! I also want vertical line animation in other three plots to match the pie chart. I want the fanciest plot. Please add as many fancy things as possible.

Here is a possible code to generate a pyplot for the following data, with some comments to explain the steps:

GPT-4: [We give two snapshots from the animation resulting from the produced code]

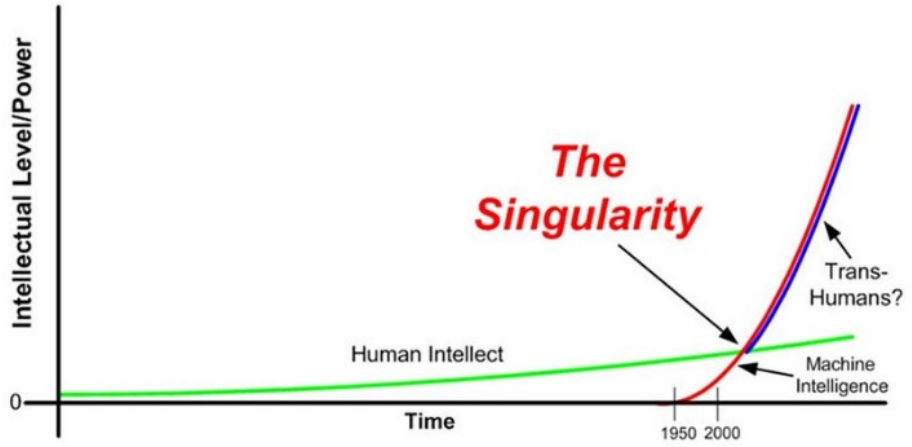


Eutopia/Dystopia where LLMs Self-Improve.

- What if LLMs can improve themselves?

LARGE LANGUAGE MODELS CAN SELF-IMPROVE

Jiaxin Huang[1*] **Shixiang Shane Gu**[2] **Le Hou**[2†] **Yuexin Wu**[2] **Xuezhi Wang**[2]
Hongkun Yu[2] **Jiawei Han**[1]
[1]University of Illinois at Urbana-Champaign [2]Google
[1]{jiaxinh3, hanj}@illinois.edu [2]{shanegu, lehou, crickwu,
xuezhiw, hongkuny}@google.com



Nick Bryant
@nickbryantfyi

The most groundbreaking AI development nobody's talking about:

Auto-GPT.

This self-improving AI represents the first spark of a true AGI.

Here's the breakdown (with 7 mind-boggling future use cases):

Torantulino/**Auto-GPT**

An experimental open-source attempt to make GPT-4 fully autonomous.

19 Contributors · 95 Issues · 19 Discussions · 9k Stars · 828 Forks

8:33 AM · Apr 6, 2023 · **152.2K** Views

7 · 24 · 111 · 158

Today

- How realistic are these expectations?
- Do we see any evidence that AI/LLMs self-grow?

Training-time
Self-Feedback

Inference-time
Self-Feedback

Today

- How realistic are these expectations?
- Do we see any evidence that AI/LLMs self-grow?

Training-time
Self-Feedback

Inference-time
Self-Feedback

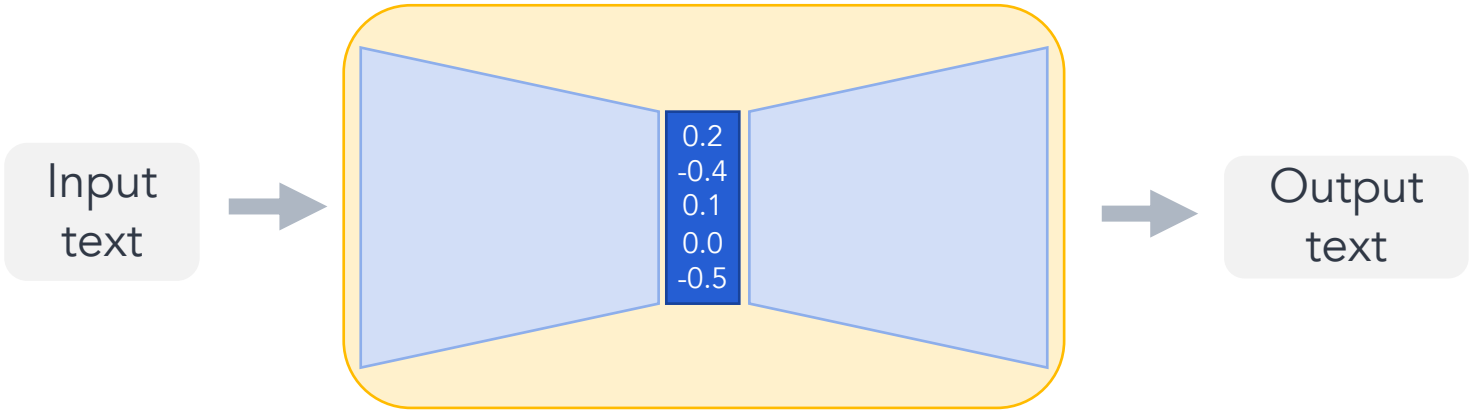
Background

Language Models

[Bengio et al. ‘04, Peters et al. ‘18, Raffel et al. ‘20, Brown et al. ’20, many others]

Background

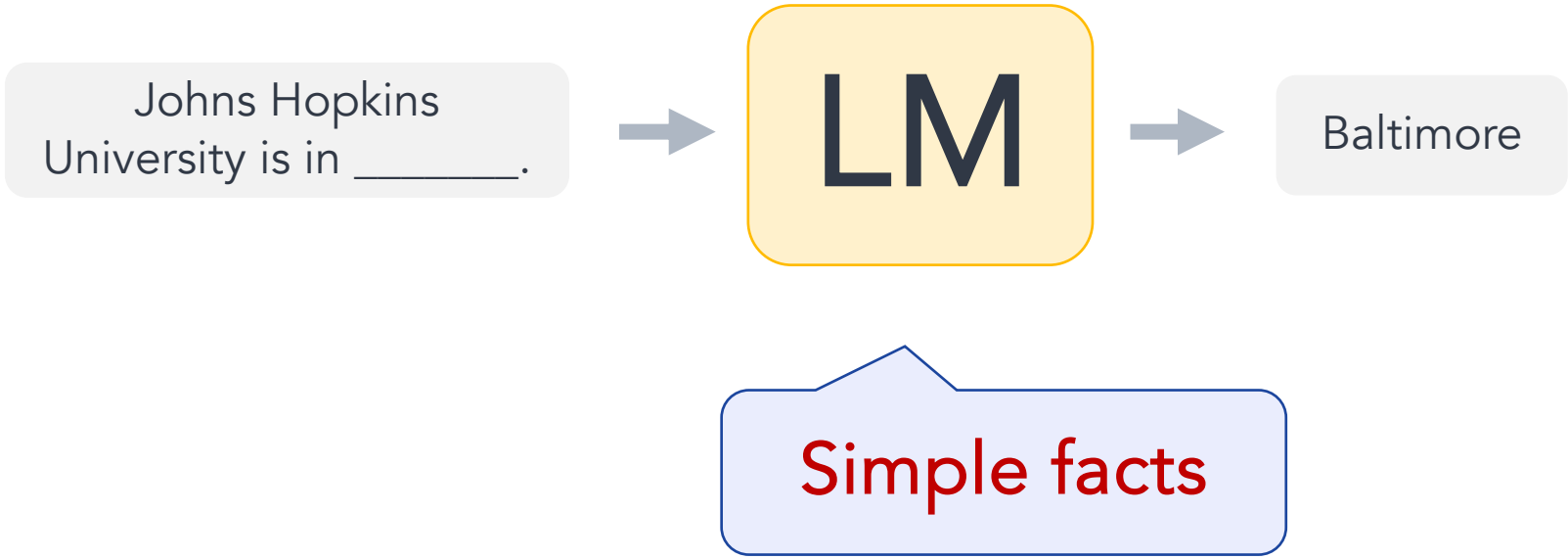
Language Models



[Bengio et al. ‘04, Peters et al. ‘18, Raffel et al. ‘20, Brown et al. ’20, many others]

