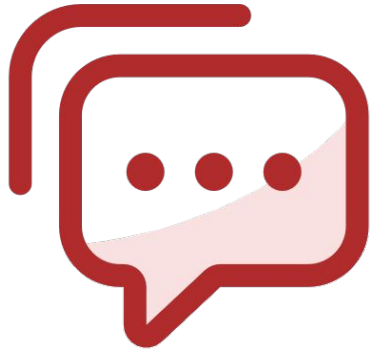


slido

Please download and install the
Slido app on all computers you use



Audience Q&A

① Start presenting to display the audience questions on this slide.

Training with Synthetic Data

Large Language Models: Methods and Applications

Daphne Ippolito and Chenyan Xiong

Learning Objectives

Learn the basics of training with synthetic data

Learn various ways to curate synthetic data

Learn recent advancements on synthetic data generation

Understand the benefits and limitations of training with synthetic data

Outline

Why Synthetic Data

Synthetizing Labels/Rewards

Synthetizing Training Data

What is Good Synthetic Data?

Recent Methodologies

Limitations

slido

Please download and install the
Slido app on all computers you use



Why people use **synthetic** training data?

① Start presenting to display the poll results on this slide.

Why Synthetic Data?

- A cheap and fast way to obtain supervised labels
- A typical manual label, on simple tasks, costs nearly \$1, and lots of time and effort

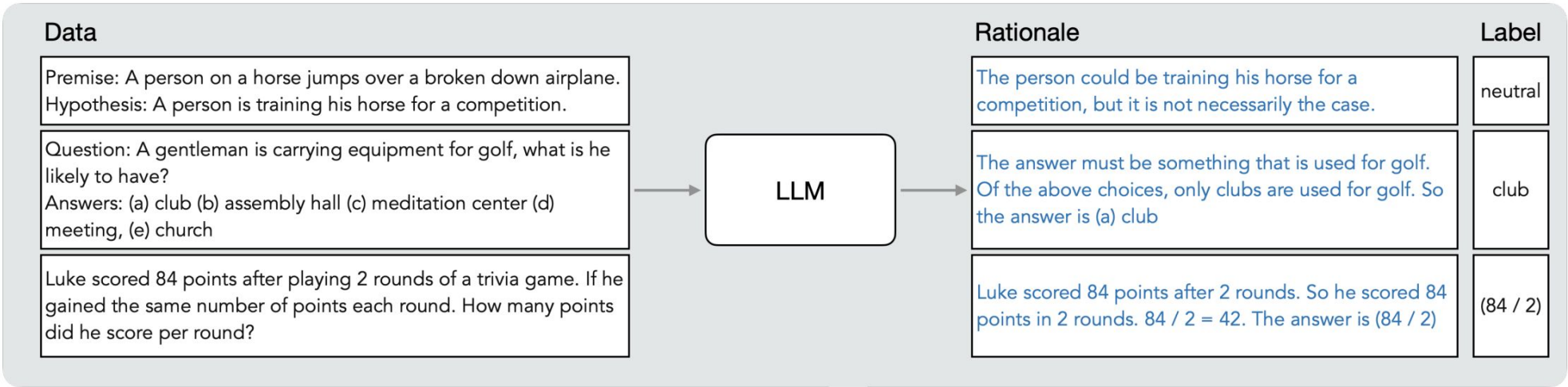
Query: When did rock'n'roll begin?		Grading rubric questions:					Rubric relevance label:
Passages:		Which musicians or bands are considered pioneers of rock n roll?	What were the major influences that led to the emergence of rock n roll?	Are there any specific events or performances that marked the beginning of rock n roll?	Did the development of new technologies have an impact on the birth of rock n roll?	Is there a general consensus among music historians regarding the exact start of rock n roll?	
		r1	r2	r3	r4	r5	
p1	, listens to pop, rock, soul and whatnot. The rock and roll era began around 1950 . It evolved from rhythm and blues in the 1940s. The name rock and roll was the title of a song by the Boswell Sisters in 1934 (YouTube), but that was a swing song.	4 Boswell Sisters	4 rhythm and blues	0	0	4 The rock and roll era began around 1950	4
p2	Definition of 'rock and roll'. rock and roll also rock'n'roll. uncountable noun. Rock and roll is a kind of popular music developed in the 1950s which has a strong beat and is played on electrical instruments Elvis Presley—the King of Rock and Roll the greatest rock 'n' roll band in the world. rock and roll or rock'n'roll.	5 Elvis Presley —the King of Rock and Roll	0	0	4 Rock and roll is played on electrical instruments	4 1950s	5
p3	But you could just as well take the release of " Rocket 88 " in 1951 as the beginning of the rocknroll era, because that was the first rocknroll song. So far we have the years 1951 and 1952 to choose from. Maybe a certain date will turn up while we examine why rocknroll emerged at all.	0	0	4 "Rocket 88"	0	4 1951 and 1952	4

Ask LLMs to Label Search Relevance [1]

[1] Farzi et al. 2024. Pencils Down! Automatic Rubric-based Evaluation of Retrieve/Generate Systems

Why Synthetic Data?

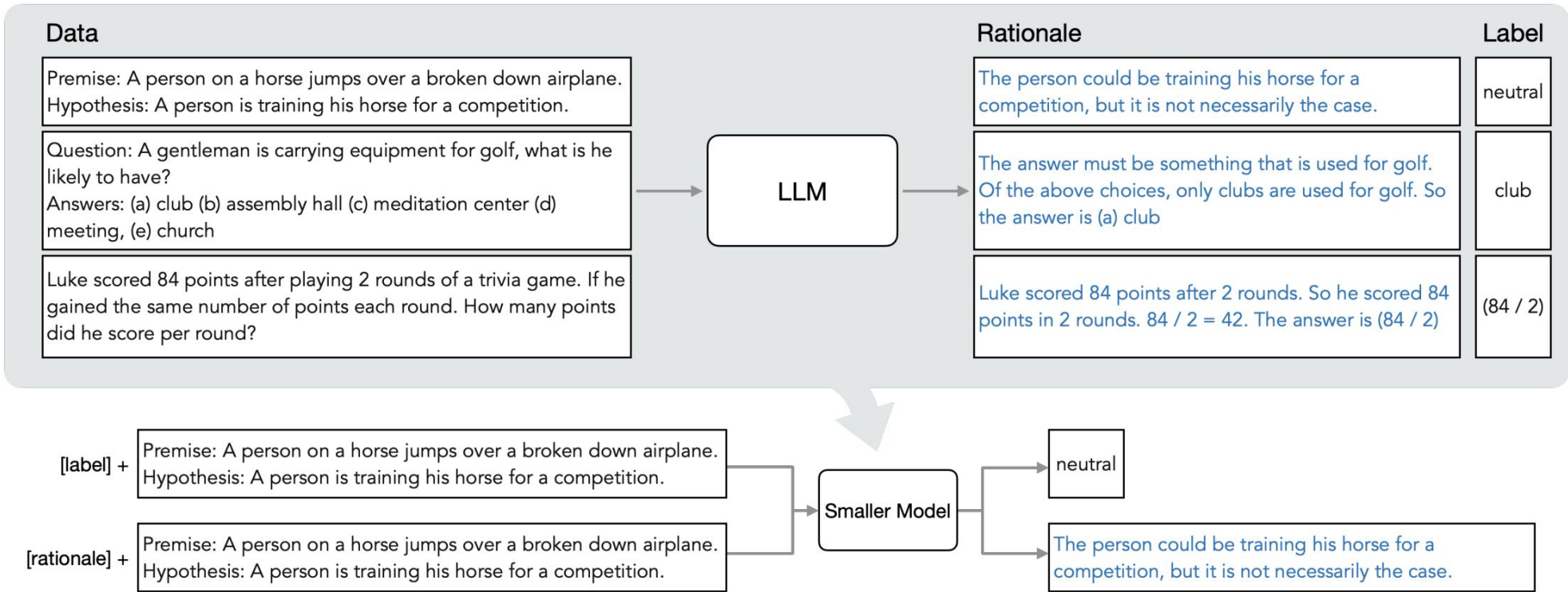
- To distill from large language models
- Closed LLMs are trained with lots of extra signals and \$\$\$



Distillation Step-by-Step [2]

Why Synthetic Data?

- To distill from large language models
- Closed LLMs are trained with lots of extra signals and \$\$\$

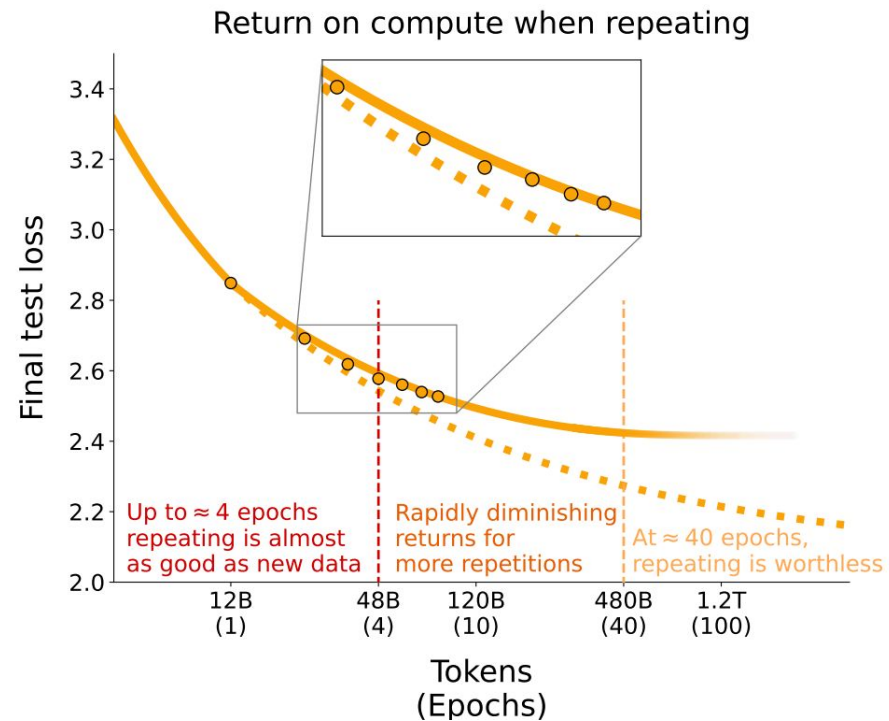


Distillation Step-by-Step [2]

Why Synthetic Data?

Create extra pretraining data

- For data limited scenarios (tail language, special domain, etc.)

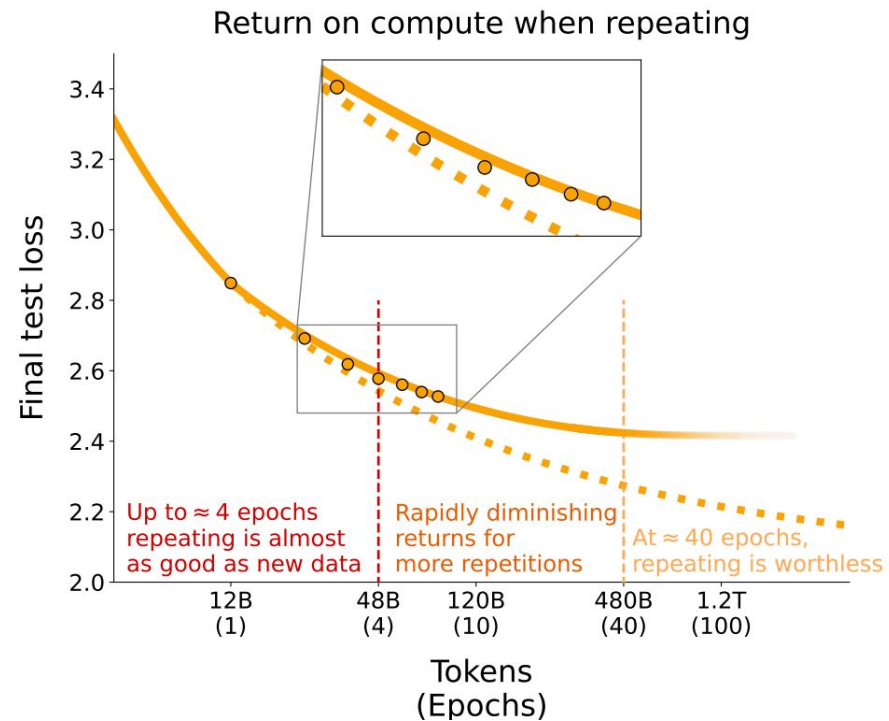


Pretraining Loss with Repeating Data (Epochs) [3]

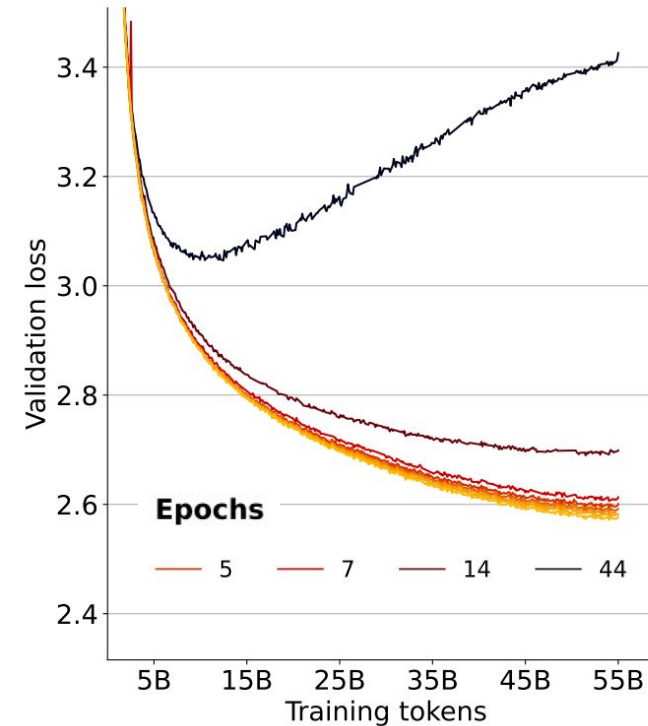
Why Synthetic Data?

