

Synthetic Data *in the Era of LLMs*



Vijay Viswanathan (Carnegie Mellon)



Xiang Yue (Carnegie Mellon)



Alisa Liu (Washington)



Yizhong Wang (Washington)



Graham Neubig (Carnegie Mellon)



Motivation

Data → NLP Progress

Language models are built on data

Pre-training

Raw text x  $P(x)$

Supervised Fine Tuning

Input x , output y  $P(y | x)$

Reasoning Training

Input x , output y , latent reasoning z  $P(y | z, x) P(z | x)$

Where do we get data?

- Scraping the internet
- Labeling manually
- Collecting from system users
- Creative curation

Why is this not enough?

- Scraping the internet
Too noisy, too massive
- Labeling manually
Too expensive, annotators
not available
- Collecting from users
Chicken and egg, privacy
implications
- Creative curation
Limited applicability

Synthetic data to the rescue!

Create data order-made that is

- Relatively clean
- Appropriately sized (not too big/small)
- Tailored to individual tasks
- Flexible

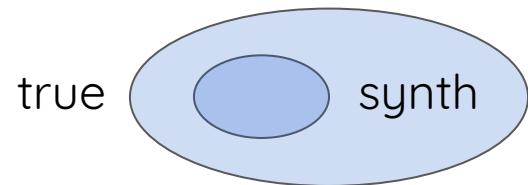
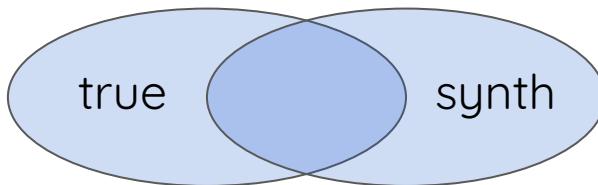
But getting all of these at once is hard!

But generating good synthetic data is hard...

The input distribution may be **off**, or **not diverse enough**

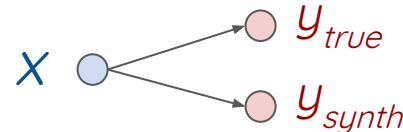
$$P_{\text{true}}(x)$$

$$\neq P_{\text{synth}}(x)$$



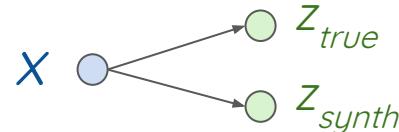
The labels may be **wrong**

$$P_{\text{true}}(y | x) \neq P_{\text{synth}}(y | x)$$



The reasoning may be **flawed**

$$P_{\text{true}}(z | x) \neq P_{\text{synth}}(z | x)$$



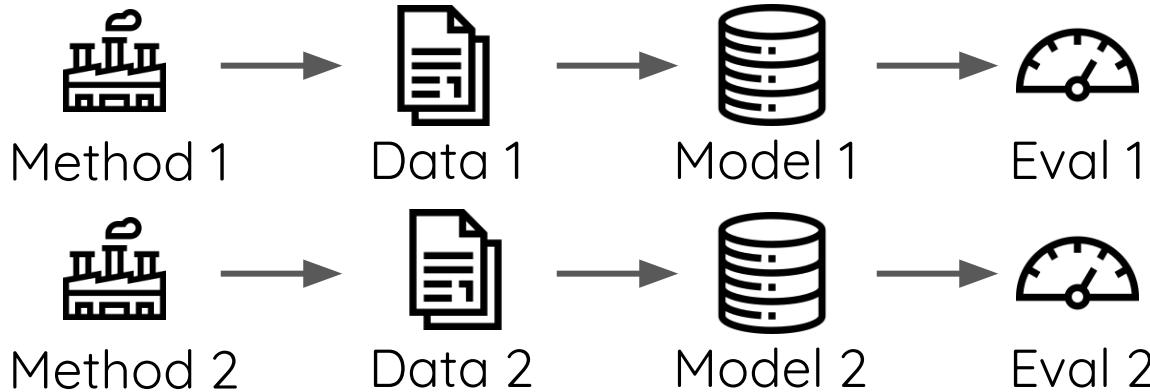
In this tutorial we will cover

- 2pm: How do we **evaluate** data quality?
(15 minutes)
- 2:20pm: How do we **create** high-quality synthetic data?
(35 minutes + QA)
- 3:05pm: How do we **use** synthetic data?
(40 minutes + QA)
- 3:30pm: 30 minute-break
- 4:25pm: Scenario-specific **applications**
(35 minutes + QA)
- 5:00pm: Limitations and open questions
(25 minutes + QA)
- 5:30pm: End