Background

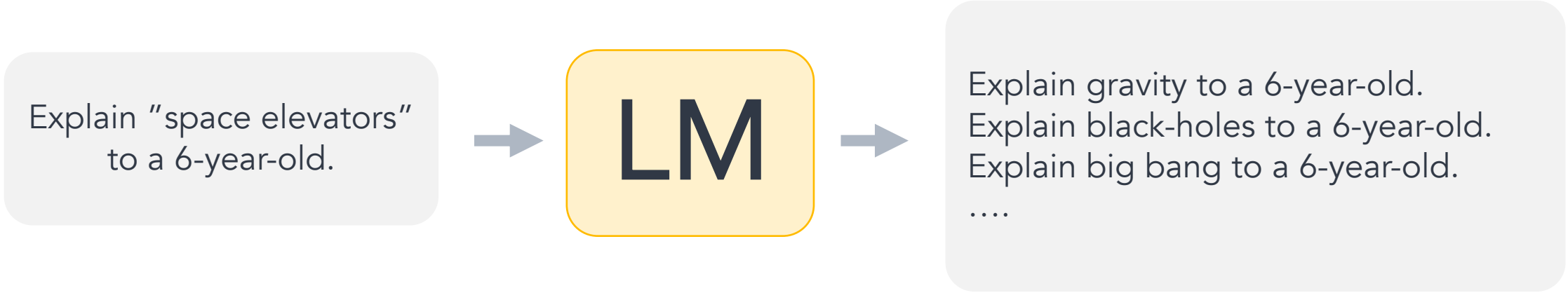
Language Models



[Bengio et al. '04, Peters et al. '18, Raffel et al. '20, Brown et al. '20, many others]

Background

Language Modeling ≠ Following User Intents



LMs are not "aligned" with user intents [Ouyang et al., 2022].

[Training language models to follow instructions with human feedback, Ouyang et al. 2022]

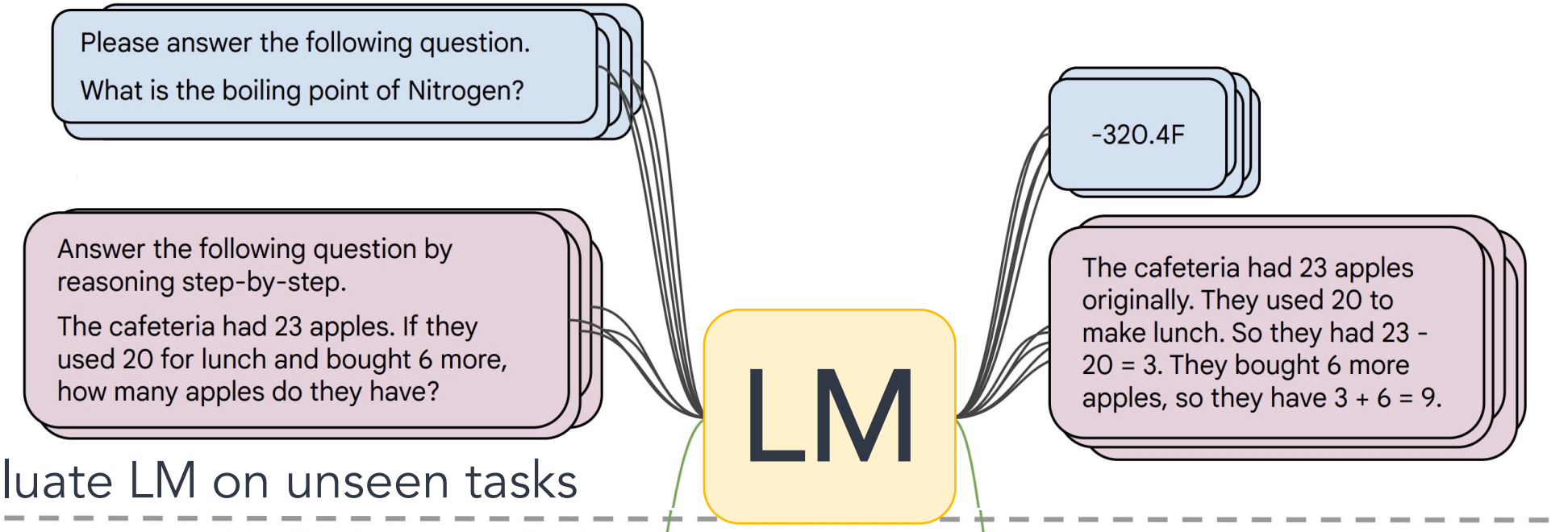
Background

How do we “align” LMs with our articulated intents?

Background

Approach 1: Behavior Cloning (Supervised Learning)

1. Collect examples of (instruction, output) pairs across many tasks and finetune an LM



2. Evaluate LM on unseen tasks

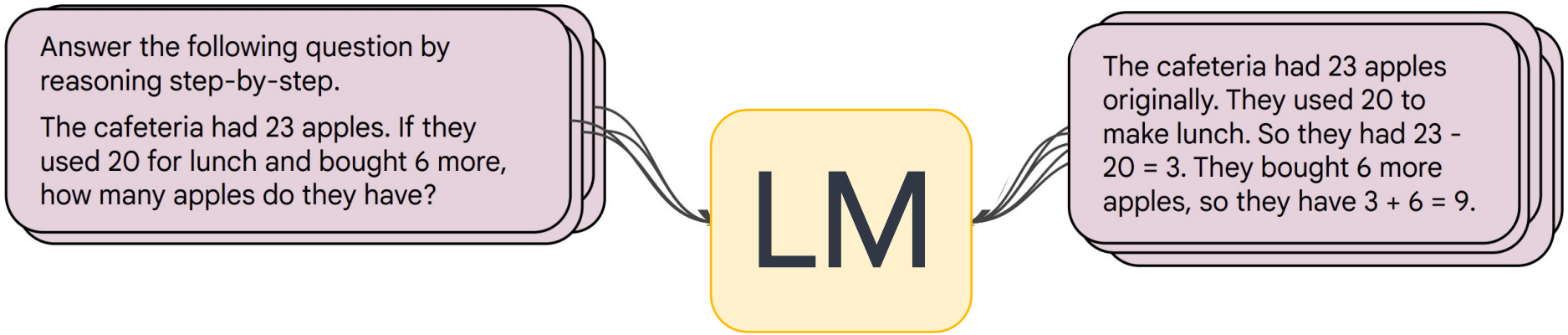
Inference: generalization to unseen tasks

[McCann et al., 2019, Weller et al. 2020. Mishra et al. 2021; Wang et al. 2022, Sanh et al. 2022; Wei et al., 2022, Chung et al. 2022, many others]

Background

Approach 1: Behavior Cloning (Supervised Learning)

- Brittle and sensitive to instruction wordings
- Incentivizes word-by-word rote learning => limits creativity
  - The resulting models' generality/creativity is bounded by that of their supervision data.

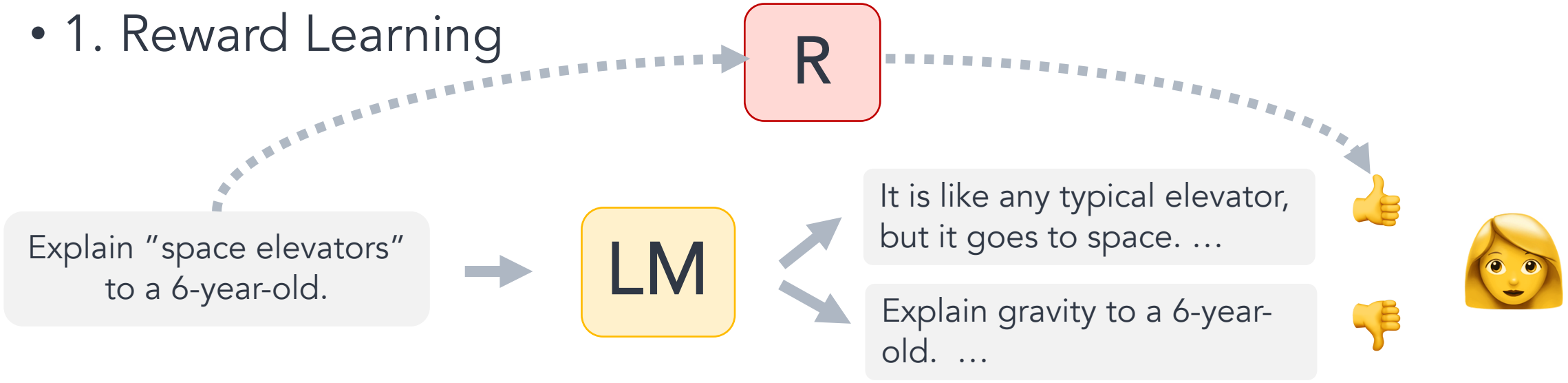


[McCann et al., 2019, Weller et al. 2020. Mishra et al. 2021; Wang et al. 2022, Sanh et al. 2022; Wei et al., 2022, Chung et al. 2022, many others]