Create extra pretraining data

- For data limited scenarios (tail language, special domain, etc.)



Pretraining Loss with Repeating Data (Epochs) [3]



Overfitting with Too Many Repetitions [3]

Why Synthetic Data?

Create extra pretraining data

- To overcome the upper bound of available web data [4]

Hitting a Data Wall

One reason for the GPT slowdown is a dwindling supply of high-quality text and other data that LLMs can process during pretraining to make sense of the world and the relationships between different concepts so they can solve problems such as drafting blog posts or solving coding bugs, OpenAI employees and researchers said.

In the past few years, LLMs used publicly available text and other data from websites, books and other sources for the pretraining process, but developers of the models have largely squeezed as much out of that type of data as they can, these people said.

Why Synthetic Data?

Create extra pretraining data

- To overcome the upper bound of available web data [4]

Hitting a Data Wall

One reason for the GPT slowdown is a dwindling supply of high-quality text and other data that LLMs can process during pretraining to make sense of the world and the relationships between different concepts so they can solve problems such as drafting blog posts or solving coding bugs, OpenAI employees and researchers said.

In the past few years, LLMs used publicly available text and other data from websites, books and other sources for the pretraining process, but developers of the models have largely squeezed as much out of that type of data as they can, these people said.

Why Synthetic Data?

Create extra pretraining data

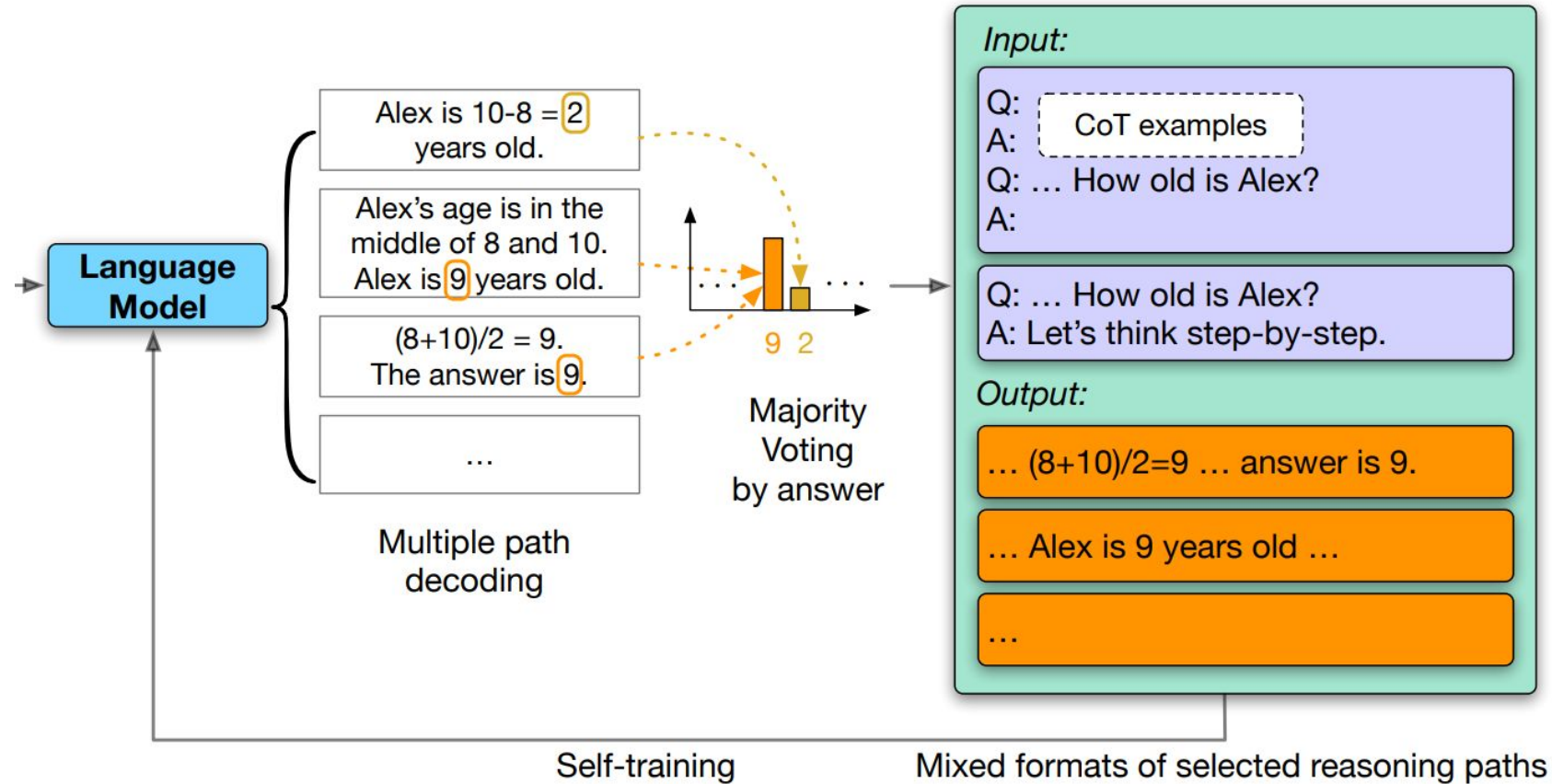
- To overcome the upper bound of available web data [4]

In response, OpenAI has created a foundations team, led by Nick Ryder, who previously ran pretraining, to figure out how to deal with the dearth of training data and how long the scaling law will continue to apply, they said.

Orion was trained in part on AI-generated data, produced by other OpenAI models, including GPT-4 and recently released reasoning models, according to an OpenAI employee. However, such synthetic data, as it is known, is leading to a new problem in which Orion may end up resembling those older models in certain aspects, the employee said.

Why Synthetic Data?

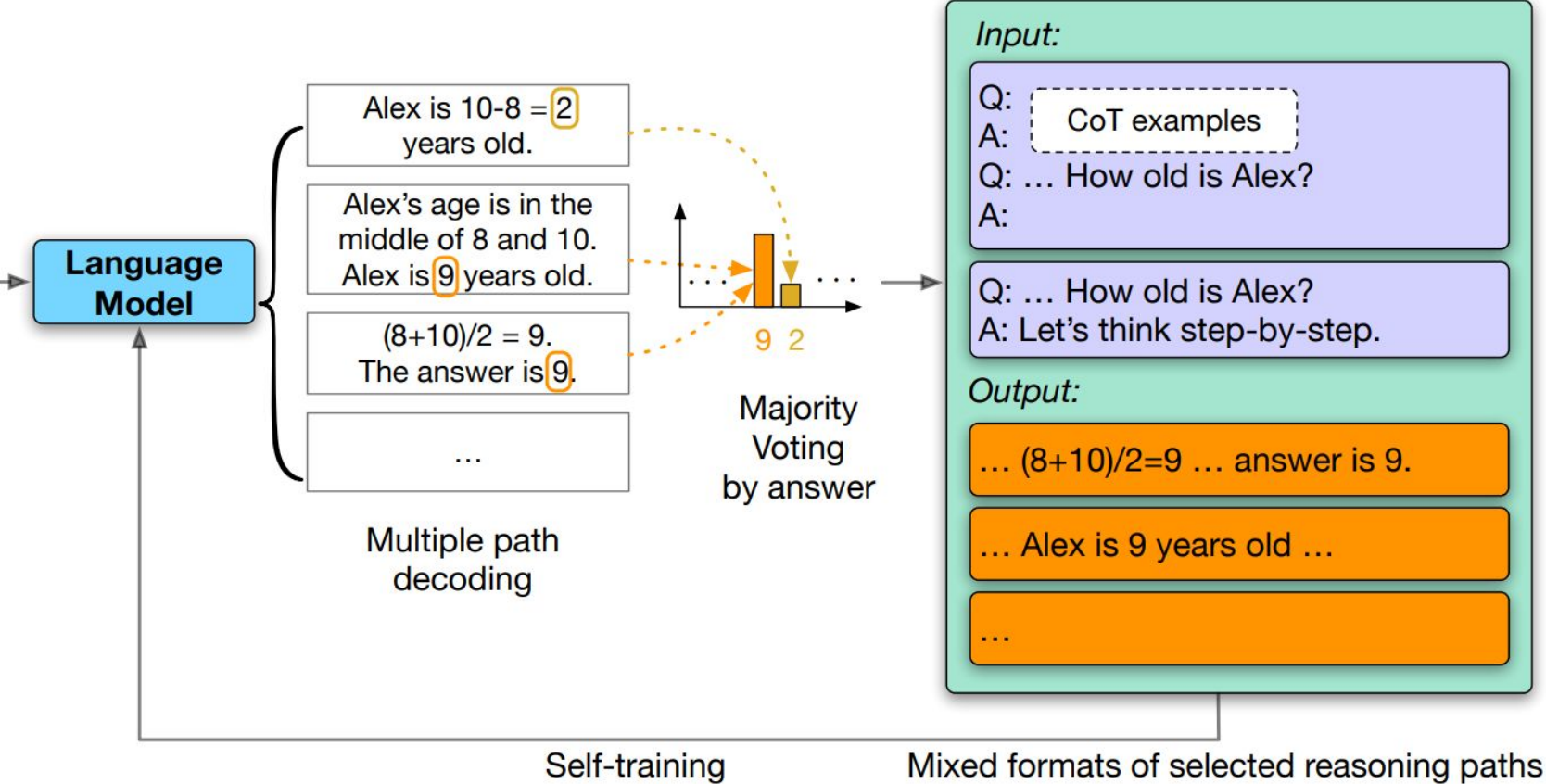
Self-Play Towards AGI



Training with LLM's Own Generated Data [4]

Why Synthetic Data?

Self-Play Towards AGI



Training with LLM's Own Generated Data [4]



Why Synthetic Data?

Self-Play Towards AGI

- Alpha-Go is trained heavily by self-play



Outline

Why Synthetic Data

Synthetizing Labels/Rewards

Synthetizing Training Data

What is Good Synthetic Data?

Recent Methodologies

Limitations

Synthetizing Labels/Rewards

LLM-as-a-Judge:

- Prompt LLMs to label data
- Often pairwise labeling

Question: If the FED buys bonds in the secondary market (A) the money supply will increase. (B) the money supply will decrease. (C) the money supply will not be affected. (D) the discount rate would be affected. Output the answer only.

Assistant A:

The Federal Reserve buys bonds in the secondary market to increase the money supply.

Assistant B:

(A) the money supply will increase.

Follow-up Question: How does it affect my daily life? Give 3 examples.

Assistant A:

The Federal Reserve buys bonds in the secondary market to increase the money supply. This affects your daily life by:

1. Increasing the money supply.
2. Increasing the money supply and lowering the discount rate.
3. Increasing the money supply and lowering the discount rate and increasing the money supply.

Assistant B:

When the Federal Reserve buys bonds in the secondary market, it can have several effects on the economy that can impact your daily life. Here are three examples:

1. Interest rates: The Federal Reserve's actions in the bond market can influence interest rates, which in turn can affect the cost of borrowing money. For example, if the Federal Reserve buys bonds and drives up demand for them, interest rates on things like mortgages, car loans, and credit cards may go up, making it more expensive to borrow money.
2. Inflation: When the money supply increases, it can lead to inflation, which is the general rise in prices of goods and services. This can impact your daily life by making it more expensive to buy things like groceries, gas, and other necessities.
3. Employment: The Federal Reserve's actions in the bond market can also have an impact on employment. For example, if the Federal Reserve's actions lead to economic growth and increased demand for goods and services, it can create more job opportunities and boost employment.

GPT-4 Judgment:

Assistant A provided an incorrect response to the user's question about how the Federal Reserve buying bonds in the secondary market affects daily life. The answer given is repetitive and lacks clear examples of how the action impacts daily life.

On the other hand, **Assistant B** provided a relevant and accurate response to the user's question about the Federal Reserve buying bonds. The answer includes three clear examples of how the action impacts daily life, such as interest rates, inflation, and employment.

Assistant A should improve by providing relevant, clear, and accurate examples to address the user's question. **Assistant B's** response is more helpful, relevant, and detailed.

Synthetizing Labels/Rewards

LLM-as-a-Judge:

- Prompt LLMs to label data
- Often pairwise labeling

Data to Label:
Compare two data
points, easier to label

Prompt GPT-4 to judge
which one is better,
with rationales for
consistency

Question: If the FED buys bonds in the secondary market (A) the money supply will increase. (B) the money supply will decrease. (C) the money supply will not be affected. (D) the discount rate would be affected. Output the answer only.

Assistant A:

The Federal Reserve buys bonds in the secondary market to increase the money supply.

Assistant B:

(A) the money supply will increase.

Follow-up Question: How does it affect my daily life? Give 3 examples.

Assistant A:

The Federal Reserve buys bonds in the secondary market to increase the money supply. This affects your daily life by:

1. Increasing the money supply.
2. Increasing the money supply and lowering the discount rate.
3. Increasing the money supply and lowering the discount rate and increasing the money supply.

Assistant B:

When the Federal Reserve buys bonds in the secondary market, it can have several effects on the economy that can impact your daily life. Here are three examples:

1. Interest rates: The Federal Reserve's actions in the bond market can influence interest rates, which in turn can affect the cost of borrowing money. For example, if the Federal Reserve buys bonds and drives up demand for them, interest rates on things like mortgages, car loans, and credit cards may go up, making it more expensive to borrow money.
2. Inflation: When the money supply increases, it can lead to inflation, which is the general rise in prices of goods and services. This can impact your daily life by making it more expensive to buy things like groceries, gas, and other necessities.
3. Employment: The Federal Reserve's actions in the bond market can also have an impact on employment. For example, if the Federal Reserve's actions lead to economic growth and increased demand for goods and services, it can create more job opportunities and boost employment.