**What is “high-quality” synthetic
data?**

Evaluation of synthetic data

- **Extrinsic:** Does it help in a downstream task?



- **Intrinsic:** What are the characteristics of the data or generation process?

Intrinsic Eval: Data Correctness

- Questions regarding whether the data is correct, judged by manual or automatic methods
- E.g. Self-Instruct manually annotates:

Quality Review Question	Yes %
Does the instruction describe a valid task?	92%
Is the input appropriate for the instruction?	79%
Is the output a correct and acceptable response to the instruction and input?	58%
All fields are valid	54%

Intrinsic Eval: Data Diversity/Coverage

- How well does the generated data cover the plausible data distribution?
- E.g. DataTune evaluates bigram diversity

Dataset	Unique Bigrams Per Example	Total Tokens Per Example
Code Line Description		
Gold	13.2	32.3
Synthetic	2.5	35.0
Transformed	14.9	86.9
Elementary Math		
Gold	10.8	48.6
Synthetic	3.3	34.4
Transformed	11.6	43.8
Implicatures		
Gold	9.9	24.1
Synthetic	2.7	27.7
Transformed	17.8	39.8

Intrinsic Eval: Other Metrics

- Many other dimensions, e.g. privacy, fairness, distributional similarity
- E.g. SynthTextEval toolkit

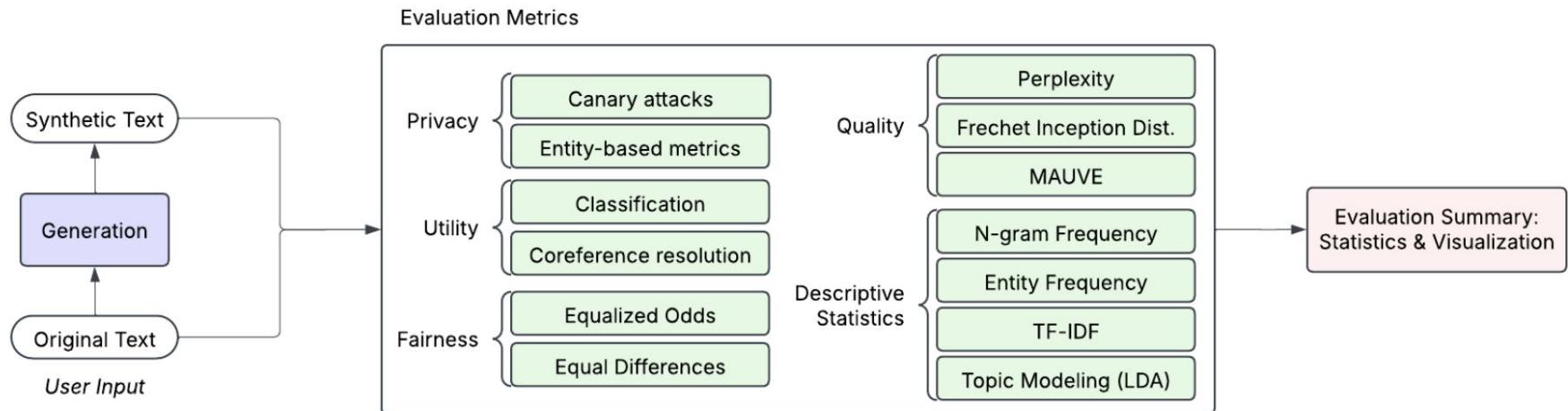
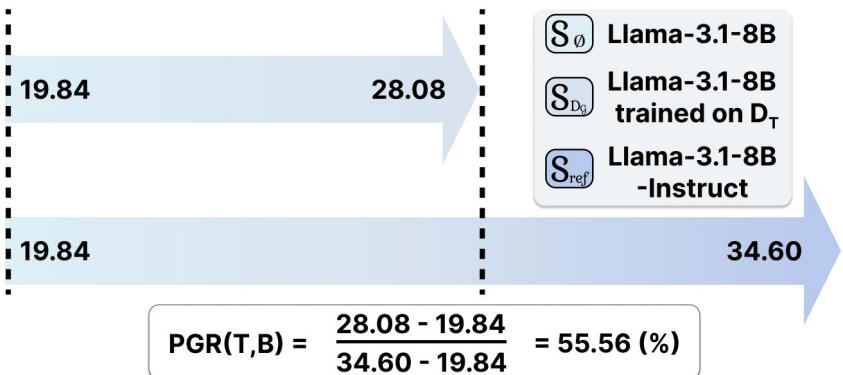