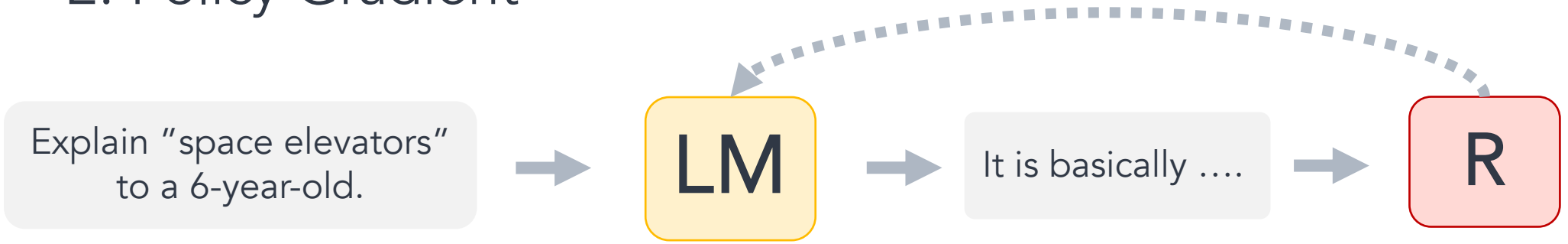
Background

Approach 2: RL w/ Ranking Feedback (RLHF)

- 1. Reward Learning



- 2. Policy Gradient



[Christiano et al. 2017; Stiennon et al. 2020; Ouyang et al., 2022]

Background

The overall recipe 👨‍🍳: Yann's Three-layered cake

Human feedback for aligning LLMs is costly.

Obtaining diverse and quality is quite difficult – not easy to crowdsource.

How far can we reduce the human annotations?

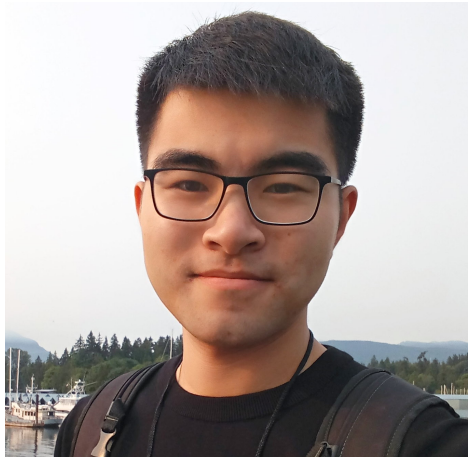
- **Goal:** reduce the role of human annotations.

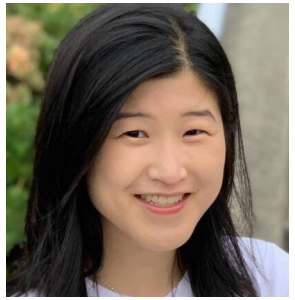
Human feedback for aligning LLMs is costly.
Obtaining diverse and quality is quite difficult – not easy to crowdsource.

Self-Instruct:

Aligning Language Models w/ Self-Generated Instructions

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu,
Noah A. Smith, Daniel Khashabi, Hannaneh Hajishirzi

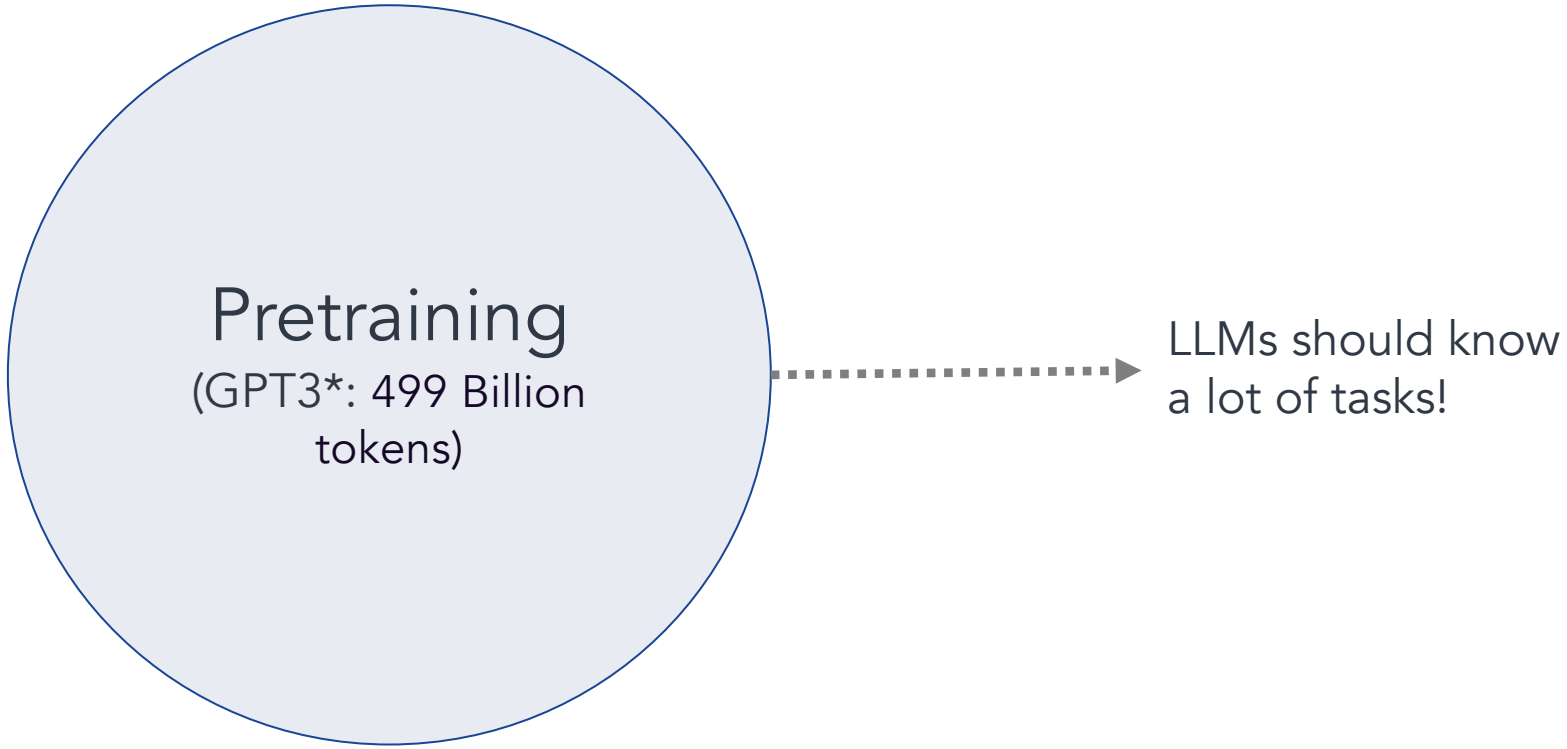



https://arxiv.org/abs/2212.10560

How far can we reduce the human annotations?

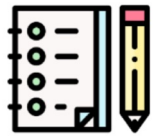
- **Goal:** reduce the role of human annotations.
- **Idea:** we can bootstrap "instruction" from off-the-shelf LMs.
  - LMs have seen humans talk about their needs and goals.



Get humans to write "seed" tasks ✍️

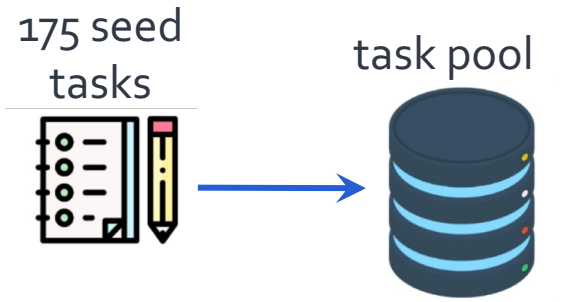
- I am planning a 7-day trip to Seattle. Can you make a detailed plan for me?
- Is there anything I can eat for breakfast that doesn't include eggs, yet includes protein and has roughly 700-100 calories?
- Given a set of numbers find all possible subsets that sum to a given number.
- Give me a phrase that I can use to express I am very happy.