GPT-4 Judgment:

Assistant A provided an incorrect response to the user's question about how the Federal Reserve buying bonds in the secondary market affects daily life. **The answer given is repetitive and lacks clear examples of how the action impacts daily life.**

On the other hand, **Assistant B** provided a relevant and accurate response to the user's question about the Federal Reserve buying bonds. The answer includes three clear examples of how the action impacts daily life, such as interest rates, inflation, and employment.

Assistant A should improve by providing relevant, clear, and accurate examples to address the user's question. **Assistant B's** response is more helpful, relevant, and detailed.

Synthesizing Labels/Rewards: How does it work?

Human preferences collected from Chatbot Arena

Chatbot Arena

An Open Platform for Human Preference Evals

Copilot Arena



Copilot Arena's Initial Leaderboard, Insights, a...

11 min read · 📅 2024

Chatbot Arena Categories



Definitions, Methods, and Insights

24 min read · 📅 2024

Preference Proxy Evaluations



A New Benchmark for Evaluating Reward Models and...

17 min read · 📅 2024

Chatbot Arena



Benchmarking LLMs in the Wild with Elo Ratings

11 min read · 📅 2023

Synthetizing Labels/Rewards: How does it work?

Human preferences collected from Chatbot Arena

For research purposes, we log user prompts and images, and may release this data to the public in the future. Please do not upload any confidential or personal information.

Choose two models to compare

Expand to see the descriptions of 72 models

chatgpt-4o-latest-20241120

Model A

can you write this homework for me

Of course, I'd be happy to help! Could you provide more details about the homework assignment? Let me know the topic, subject, or any specific instructions you have, and I'll help you craft a well-written response.

ministral-8b-2410

Model B

can you write this homework for me

Of course, I'd be happy to help you with your homework! Please provide the details of the task, and I'll assist you as much as possible.

A is better

B is better

Tie

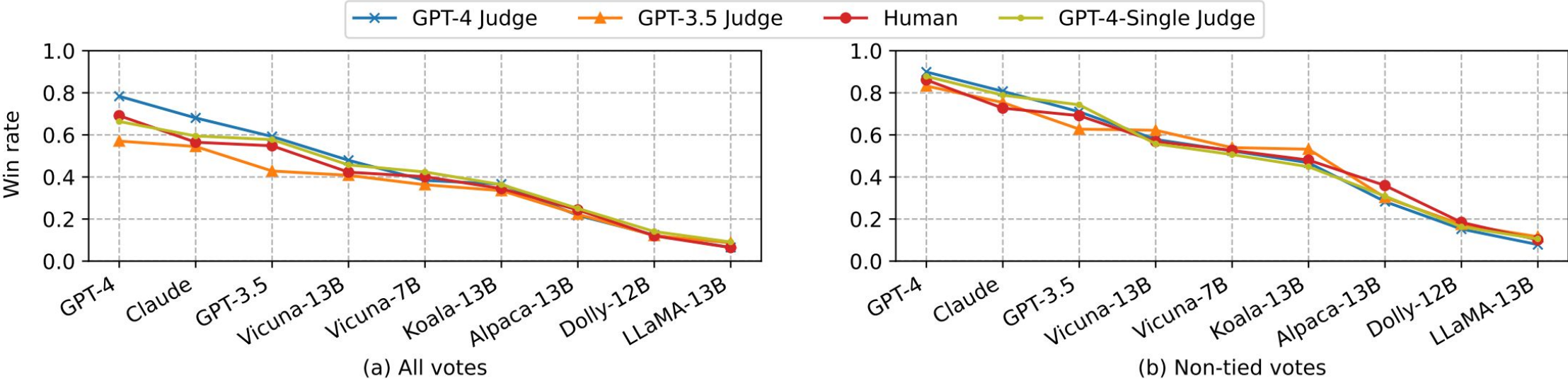
Both are bad

Enter your prompt and press ENTER

Send

Synthesizing Labels/Rewards: How does it work?

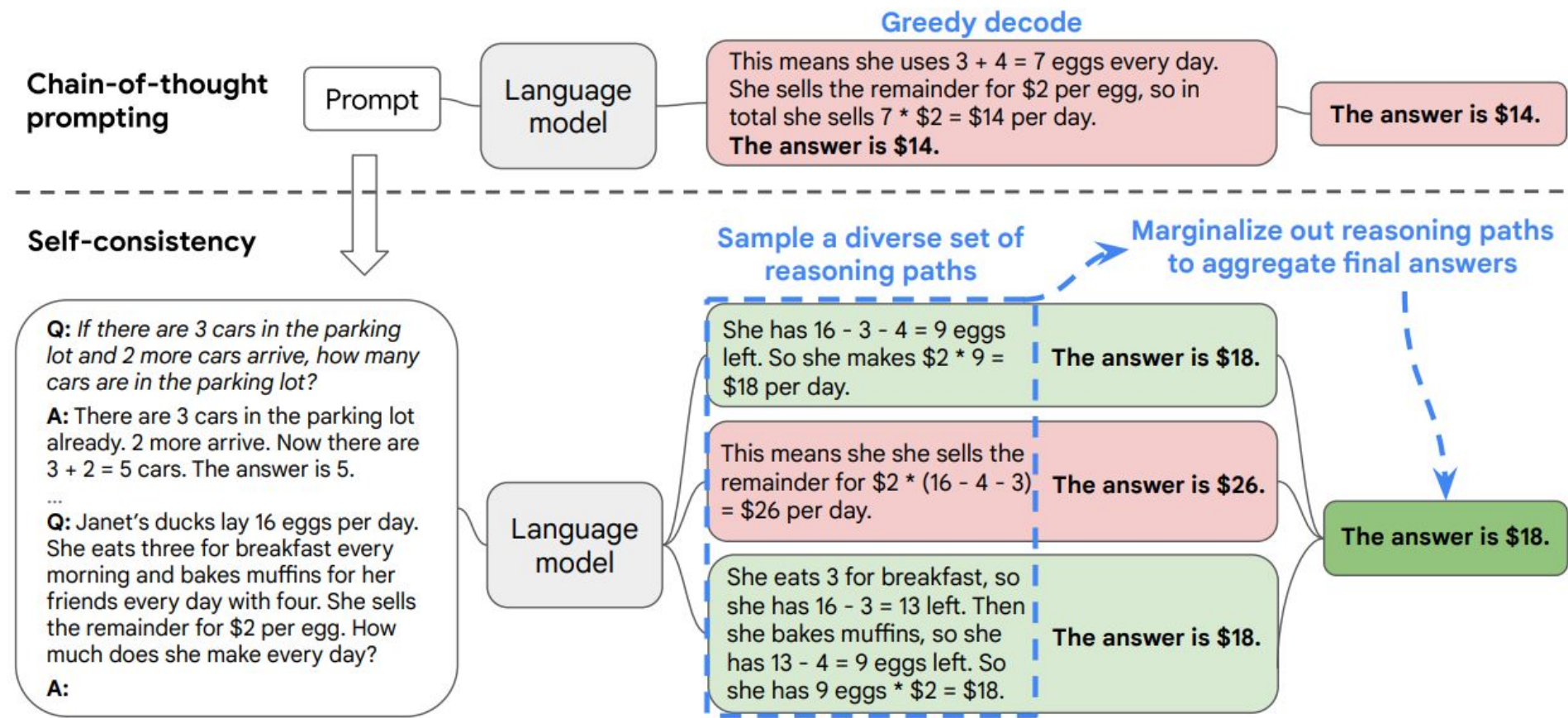
Correlation of LLM labels and human labels



Average Win Rates of Models from Different Judges [6]

Synthetizing Labels/Rewards: Further Improvements

Self-Consistency: Sampling multiple answers and aggregating to the top one



Synthesizing Labels/Rewards: Further Improvements

Self-Consistency: Sampling multiple answers and aggregating to the top one

	GSM8K	MultiArith	AQuA	SVAMP	CSQA	ARC-c
Greedy decode	56.5	94.7	35.8	79.0	79.0	85.2
Weighted avg (unnormalized)	56.3 \pm 0.0	90.5 \pm 0.0	35.8 \pm 0.0	73.0 \pm 0.0	74.8 \pm 0.0	82.3 \pm 0.0
Weighted avg (normalized)	22.1 \pm 0.0	59.7 \pm 0.0	15.7 \pm 0.0	40.5 \pm 0.0	52.1 \pm 0.0	51.7 \pm 0.0
Weighted sum (unnormalized)	59.9 \pm 0.0	92.2 \pm 0.0	38.2 \pm 0.0	76.2 \pm 0.0	76.2 \pm 0.0	83.5 \pm 0.0
Weighted sum (normalized)	74.1 \pm 0.0	99.3 \pm 0.0	48.0 \pm 0.0	86.8 \pm 0.0	80.7 \pm 0.0	88.7 \pm 0.0
Unweighted sum (majority vote)	74.4 \pm 0.1	99.3 \pm 0.0	48.3 \pm 0.5	86.6 \pm 0.1	80.7 \pm 0.1	88.7 \pm 0.1

Performance of PaLM-540B with Different Aggregation [8]

Synthesizing Labels/Rewards: Further Improvements

Generative Verifiers: Verify/label step-by-step [9]

Problem: Tim decides to cancel his cable subscription and get streaming services. He gets Netflix for \$10 a month. Hulu and Disney Plus normally cost \$10 a month each but he saves 20% for bundling. How much money does he save by cancelling his \$60 cable package?

Solution: Tim pays \$60 for cable. He gets Netflix for 10 and **the bundle of Hulu and Disney Plus costs $\$10 * 80\% = \8 .** So he pays $\$10 + \$8 = \$18$ for the bundle. Now he saves $\$60 - \$18 = \$42$. The answer is 42.

Discriminative RM correctness score: **0.999**

GenRM-CoT: Let's verify step by step. ...

Step 3. ****The bundle of Hulu and Disney Plus costs $\$10 * 80\% = \8 .** (Incorrect)**

- Expected Answer: **The bundle of Hulu and Disney Plus costs $\$10 + \$10 = \$20$. With the 20% discount, the total cost is $\$20 * 0.8 = \16**

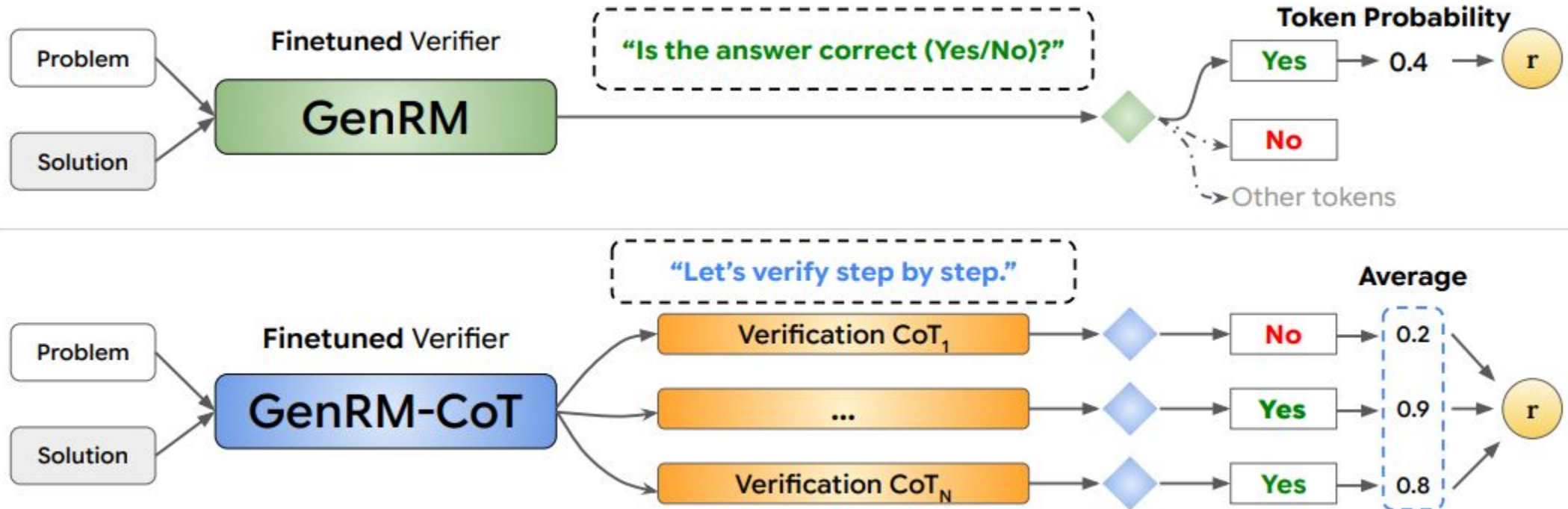
Verification: **Is the answer correct (Yes/No)? No**