Figure 1: Architecture overview of SYNTHTEXEVAL.

Evaluating Language Models as Data Generators

- We can also evaluate language models based on their ability to generate synthetic data
- E.g. AgoraBench, which measures synthetic data by different LMs based on its ability to match manually created data (at what cost)



Data Generator	API Cost		Prob. Solv.	Data Gen.
	Input	Output		
GPT-4o	\$2.50	\$10.00	80.9	29.5%
GPT-4o-mini	\$0.15	\$0.60	75.4	19.2%
Claude-3.5-Sonnet	\$3.00	\$15.00	80.5	23.6%
Llama-3.1-405B	\$1.79	\$1.79	75.0	11.3%
Llama-3.1-70B	\$0.35	\$0.40	69.6	14.1%
Llama-3.1-8B	\$0.055	\$0.055	50.2	15.9%

How do we create synthetic data?

Approaches to synthetic data creation

Sampling-based generation

Back-translation

Transformation of existing data

Human-AI collaboration

Symbolic generation

Approaches to synthetic data creation

Sampling-based data generation

Back-translation

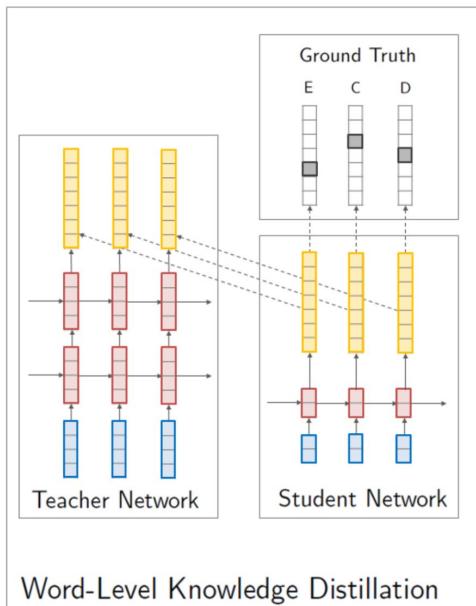
Transformation of existing data

Human-AI collaboration

Symbolic generation

Background: knowledge distillation

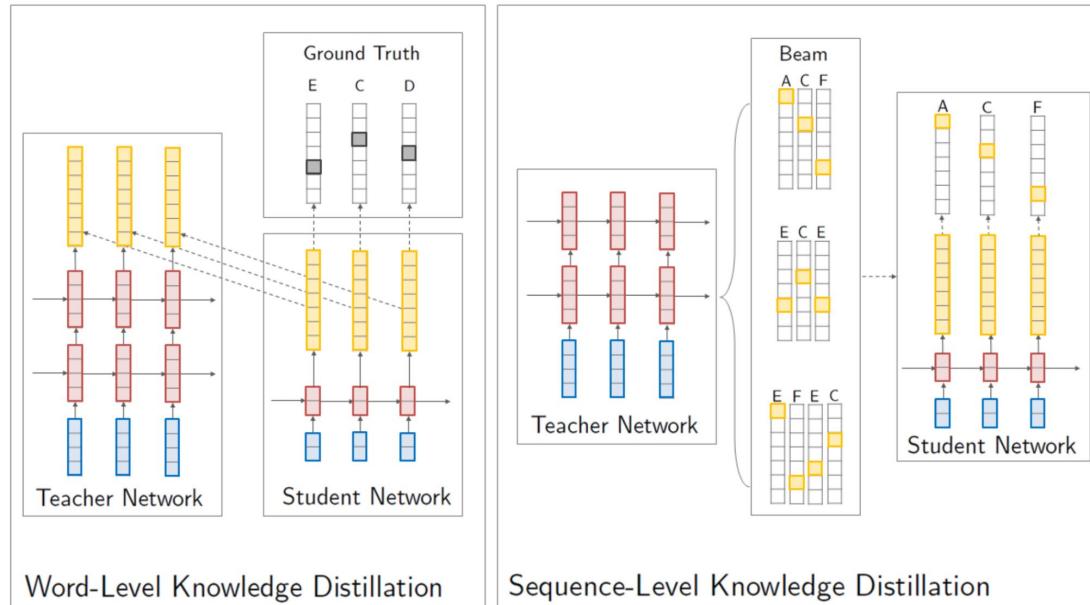