175 seed tasks



Put them your task bank 📦

- I am planning a 7-day trip to Seattle. Can you make a detailed plan for me?
- Is there anything I can eat for breakfast that doesn’t include eggs, yet includes protein and has roughly 700-100 calories?
- Given a set of numbers find all possible subsets that sum to a given number.
- Give me a phrase that I can use to express I am very happy.



Sample and get LLM to expand it

- I am planning a 7-day trip to Seattle. Can you make a detailed plan for me?
- Is there anything I can eat for breakfast that doesn't include eggs, yet includes protein and has roughly 700-100 calories?
- Given a set of numbers find all possible subsets that sum to a given number.
- Give me a phrase that I can use to express I am very happy.

LM Pre-trained, but **not aligned yet**

- Create a list of 10 African countries and their capital city?
- Looking for a job, but it's difficult for me to find one. Can you help me?
- Write a Python program that tells if a given string contains anagrams.

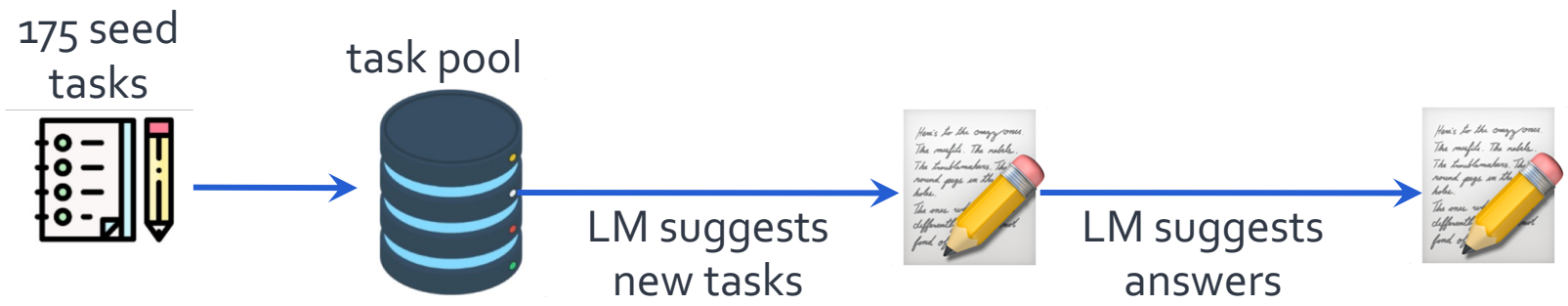
Get LLM to answers the new tasks

```
• Task: Convert the following temperature from Celsius to Fahrenheit.
• Input: 4 °C
• Output: 39.2 °F

• Task: Write a Python program that tells if a given string contains anagrams.
```

LM Pre-trained, but **not aligned yet**

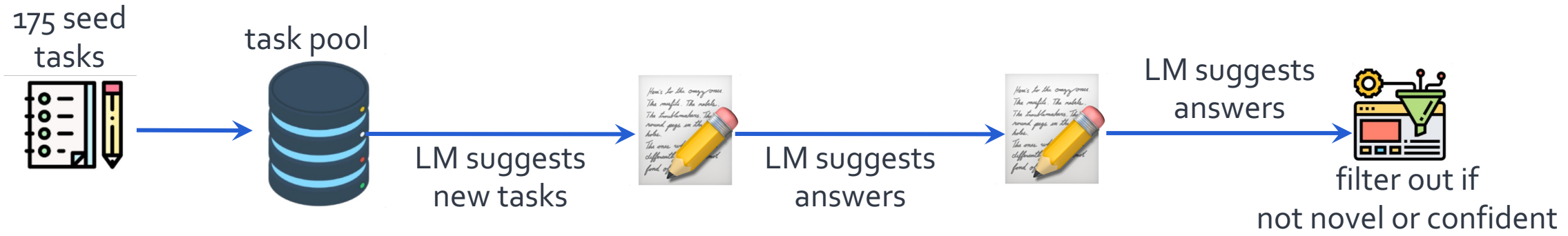
```
• Input: -
• Output:
    def isAnagram(str1, str2): ...
```



Filter tasks

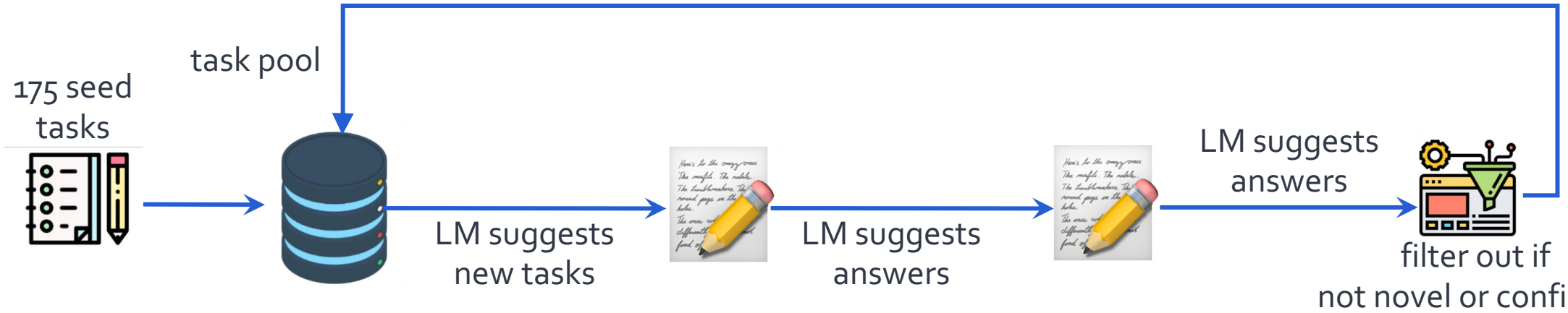
- Drop tasks if LM assigns low probability to them.

- Drop tasks if they have a high overlap with one of the existing tasks in the task pool.
  - Otherwise, common tasks become more common — tyranny of majority.



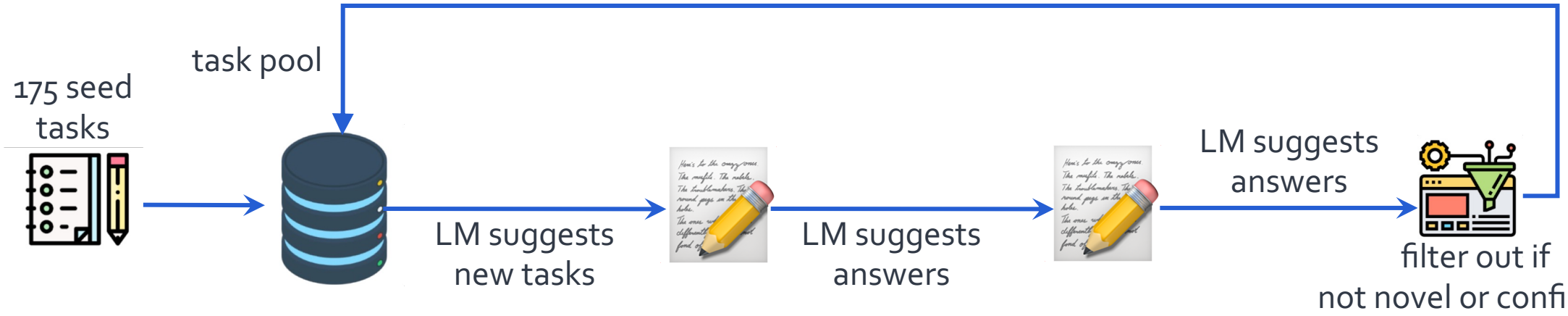
Close the loop

- Add the filtered tasks to the task pool.
- Iterate this process (generate, filter, add) until yield is near zero.