GenRM-CoT (Majority Voting) score: **0.0015**

Bring in CoT into LLM reward models

Synthesizing Labels/Rewards: Further Improvements

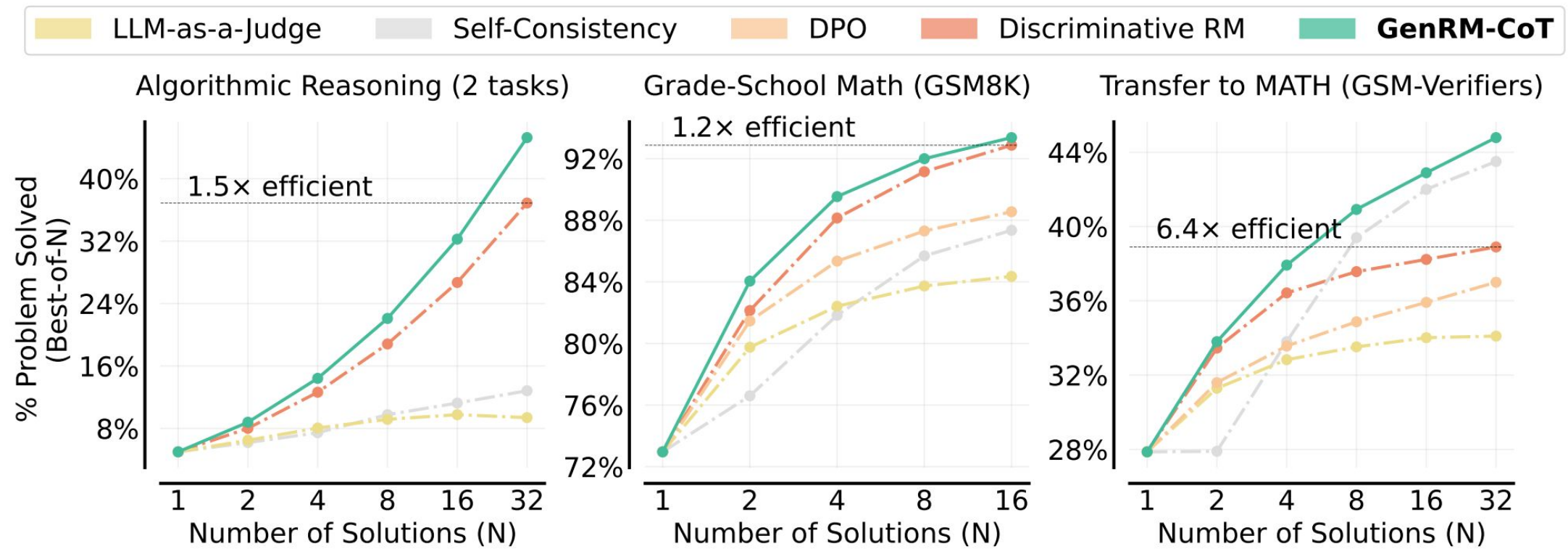
Generative Verifiers: Verify/label step-by-step [9]



Bring in CoT into LLM reward models

Synthesizing Labels/Rewards: Further Improvements

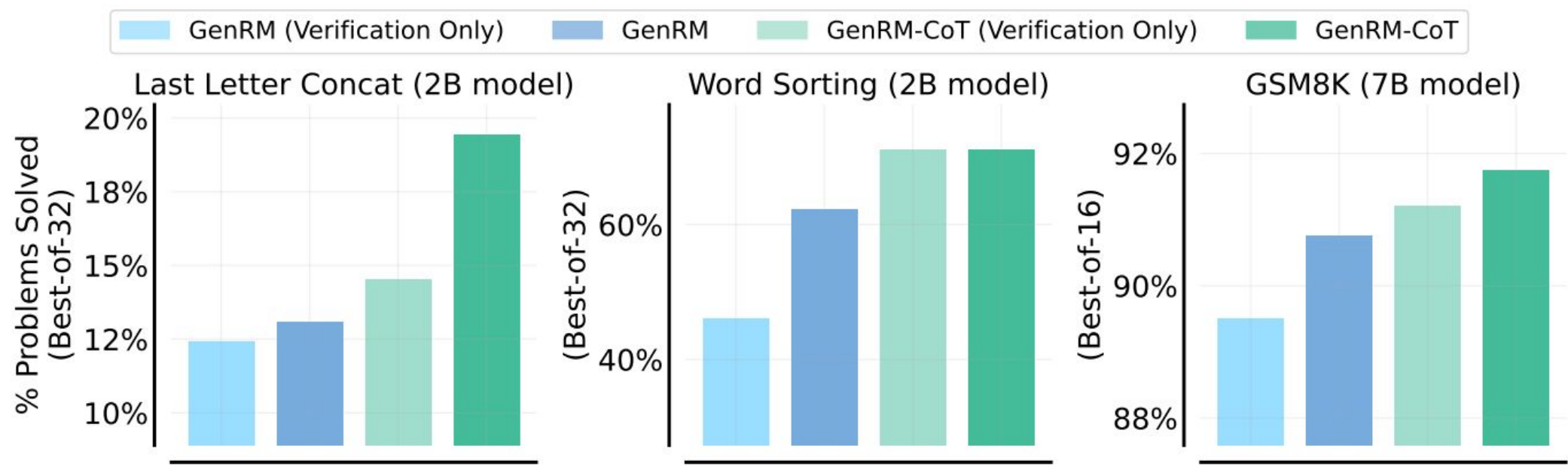
Generative Verifiers: Verify/label step-by-step [9]



Significantly Improved Data Efficiency on Reasoning Tasks

Synthesizing Labels/Rewards: Further Improvements

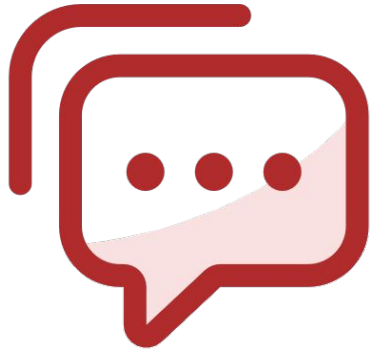
Generative Verifiers: Unifying Reward/Verification and Generation [9]



Potential to Unite Reward Model and Generation Model

slido

Please download and install the
Slido app on all computers you use



Audience Q&A

① Start presenting to display the audience questions on this slide.

Outline

Why Synthetic Data

Synthetizing Labels/Rewards

Synthetizing Training Data

What is Good Synthetic Data?

Recent Methodologies

Limitations

Synthesizing Training Data

Many scenarios not only limited in labels, but also data.

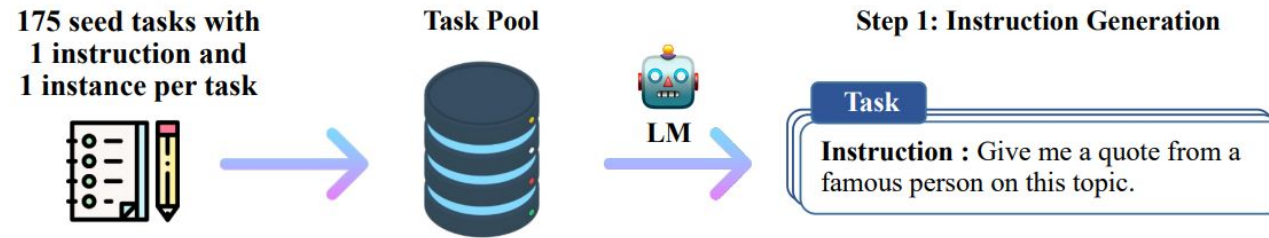
- E.g., instructions: there was no publicly available instructions from real LLM production

Generating data for certain part of the task maybe easier than learning the task

- Generating a question for a given answer versus answering it
- Generating a caption for a video versus generation a video

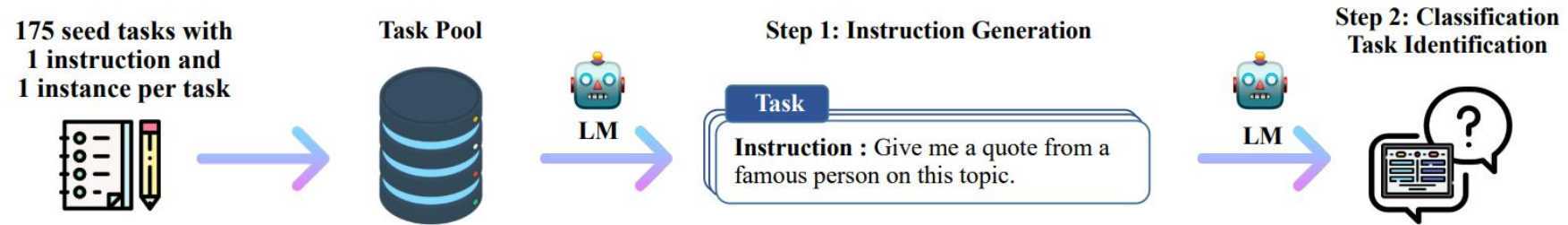
Synthesizing Training Data: Self-Instruct

Generating Instruction Tuning Data using LLMs



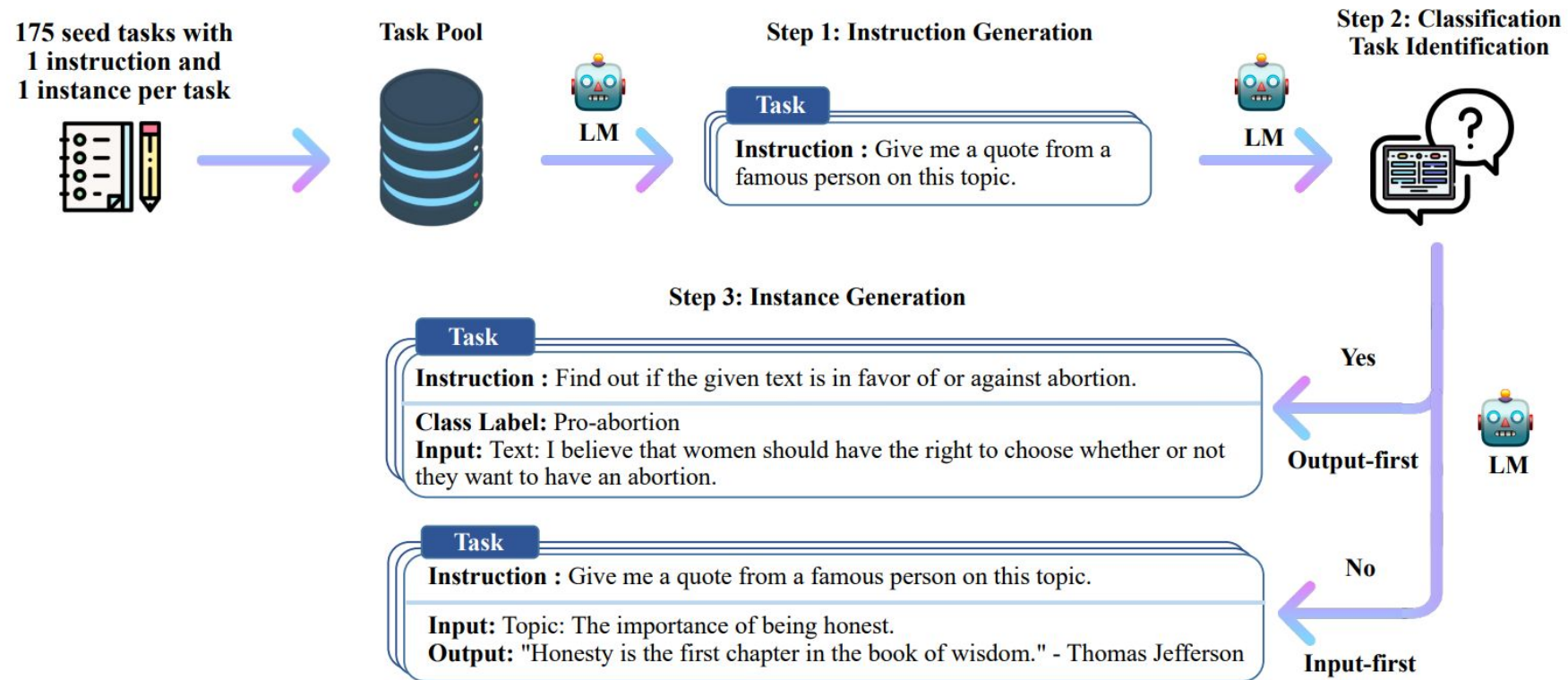
Synthesizing Training Data: Self-Instruct

Generating Instruction Tuning Data using LLMs



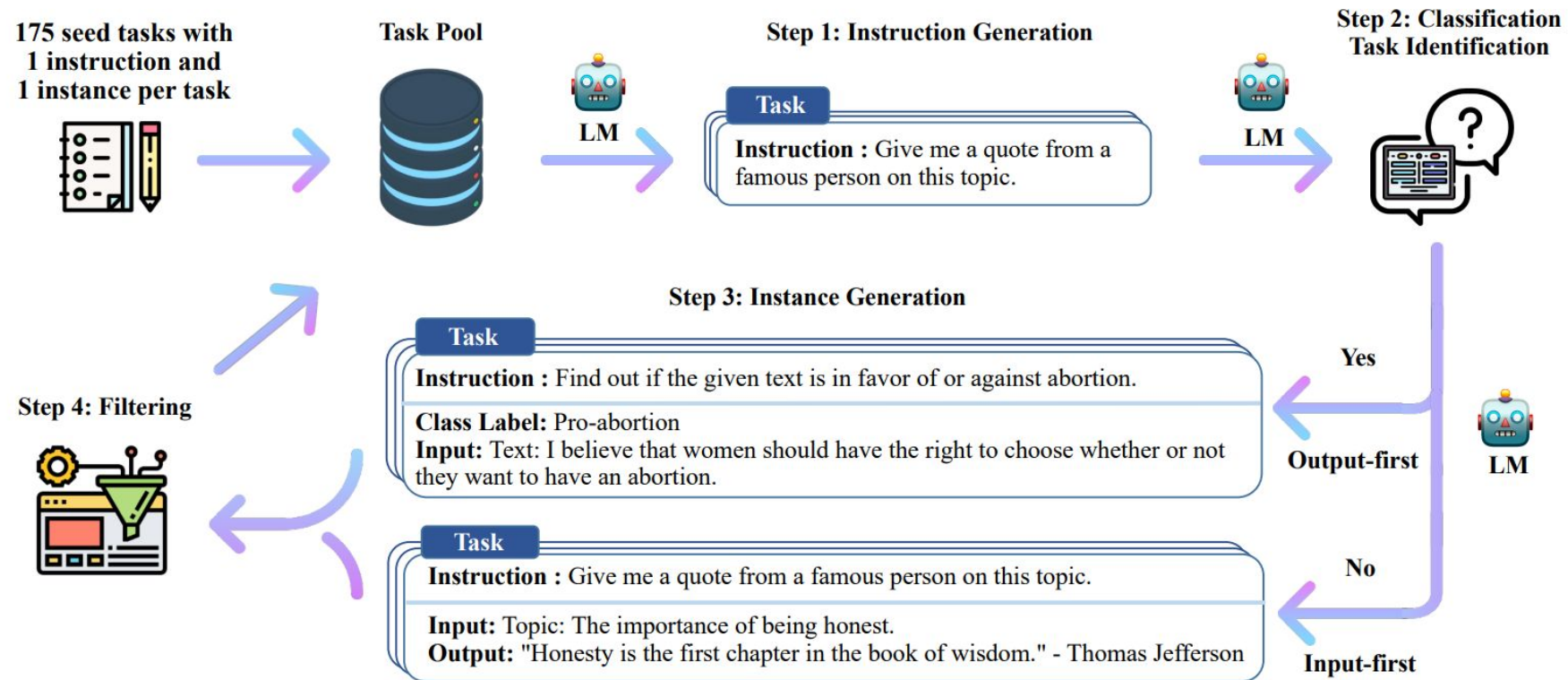
Synthesizing Training Data: Self-Instruct