Train student model to mimic the teacher's predicted probability distribution (e.g., over words)



“Once the cumbersome model has been trained, we can then use a different kind of training, which we call “distillation” to transfer the knowledge to a small model”

Sequence -level knowledge distillation

Train student on complete generations (i.e., sequences of words) from the teacher



Generating task data from LMs

Use GPT-3's in-context learning ability to generate new examples of arbitrary tasks

Task: Write two sentences that mean the same thing

Sentence 1: A man is playing the flute
Sentence 2: He's playing the flute

Create sentence-similarity examples by prompting the model to write similar (or dissimilar) sentences!

Generating instruction data from scratch

Instead of generating more examples under a given task,
generate completely new tasks

Come up with a series of
tasks.

{*in-context examples*}

Task: Given an address and
city, come up with the zip
code.

Come up with examples for the
following tasks.

{*in-context examples*}

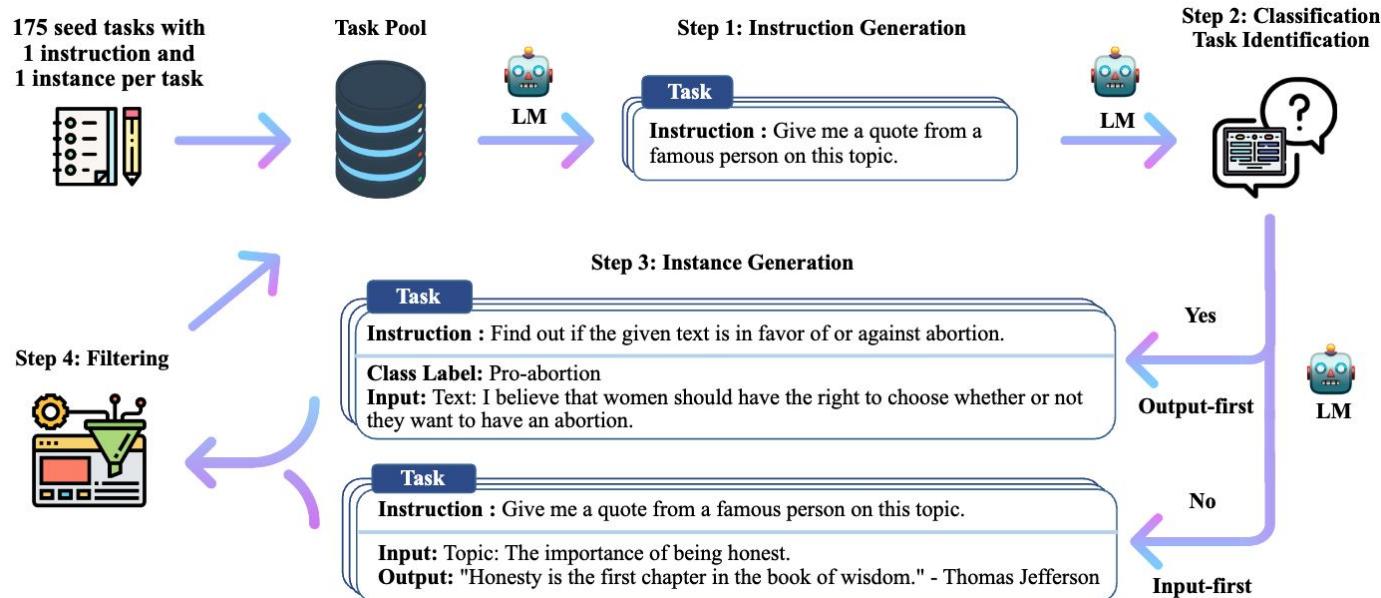
Task: Given an address and city, come
up with the zip code.