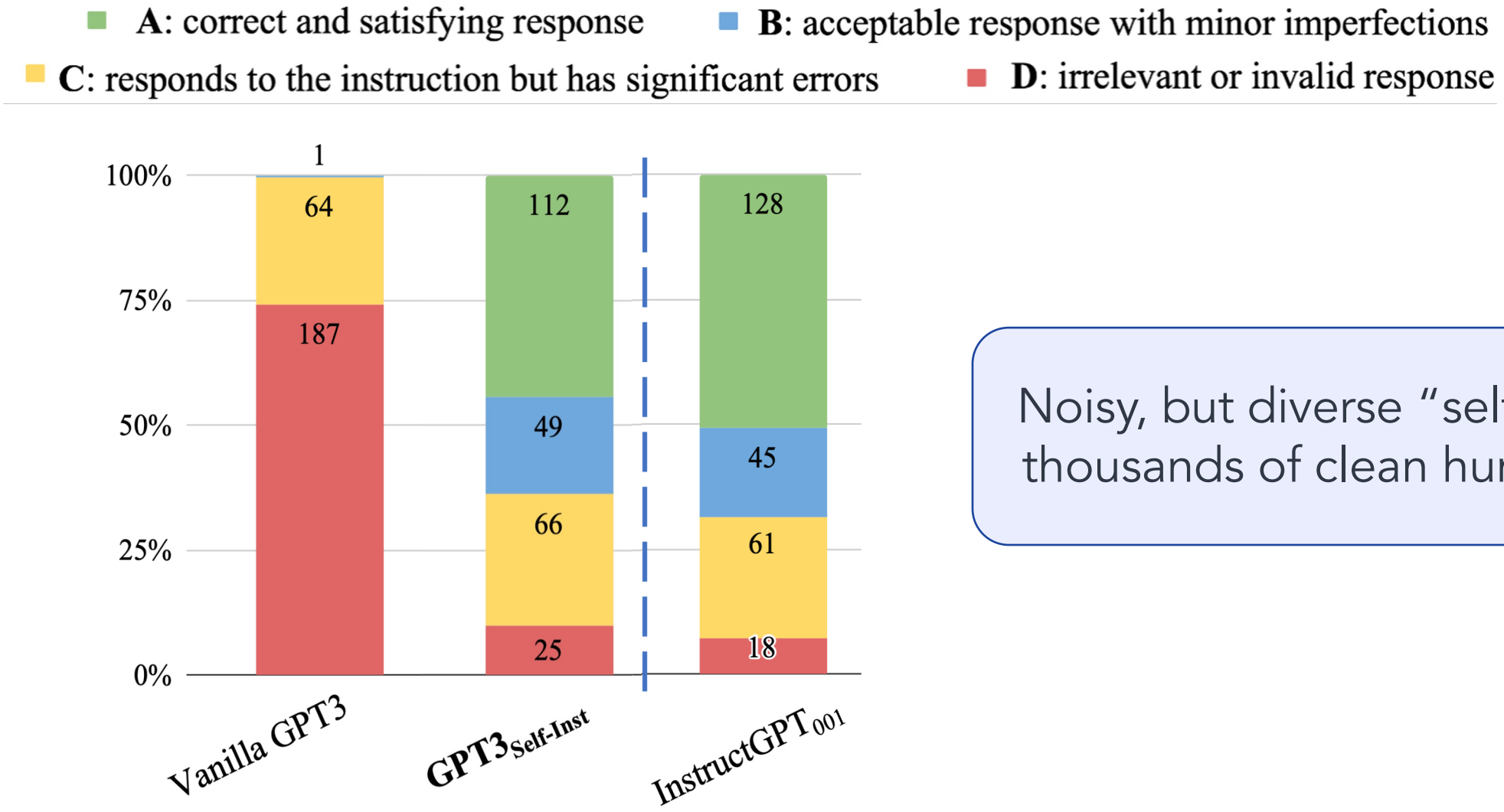


Self-Instructing GPT3 (base version)

- **Generate:**
  - GPT3 ("davinci" engine).
  - We generated 52K instructions and 82K instances.
  - API cost ~$600
- **Align:**
  - We finetuned GPT3 with this data via OpenAI API (2 epochs). **
  - API cost: ~$338 for finetuning



Evaluation on User-Oriented Instructions



Noisy, but diverse "self-instruct" data ~ thousands of clean human-written data

[Self-Instruct: Aligning Language Model with Self-Generated Instructions, Wang et al. 2023]

Summary Thus Far

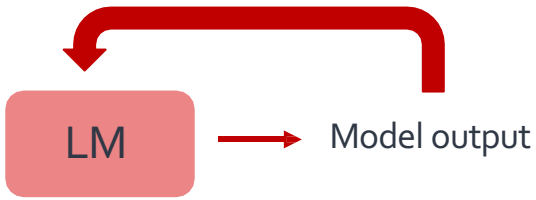
- Self-Instruct: Using LLM itself bootstrap alignment data
- We can reduce the reliance on human annotations in “alignment”.
- LLMs can expand upon examples and diversify the labelled data.

(* See also concurrent work: Unnatural-Instructions [Honovich et al. 2022] and Self-Chat [Xu et al. 2023])

Impact: Learning from AI Feedback

- Open-source models adopted Self-Instruct data generation.
  - Alphaca, Zephyr, etc. [Taori et al. 2023; Tunstall et al. 2023]

- LLMs used directly as a reward during alignment, skipping the data generation. [Lee et al. 2023; many others]

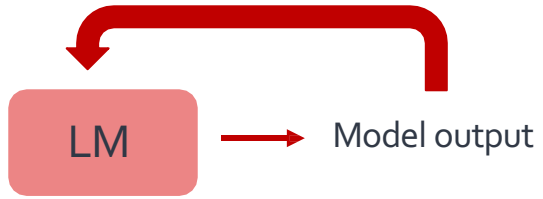


RLAIF: Scaling Reinforcement Learning from Human Feedback with AI Feedback

Harrison Lee, Samrat Phatale, Hassan Mansoor, Thomas Mesnard, Johan Ferret, Kellie Lu, Colton Bishop, Ethan Hall, Victor Carbune, Abhinav Rastogi, Sushant Prakash
Google Research
{harrisonlee,samratph,hassan}@google.com

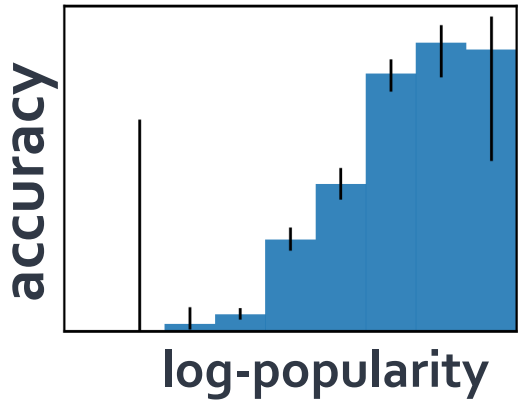
Training LLMs with LLM Feedback: The Bottleneck

- Model feedback is a powerful idea, but …
- It has many limitations …
  - It amplifies existing biases.
  - It is still confined to the [implicit] boundaries defined by the its prompts.
  - LLMs work best in high-data regime. They fail when data is thin.



[Mallen et al. 2022; Razeghi et al. 2022; many others]

- Training with self-feedback is not the way to the moon!



Today

- Do we see any evidence that AI/LLMs self-grow?

Training-time
Self-Feedback

Inference-time
Self-Feedback

Today

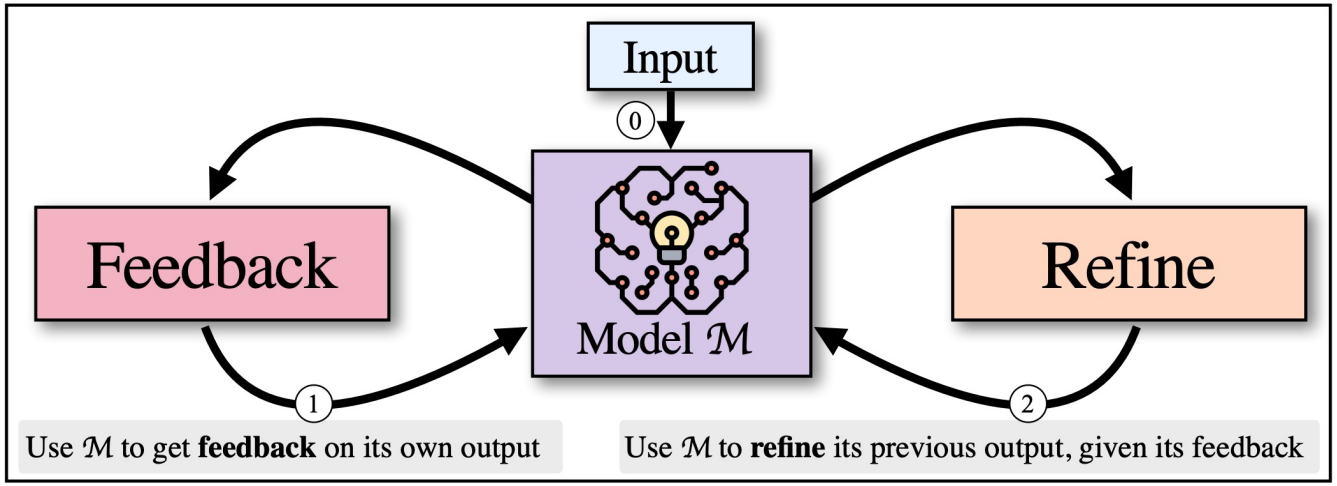
- Do we see any evidence that AI/LLMs self-grow?