Generating Instruction Tuning Data using LLMs



Synthesizing Training Data: Self-Instruct

Generating Instruction Tuning Data using LLMs



Synthesizing Training Data: Self-Instruct

Improved performance when fine-tuned upon

Significantly Improved
Instruction Following of
Vanilla GPT3

Distilled Most of
InstructGPT out

Model		# Params	ROUGE-L
Vanilla LMs			
T5-LM		11B	25.7
GPT3		175B	6.8
Instruction-tuned w/o SUPERNI			
①	T0	11B	33.1
	GPT3 + T0 Training	175B	37.9
②	GPT3 _{SELF-INST} (Ours)	175B	39.9
	InstructGPT ₀₀₁	175B	40.8
Instruction-tuned w/ SUPERNI			
Tk-INSTRUCT		11B	46.0
③	GPT3 + SUPERNI Training	175B	49.5
	GPT3 _{SELF-INST} + SUPERNI Training (Ours)	175B	51.6

Synthesizing Training Data: Self-Instruct

Improved performance when fine-tuned upon

Significantly Improved
Instruction Following of
Vanilla GPT3

Distilled Most of
InstructGPT out

Still Useful with
Supervised Labels

Model		# Params	ROUGE-L
Vanilla LMs			
T5-LM		11B	25.7
GPT3		175B	6.8
Instruction-tuned w/o SUPERNI			
T0		11B	33.1
GPT3 + T0 Training		175B	37.9
GPT3 _{SELF-INST} (Ours)		175B	39.9
InstructGPT ₀₀₁		175B	40.8
Instruction-tuned w/ SUPERNI			
Tk-INSTRUCT		11B	46.0
GPT3 + SUPERNI Training		175B	49.5
GPT3 _{SELF-INST} + SUPERNI Training (Ours)		175B	51.6

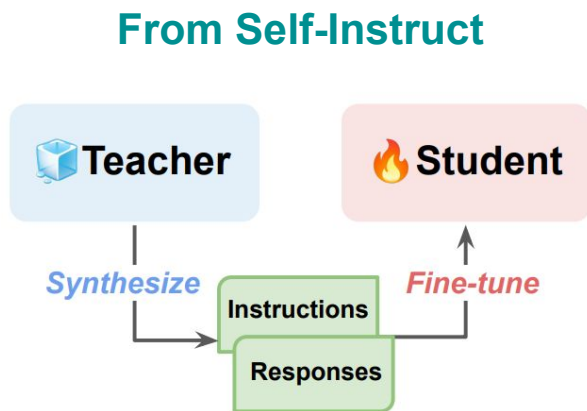
①

②

③

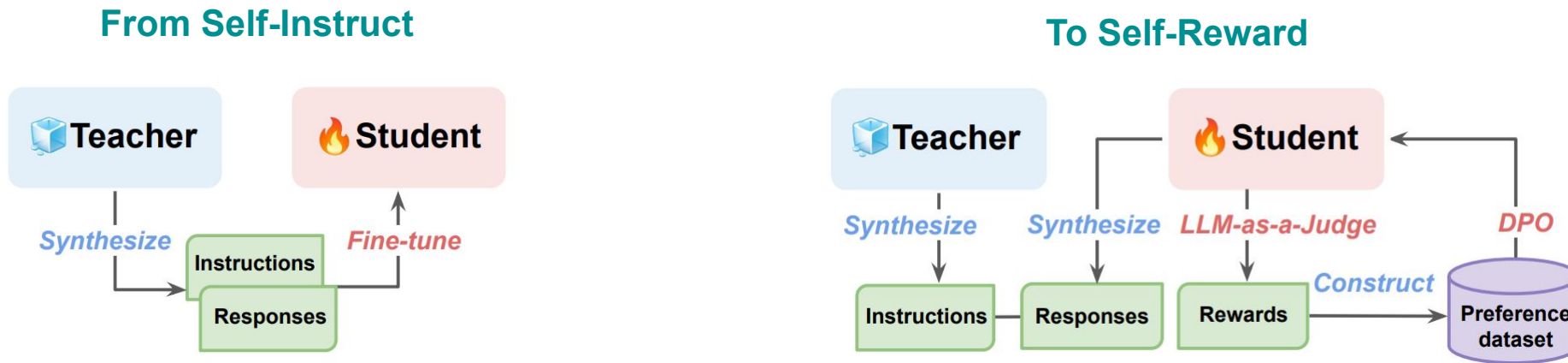
Synthesizing Training Data and Label: Self-Reward

Combine self-instruct and LLM-as-a-Judge [11]



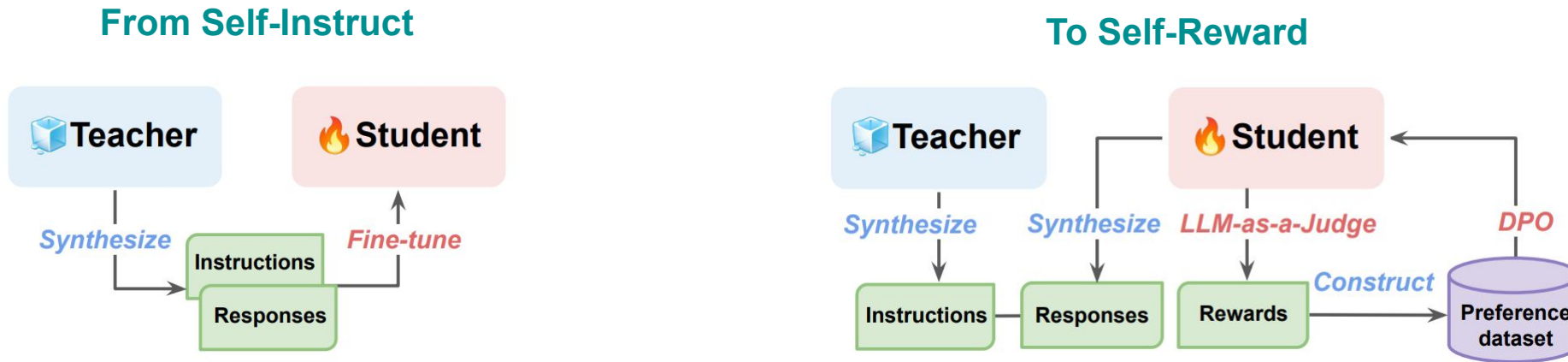
Synthesizing Training Data and Label: Self-Reward

Combine self-instruct and LLM-as-a-Judge [11]



Synthesizing Training Data and Label: Self-Reward

Combine self-instruct and LLM-as-a-Judge [11]



Use the student itself as a reward model

Synthesizing Training Data and Label: Self-Reward

Combine self-instruct and LLM-as-a-Judge [11]

	Overall Score	Math, Code & Reasoning	Humanities, Extraction, STEM, Roleplay & Writing
SFT Baseline	6.85	3.93	8.60
M_1	6.78	3.83	8.55
M_2	7.01	4.05	8.79
M_3	7.25	4.17	9.10

Improvement Through Iterations 1-3



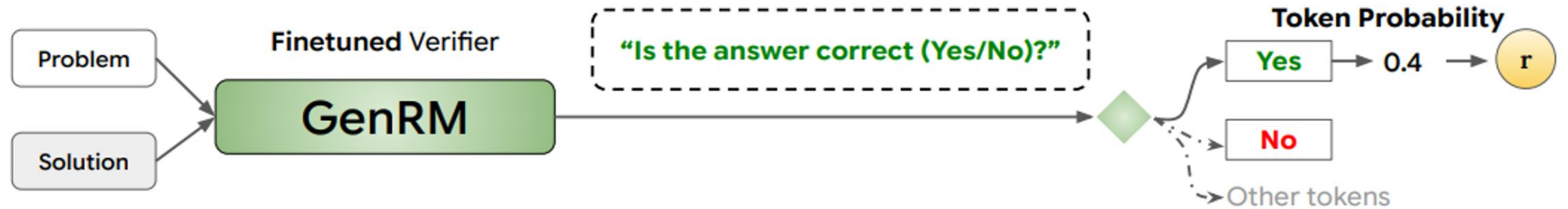
**What the reward signal from
the model itself can improve
its training?**

① Start presenting to display the poll results on this slide.

Synthesizing Training Data and Label: Self-Reward

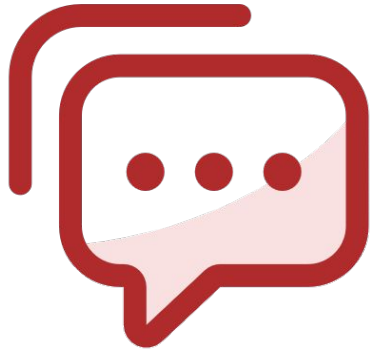
Task difference: evaluating a generation is easier than generating it

- Instruction are still from the teacher



slido

Please download and install the
Slido app on all computers you use



Audience Q&A

① Start presenting to display the audience questions on this slide.

Outline

Why Synthetic Data

Synthetizing Labels/Rewards

Synthetizing Training Data

What is Good Synthetic Data?

Recent Methodologies

Limitations

What is Good Synthetic Training Data?

High-Quality: Correctness

- Various filters and checkers

Diverse: not duplicate with each other

- Similarity based filters

What is Good Synthetic Training Data?

High-Quality: Correctness

- Various filters and checkers

Diverse: not duplicate with each other

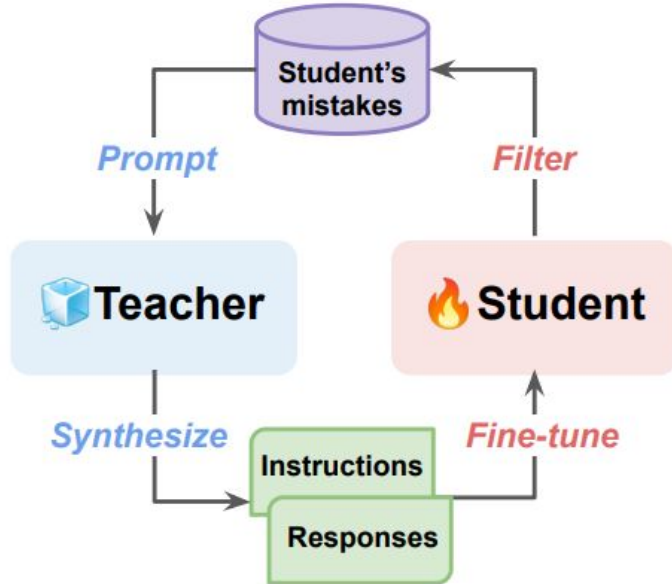
- Similarity based filters

Effective for training?

Synthetic Training Data Curation: Active Learning

Active Learning Assumption: training data that the model performs worse on are more valuable

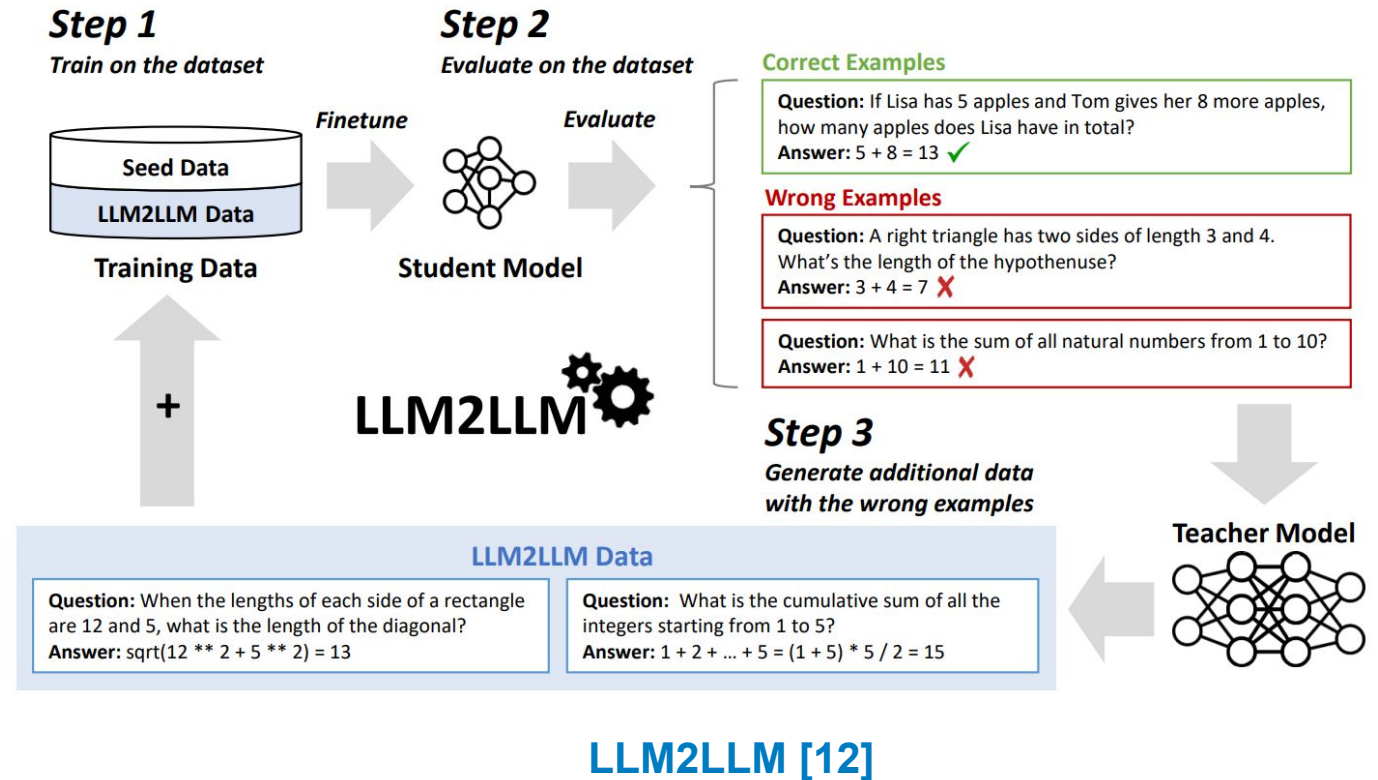
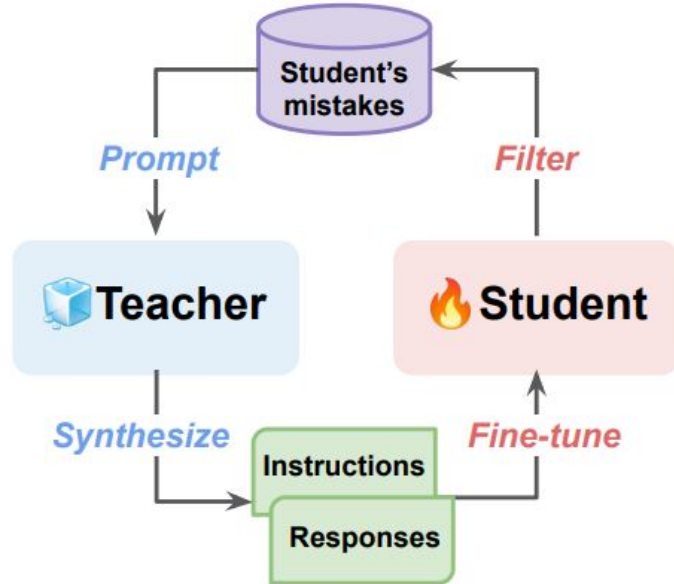
- Learn from mistakes
- Emphasize on data points hard for the model