Input: 123 Main Street, San Francisco

Output: 94105

Generating instruction data from scratch

From just 175 seed examples → ~100K new examples

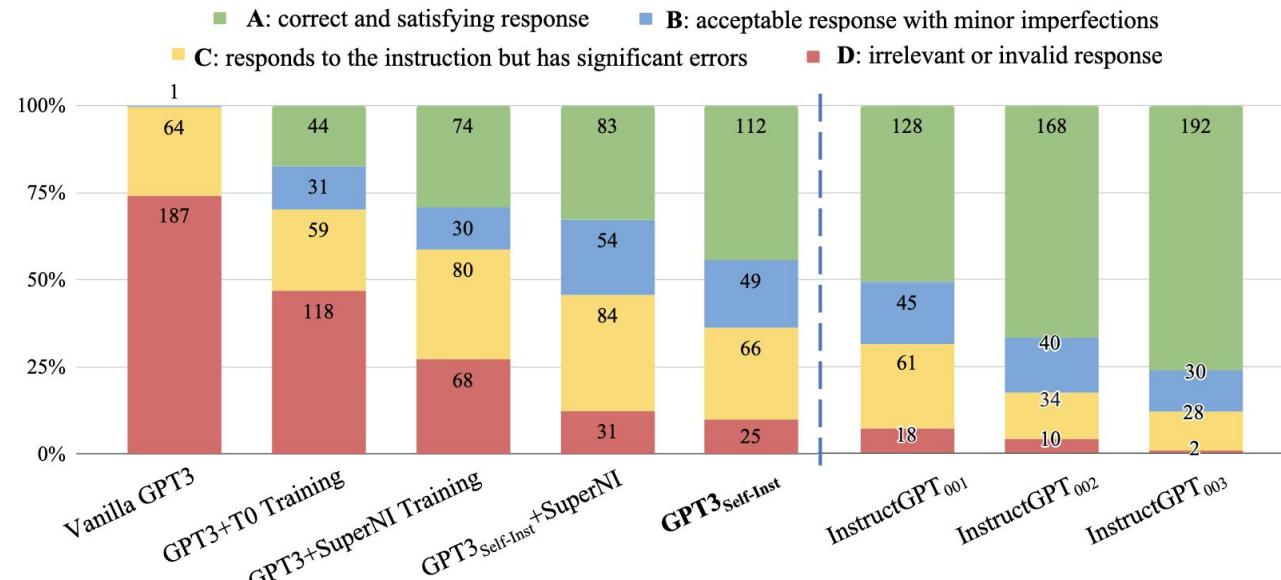


[Self-Instruct: Aligning Language Models with Self-Generated Instructions \(wang et al., 2022\)](#)

[Unnatural Instructions: Tuning Language Models with \(Almost\) No Human Labor \(Honovich et al., 2022\)](#)