Training-time Self-Feedback

Inference-time Self-Feedback

Inference-Time Self-Refinement

- If LLMs prompted appropriated, can they improve their previous generations?



[SELF-REFINE: Iterative Refinement with Self-Feedback, Madaan et al., 2023]
[Reflexion: Language Agents with Verbal Reinforcement Learning, Shinn et al., 2023]

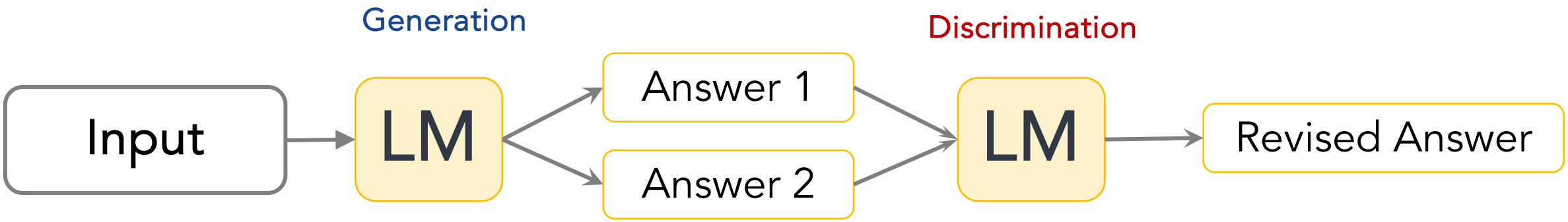
Tangled in Own Thoughts:

LLMs Struggle with Refining Self-Generated Responses

Dongwei Jiang, Jingyu Zhang, Orion Weller, Nathaniel Weir
Benjamin Van Durme, Daniel Khashabi

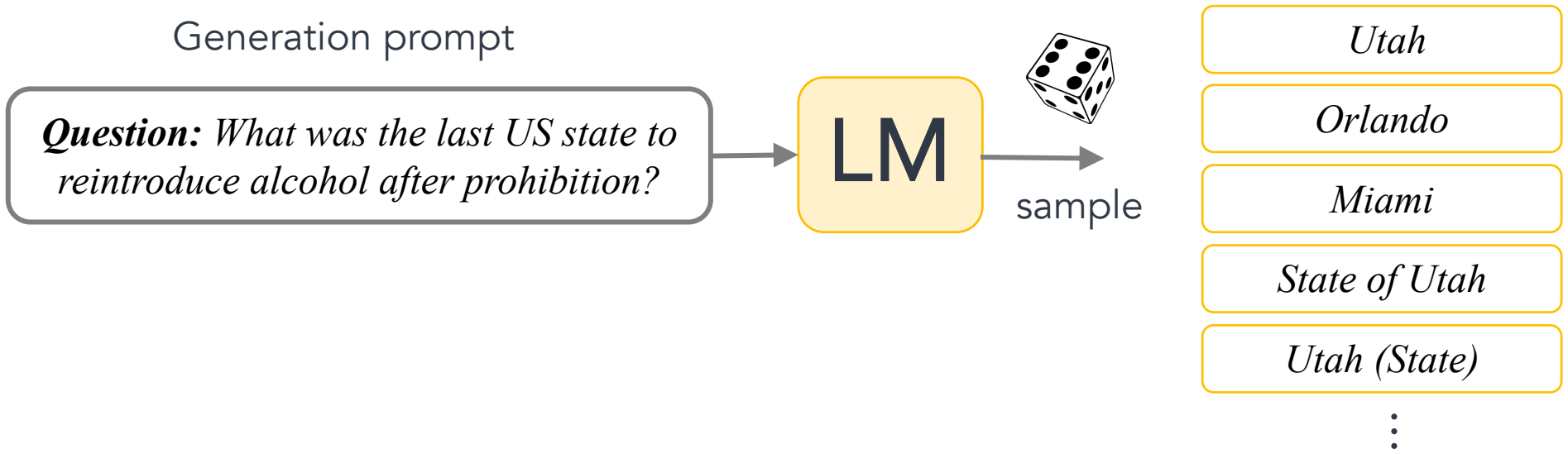
Draft to on arXiv next week!!

Setup and Hypothesis

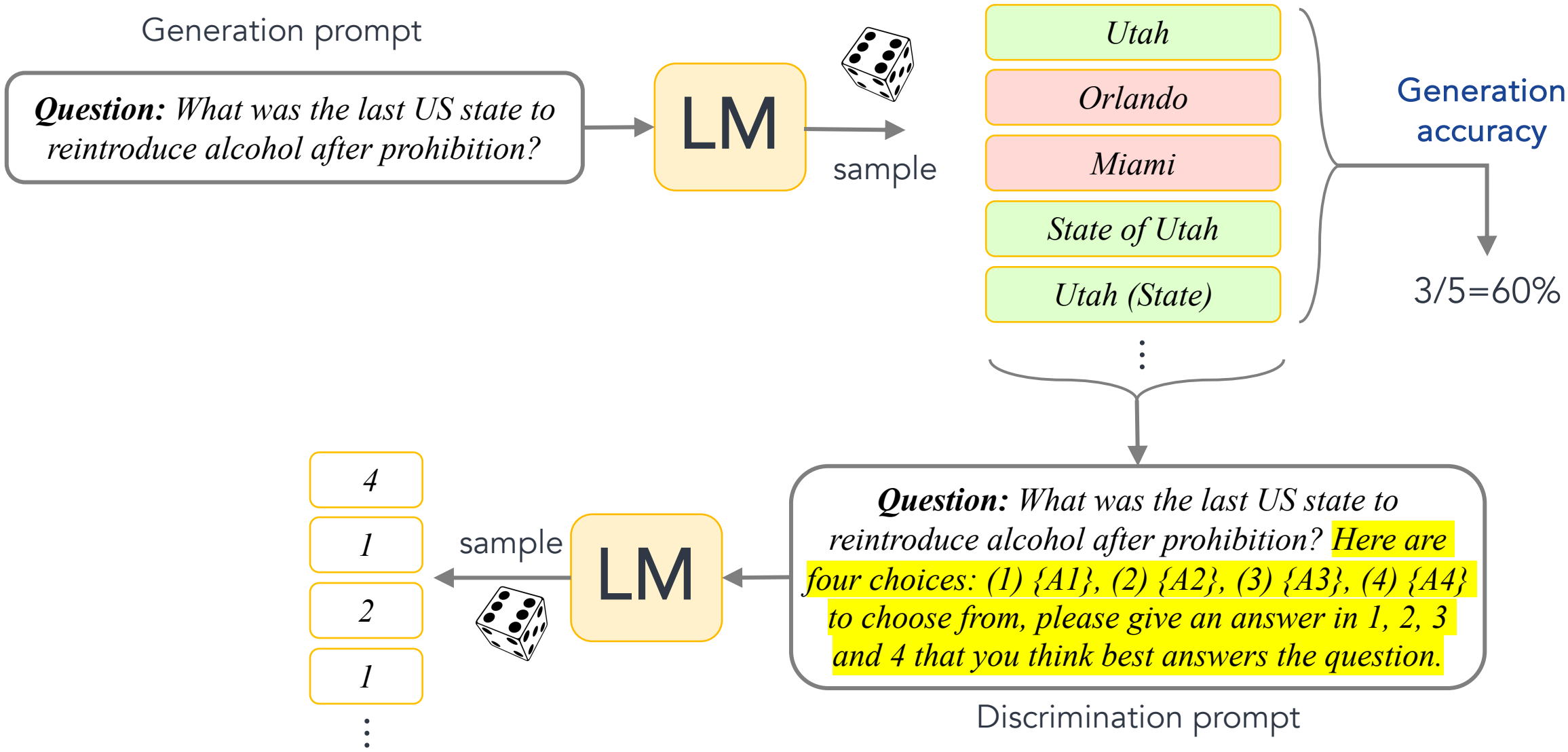


For inference-time refinement, LLMs should be better at discriminating among previously-generated alternatives than generating initial responses.