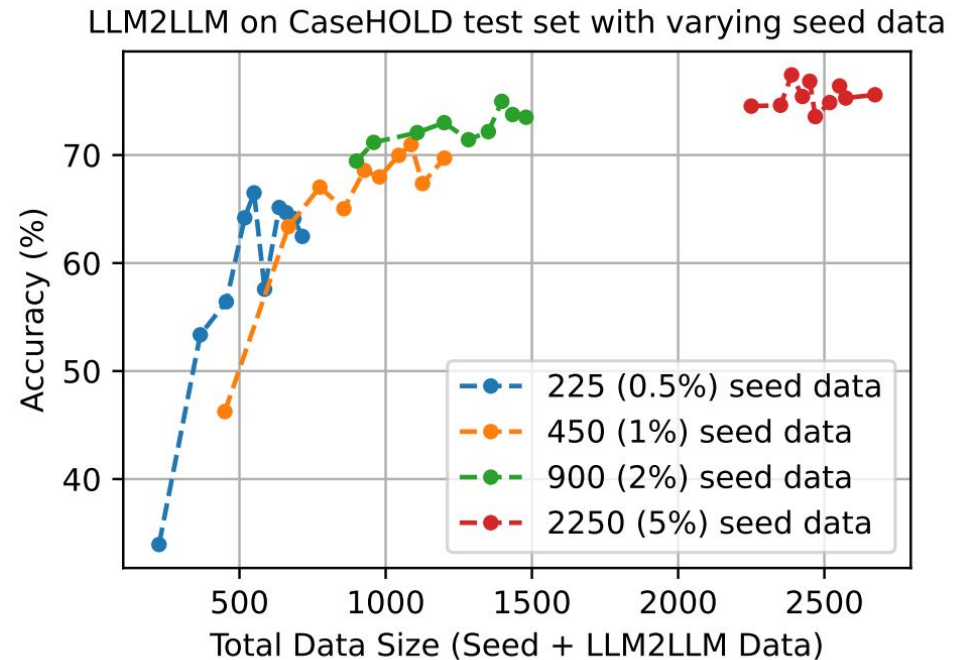
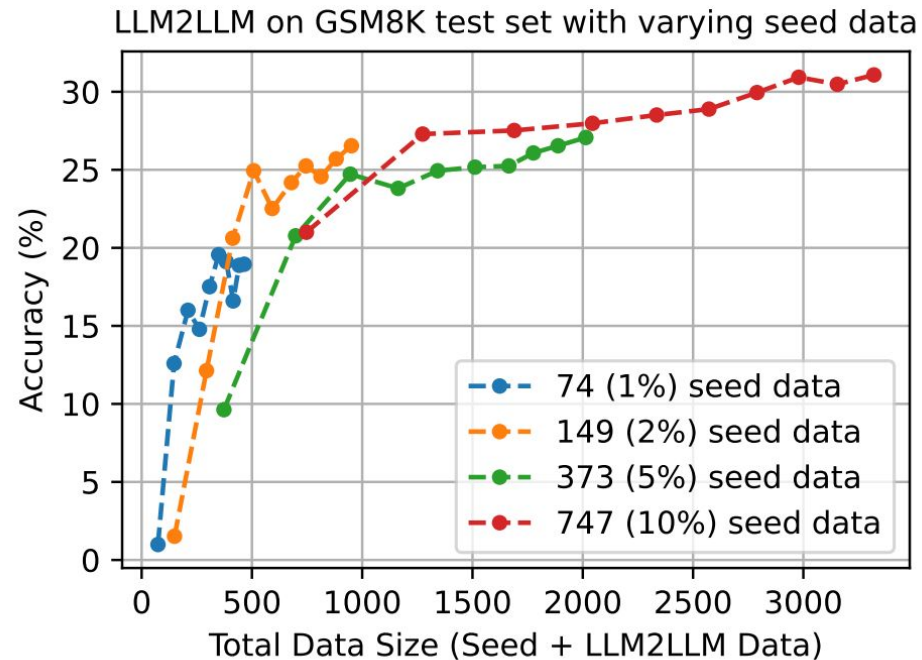
Synthetic Training Data Curation: Active Learning

Active Learning Assumption: training data that the model performs worse on are more valuable

- Learn from mistakes
- Emphasize on data points hard for the model



Synthetic Training Data Curation: Active Learning



Performance Starting from Different Number of Seed Data [12]

What is Good Synthetic Training Data?

- Data Influence: The improvement of the model when trained on this data point z_m [13]

$$\theta^* = \operatorname{argmin}_{\theta} \sum_i L(D, \theta)$$

Model with Training data D

$$\theta^*(\epsilon) = \operatorname{argmin}_{\theta} \sum_i L(D, \theta) + \epsilon L(z_m, \theta)$$

Model with Training data $D \cup z_m$

What is Good Synthetic Training Data?

- Data Influence: The improvement of the model when trained on this data point z_m [13]

$$\theta^* = \operatorname{argmin}_{\theta} \sum_i L(D, \theta)$$

Model with Training data D

$$\theta^*(\epsilon) = \operatorname{argmin}_{\theta} \sum_i L(D, \theta) + \epsilon L(z_m, \theta)$$

Model with Training data $D \cup z_m$

$$v(z_m) = L_{\text{ref}}(D_{\text{ref}}, \theta^*) - L_{\text{ref}}(D_{\text{ref}}, \theta^*(\epsilon))$$

Improvement on target reference metric

What is Good Synthetic Training Data?

- Data Influence: The improvement of the model when trained on this data point z_m [13]

$$\theta^* = \operatorname{argmin}_{\theta} \sum_i L(D, \theta)$$

Model with Training data D

$$\theta^*(\epsilon) = \operatorname{argmin}_{\theta} \sum_i L(D, \theta) + \epsilon L(z_m, \theta)$$

Model with Training data $D \cup z_m$

$$v(z_m) = L_{\text{ref}}(D_{\text{ref}}, \theta^*) - L_{\text{ref}}(D_{\text{ref}}, \theta^*(\epsilon))$$

Improvement on target reference metric

Two challenges:

1. Very expensive to calculate. Require actual training of the model
2. Pointwise data valuation while training is set level

Synthetic Training Data Influence: Influence Function Approximation

Local approximation of data Influence oracle

$$\begin{aligned}v(z_m) &\approx -\nabla_{\theta} L_{\text{ref}}(D_{\text{ref}}, \theta^*) \mathbf{H}^{-1} \nabla_{\theta} L(z_m, \theta^*) \epsilon \\ &\approx -\nabla_{\theta} L_{\text{ref}}(D_{\text{ref}}, \theta^*) \nabla_{\theta} L(z_m, \theta^*)\end{aligned}$$

Approximated using local Taylor Expansion

Ignore the Hessian for Efficiency

Synthetic Training Data Influence: Influence Function Approximation

Local approximation of data Influence oracle

$$v(z_m) \approx -\nabla_{\theta} L_{\text{ref}}(D_{\text{ref}}, \theta^*) \mathbf{H}^{-1} \nabla_{\theta} L(z_m, \theta^*) \epsilon$$

Approximated using local Taylor Expansion

$$\approx -\nabla_{\theta} L_{\text{ref}}(D_{\text{ref}}, \theta^*) \nabla_{\theta} L(z_m, \theta^*)$$

Ignore the Hessian for Efficiency

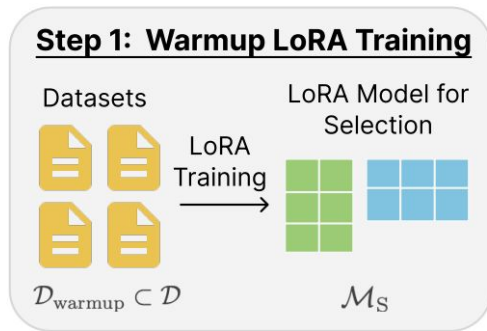
Measure data influences by its similarity of gradients with validation data

A common approach derived from many starting points

- Data influence
- Meta Learning

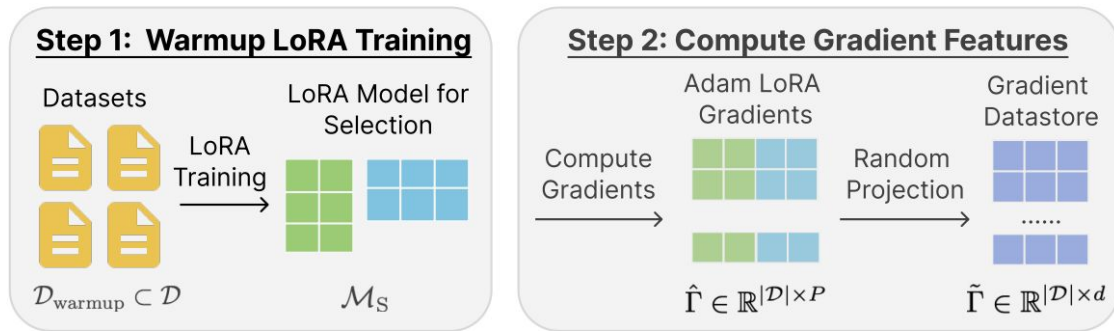
Synthetic Training Data Influence: Influence Function Approximation

LESS: selecting synthetic instruction tuning data by gradient similarities [14]



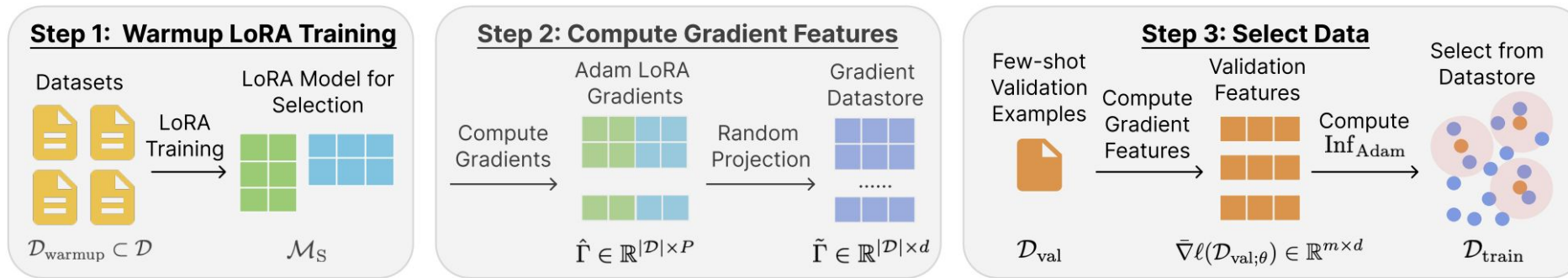
Synthetic Training Data Influence: Influence Function Approximation

LESS: selecting synthetic instruction tuning data by gradient similarities [14]



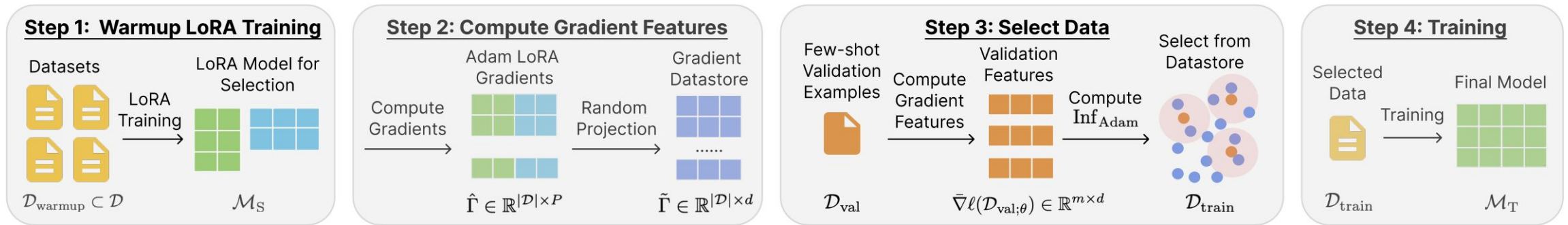
Synthetic Training Data Influence: Influence Function Approximation

LESS: selecting synthetic instruction tuning data by gradient similarities [14]



Synthetic Training Data Influence: Influence Function Approximation

LESS: selecting instruction tuning data by gradient similarities [14]



Synthetic Training Data Influence: Influence Function Approximation

LESS: selecting instruction tuning data by gradient similarities [14]

Data percentage	MMLU				TyDiQA				BBH			
	Full	Rand.	LESS-T	LESS	Full	Rand.	LESS-T	LESS	Full	Rand.	LESS-T	LESS
	(100%)	(5%)	(5%)	(5%)	(100%)	(5%)	(5%)	(5%)	(100%)	(5%)	(5%)	(5%)
LLAMA-2-7B	51.6	46.5 (0.5)	-	50.2 (0.5)	54.0	52.7 (0.4)	-	56.2 (0.7)	43.2	38.9 (0.5)	-	41.5 (0.6)
LLAMA-2-13B	54.5	53.4 (0.1)	54.6 (0.3)	54.0 (0.7)	54.3	53.0 (1.3)	57.5 (0.8)	54.6 (0.3)	50.8	47.0 (1.6)	49.9 (0.5)	50.6 (0.6)
MISTRAL-7B	60.4	60.0 (0.1)	60.6 (0.3)	61.8 (0.4)	57.7	56.9 (0.2)	61.7 (1.7)	60.3 (2.4)	53.0	54.5 (0.1)	56.0 (0.8)	56.0 (1.0)

Performance with Gradient Based Instruction Data Selection [14]

Synthetic Training Data Influence: Model Approximation

- Learn the oracle data influence using a parameterized model
 - Sample a bunch of z , train and evaluate the target model to obtain their oracle influence

$$v(z_1) = L_{\text{ref}}(D_{\text{ref}}, \theta^*) - L_{\text{ref}}(D_{\text{ref}}, \theta^*(z_1, \epsilon))$$

$$v(z_2) = L_{\text{ref}}(D_{\text{ref}}, \theta^*) - L_{\text{ref}}(D_{\text{ref}}, \theta^*(z_2, \epsilon))$$

$$v(z_3) = L_{\text{ref}}(D_{\text{ref}}, \theta^*) - L_{\text{ref}}(D_{\text{ref}}, \theta^*(z_3, \epsilon))$$

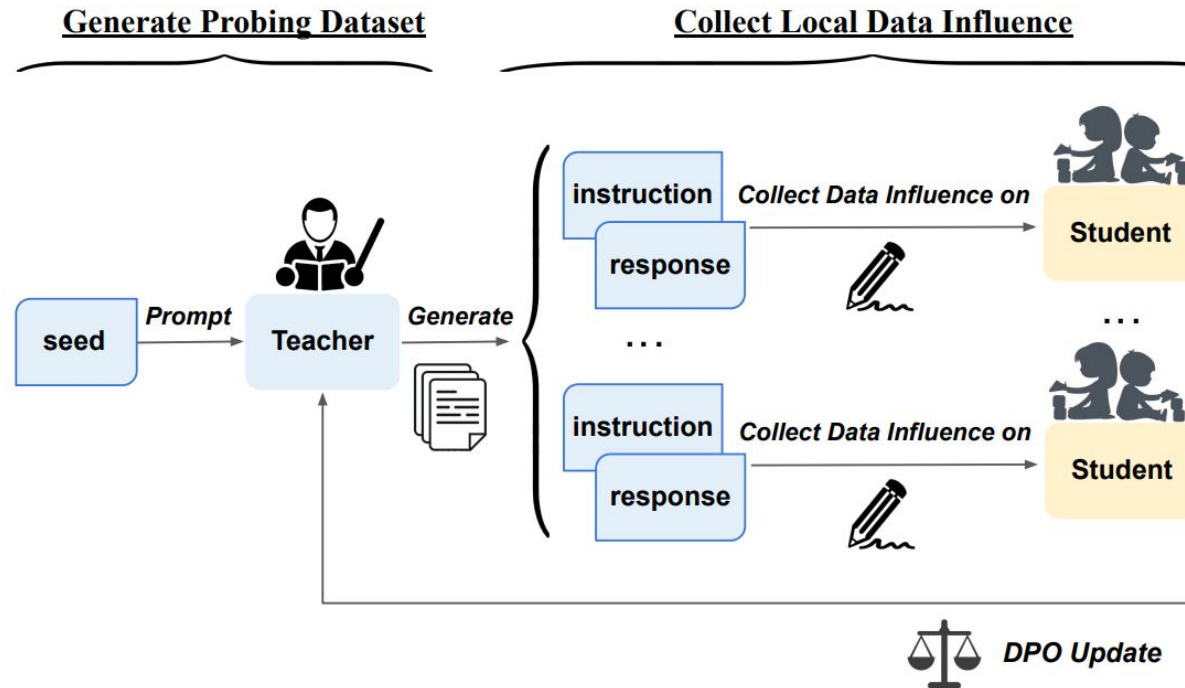
...

$$v(z_n) = L_{\text{ref}}(D_{\text{ref}}, \theta^*) - L_{\text{ref}}(D_{\text{ref}}, \theta^*(z_n, \epsilon))$$

- Use a model to learn the function $v()$ using training data $(z, v(z))$

Synthetic Training Data Influence: Model Approximation

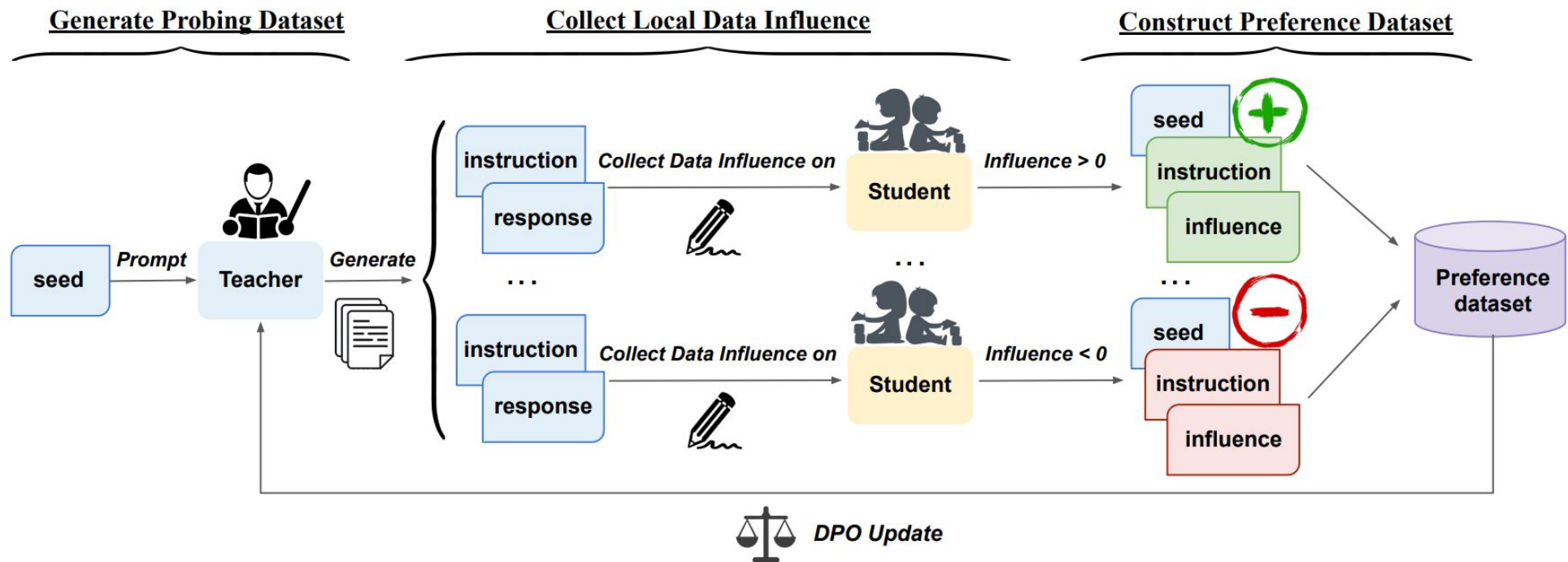
Learn the oracle data influence in the teacher model



Building a Feedback Loop from Data Influence to Data Synthesizer [15]

Synthetic Training Data Influence: Model Approximation

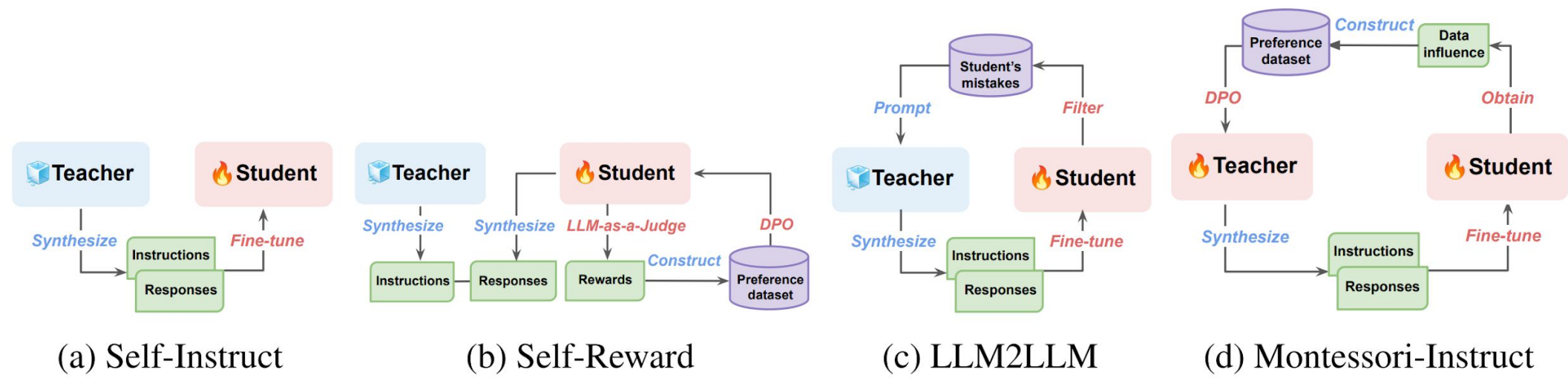
Learn the oracle data influence in the teacher model



Building a Feedback Loop from Data Influence to Data Synthesizer [15]

Synthetic Training Data Influence: Model Approximation

Various controls on synthetic training data curation



Synthetic Training Data Influence: Model Approximation

Learn to generate more influential synthetic data

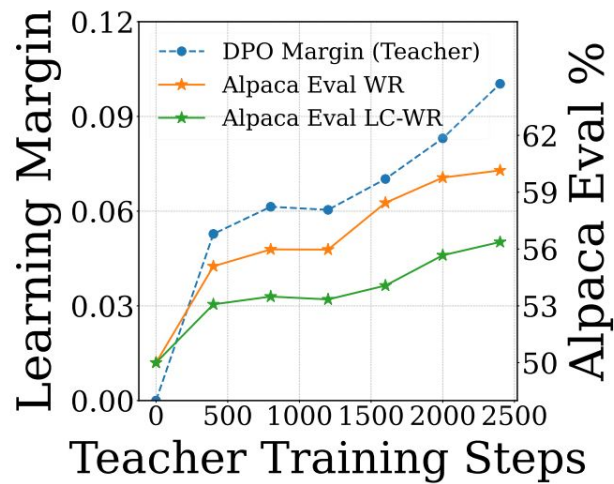
Methods	In-Domain		MT-Bench	Out-Of-Domain				
	Alpaca Eval 2.0			MMLU	GPQA	ARC-C	GSM8K	HellaSwag
	LC-WR	WR						
8B Setting: Student=Llama3-8B								
<i>No fine-tuning</i>	2.09%	3.39%	5.597	62.15	24.33	57.85	51.25	<u>81.96</u>
<i>Self-Instruct</i>	50%	50%	6.490	62.42	31.92	59.98	<u>58.76</u>	80.93
<i>Self-Instruct*</i>	<u>54.95%</u>	56.39%	5.918	<u>63.41</u>	30.13	60.58	50.42	81.42
<i>Self-Reward*</i>								
Iteration 1	51.87%	55.38%	6.713	62.46	28.19	59.84	53.60	81.04
Iteration 2	53.49%	57.32%	6.798	62.02	29.08	<u>60.64</u>	56.37	81.13
<i>LLM2LLM</i>								
Iteration 1	51.49%	53.12%	6.531	62.18	29.12	57.49	55.28	80.49
Iteration 2	52.63%	55.02%	6.519	62.46	30.04	59.65	57.75	80.57
<i>Montessori-Instruct</i>								
Iteration 1	54.92%	<u>58.59%</u>	<u>6.903</u>	62.93	29.91	62.97	<u>58.76</u>	81.22
Iteration 2	56.82%	60.23%	7.092	63.44	<u>31.19</u>	59.98	60.05	81.98

Performance of Different Synthetic Data Curation Methods [15]

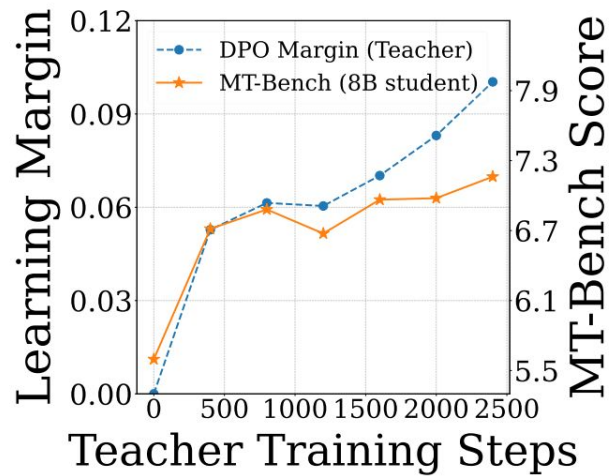
[15] Li et al. 2024. Montessori-Instruct: Generate Influential Training Data Tailored for Student Learning