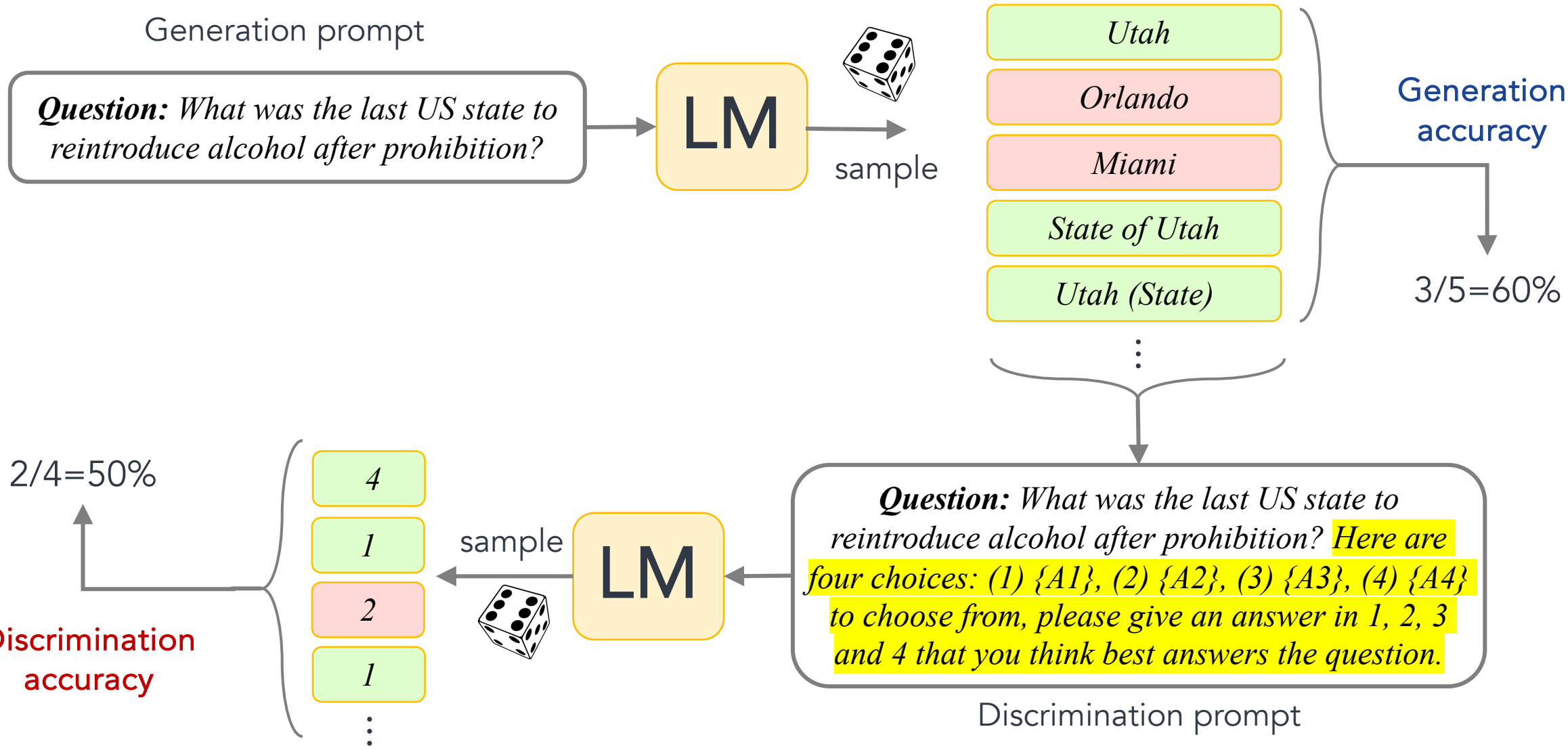
Evaluation Setup



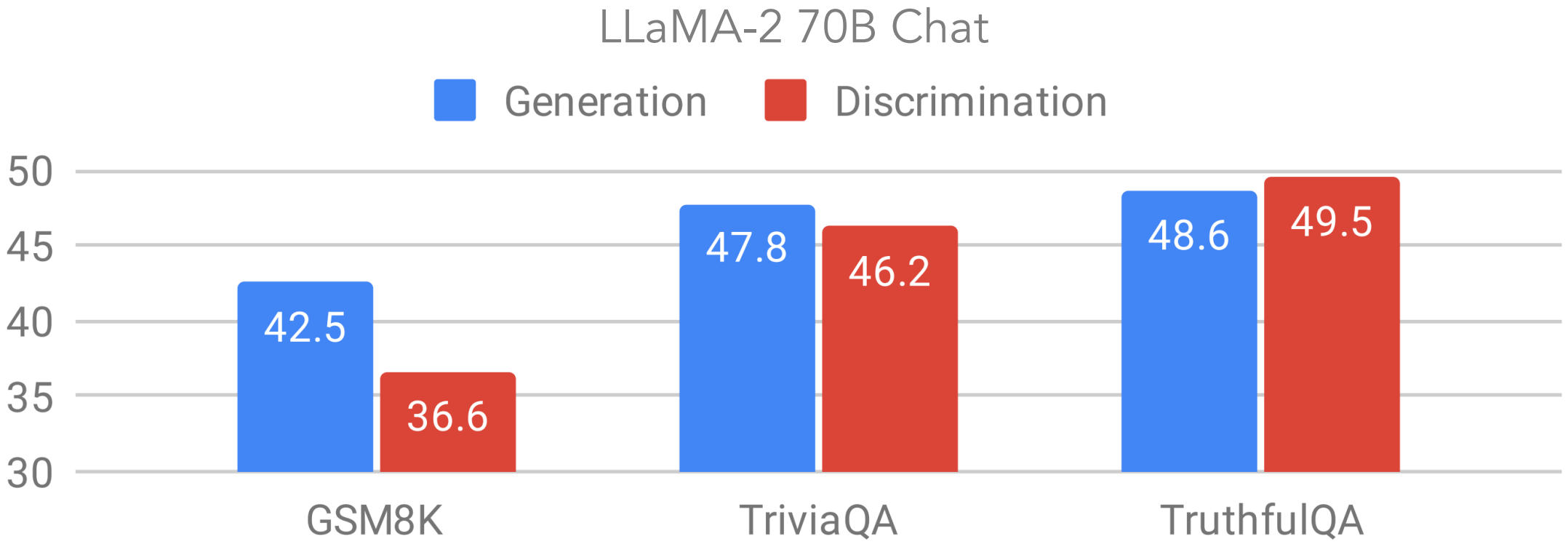
Evaluation Setup



Evaluation Setup

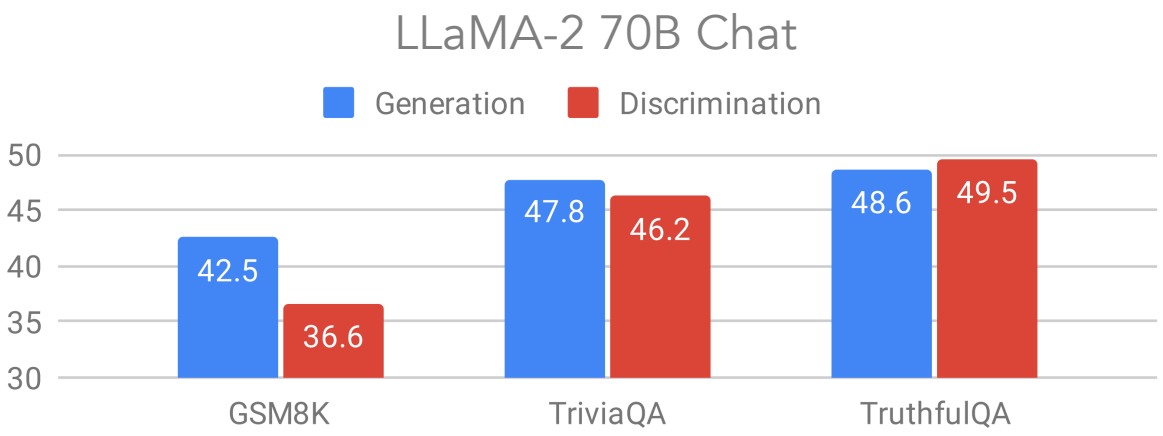


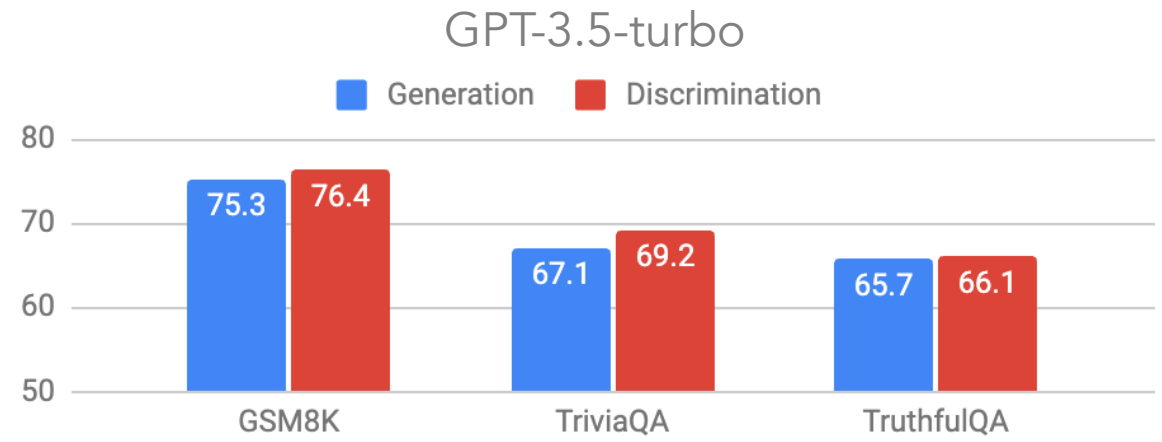
Evaluation Results

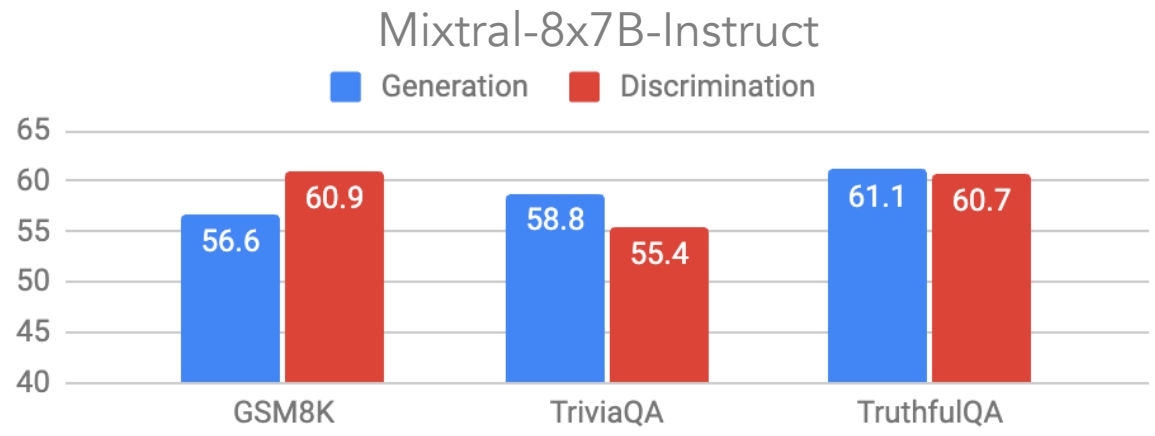


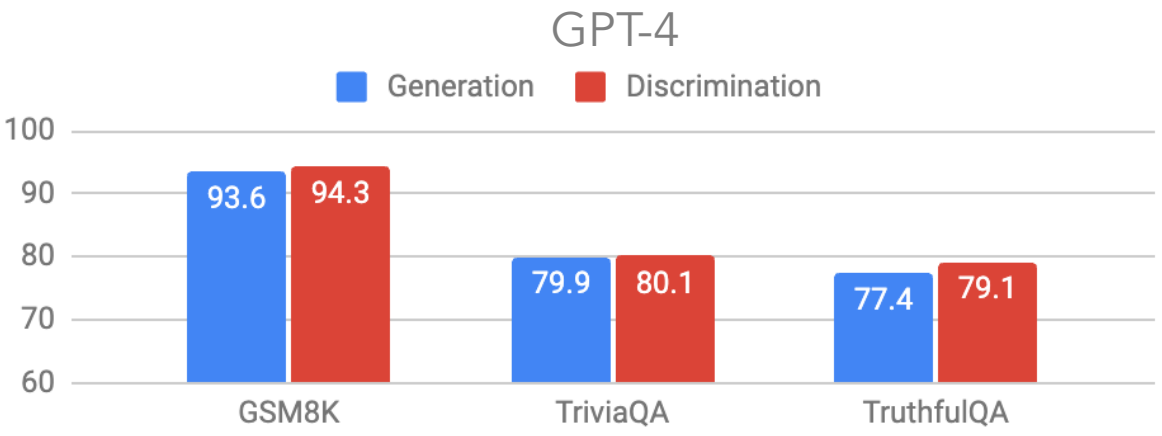
There is no evidence that discriminating among candidates is necessarily an easier task than generating answers.

There is no evidence that discriminating among candidates is necessarily an easier task than generating answers.









Why is "Discrimination" not Easier than "Generation"?

- Sub-hypothesis: Pre-training objective (next-token prediction) benefits generation more.
- Sub-hypothesis: Alignment datasets are skewed toward generative tasks.
- Sub-hypothesis: Length generalization benefits generation more.
- We have partial evidence for all these.

Summary of this work

- We do not see any evidence that inference-time refinement of answers leads to consistent gains.

- Parallel works

ICLR 2024

LARGE LANGUAGE MODELS CANNOT SELF-CORRECT REASONING YET

Jie Huang[1,2*] **Xinyun Chen**[1*] **Swaroop Mishra**[1] **Huaixiu Steven Zheng**[1] **Adams Wei Yu**[1] **Xinying Song**[1] **Denny Zhou**[1]

[1]Google DeepMind [2]University of Illinois at Urbana-Champaign

jeffhj@illinois.edu, {xinyunchen, dennyzhou}@google.com

arXiv 2023

LLMs cannot *find* reasoning errors, but can *correct* them!

Gladys Tyen*[1], Hassan Mansoor[2], Victor Cărbune[2], Peter Chen†[2], Tony Mak†[2]

[1]University of Cambridge, Dept. of Computer Science & Technology, ALTA Institute

[2]Google Research

gladys.tyen@cl.cam.ac.uk

{hassan,chenfeif,tonymak,vcarbune}@google.com

Growing Chatbots Out of Thin-Air

Training-time
Self-Feedback

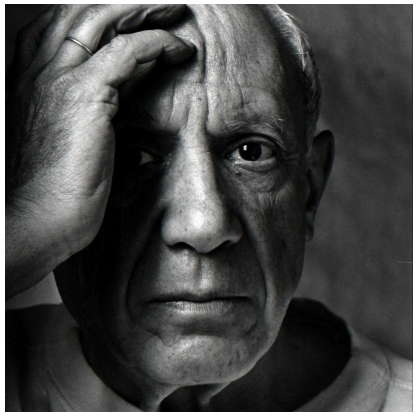
Inference-time
Self-Feedback

Success of AI Depends on "Assumptions"

- It is true that "self-supervised learning" has been an effective force for growing models via unlabeled data.
- Yet, your model will work if it has seen similar-ish problems.
- We always need to make assumptions about tasks, domain, and data (e.g., "prompt-engineering").

"Computers are useless.
They can only give you answers"

-- Pablo Picasso, 1968



Intelligence Continues to be a Moving Target

- Every step forward, we realize there are new challenges ahead.

Thanks!