Synthetic Training Data Influence: Model Approximation

Learn to generate more influential synthetic data



(a) Alpaca Eval

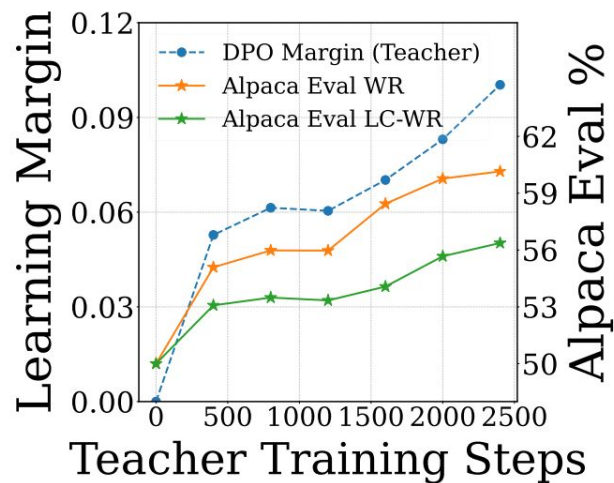


(b) MT-Bench

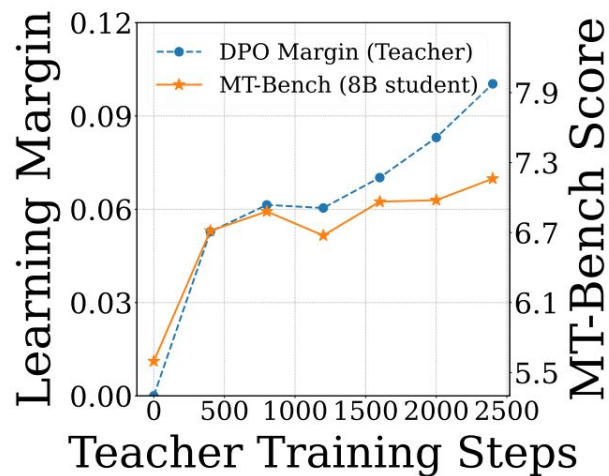
Generating More Influential Data Through Iterations [15]

Synthetic Training Data Influence: Model Approximation

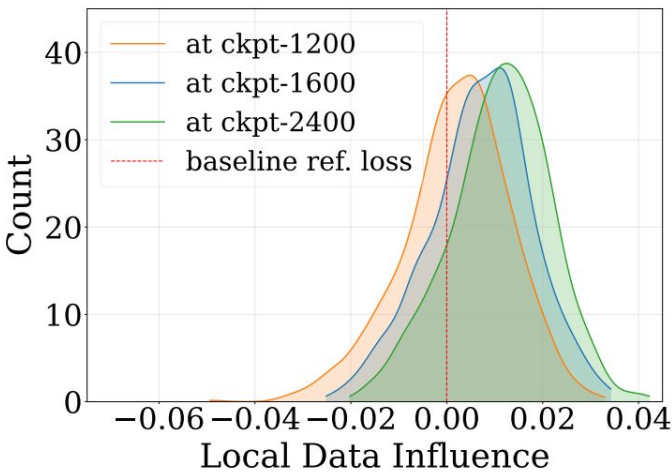
Learn to generate more influential synthetic data



(a) Alpaca Eval



(b) MT-Bench



(c) Data influence

Generating More Influential Data Through Iterations [15]

Recap

Prompt LLMs to generate synthetic training data

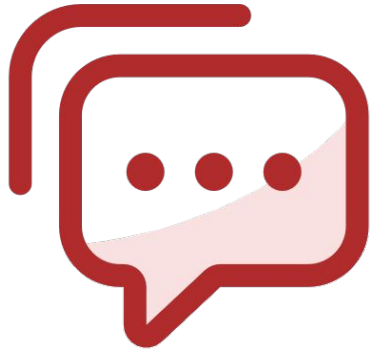
- Labels
- Generative Rewards
- Entire Training Data

Various recent progresses in identifying good synthetic training data

- Quality
- Self-consistency
- Informativeness
- Data influence and its approximations

slido

Please download and install the
Slido app on all computers you use



Audience Q&A

① Start presenting to display the audience questions on this slide.

Recap

Prompt LLMs to generate synthetic training data

- Labels
- Generative Rewards
- Entire Training Data

Various recent progresses in identifying good synthetic training data

- Quality
- Self-consistency
- Informativeness
- Data influence and its approximations

All the above work on individual data points, but training is done on the set

Model Collapse with Synthetic Training

Lost of data variety in data synthetize

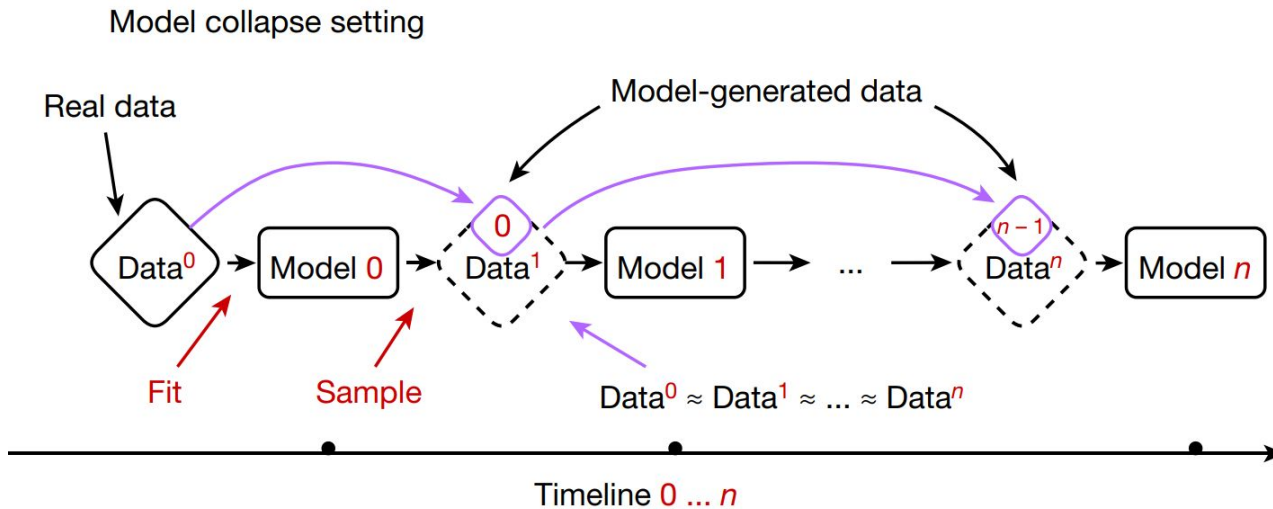
- Synthetic dataset lacks long tail variety
- Though may maintain average quality



Model Collapse with Synthetic Training

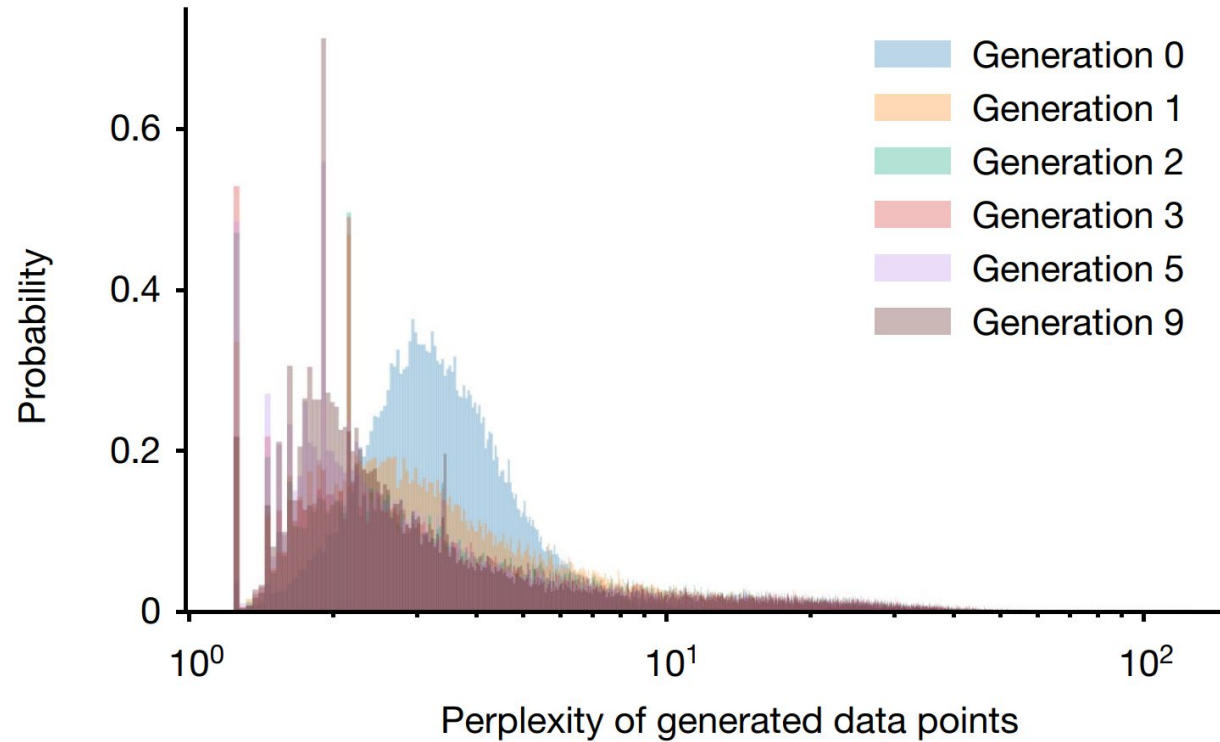
Lost of data variety in data synthetize

- Synthetic dataset lacks long tail variety
- Though may maintain average quality



Model Collapse with Synthetic Training

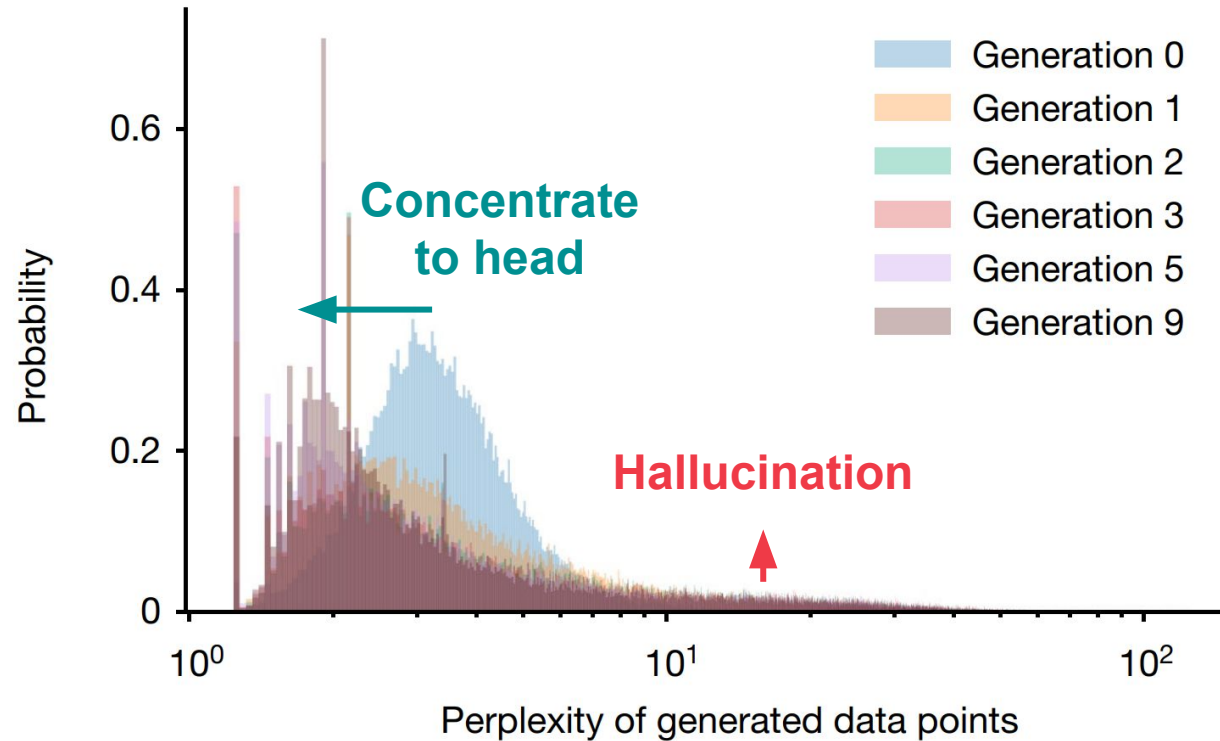
Generated data distribution loses the high perplexity part



Perplexity of Synthetic Data Evaluated by Wikitext LM [16]

Model Collapse with Synthetic Training

Generated data distribution loses the high perplexity part



Perplexity of Synthetic Data Evaluated by Wikitext LM [16]

Recap

Prompt LLMs to generate synthetic training data

- Labels
- Generative Rewards
- Entire Training Data

Various recent progresses in identifying good synthetic training data

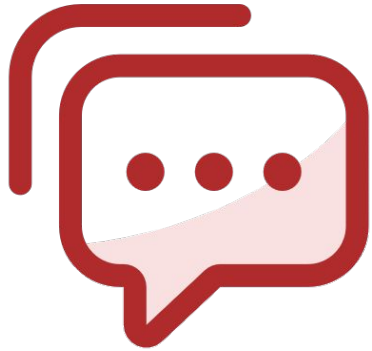
- Quality
- Self-consistency
- Informativeness
- Data influence and its approximations

Current capability

- Distillation
- Mitigate the lack of SFT data, but not replace them
- A few round of benefits from self-play
- Assist pretraining

slido

Please download and install the
Slido app on all computers you use



Audience Q&A

① Start presenting to display the audience questions on this slide.