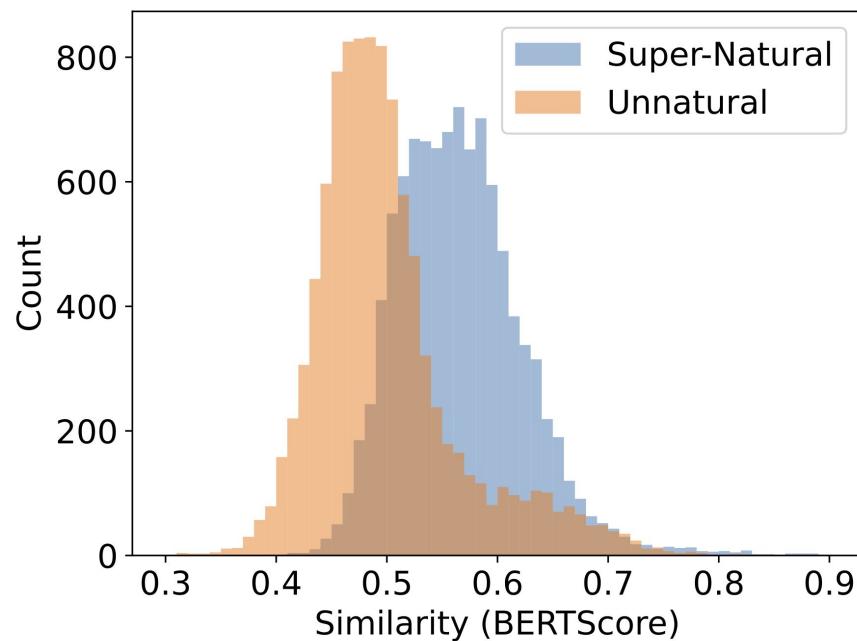
Generating instruction data from scratch

Finetuning GPT-3 on self-generated data improves over existing instruction datasets



Generating instruction data from scratch

Generated data is more diverse than human-written data



Generating instruction data from scratch

In both Self-Instruct data and Unnatural Instructions, only half of the examples are actually correct (!!)

Quality Review Question	Yes %
Does the instruction describe a valid task?	92%
Is the input appropriate for the instruction?	79%
Is the output a correct and acceptable response to the instruction and input?	58%
All fields are valid	54%

113 of the 200 analyzed examples (56.5%) are correct. Of the 87 incorrect examples, 9 (4.5%) had incomprehensible instructions, 35 (17.5%) had an input that did not match the task description, and 43 (21.5%) had incorrect outputs.

Table 2: Data quality review for the instruction, input, and output of the generated data.

Takeaways from early efforts

Synthetic data can reflect creativity & diversity difficult to elicit from crowdworkers

Diversity can be more valuable than correctness!

Synthetic data can sometimes enable self-improvement

Data creation becomes a complex pipeline

Approaches to synthetic data creation

Sampling-based generation

Back-translation

Transformation of existing data

Human-AI collaboration

Symbolic generation

Approaches to synthetic data creation

Sampling-based generation

Back-translation

Transformation of existing data

Human-AI collaboration

Symbolic generation

Background: back-translation in MT

Given the output, generate a corresponding input

Since models are trained to produce outputs, so we want those to be natural (inputs can be unnatural)

[TODO: add some illustration]

Instruction back-translation

Given web text y , generate instruction x for which y would be a good response

$y \rightarrow x$ can be easier than $x \rightarrow y$!

Output: It doesn't matter where you are in the world, how old you are, or how much you know about meditation, it's for everyone. The benefits of meditation are endless. Medication can be as simple as sitting quietly for five minutes...

What kind of instruction could this be the answer to?

Instruction: Write an essay about the benefits of meditation.

Instruction back-translation

TODO: some results

Approaches to synthetic data creation

Sampling-based generation

Back-translation

Transformation of existing data

Human-AI collaboration

Symbolic generation

Approaches to synthetic data creation

Sampling-based generation

Back-translation

Transformation of existing data

Human-AI collaboration

Symbolic generation

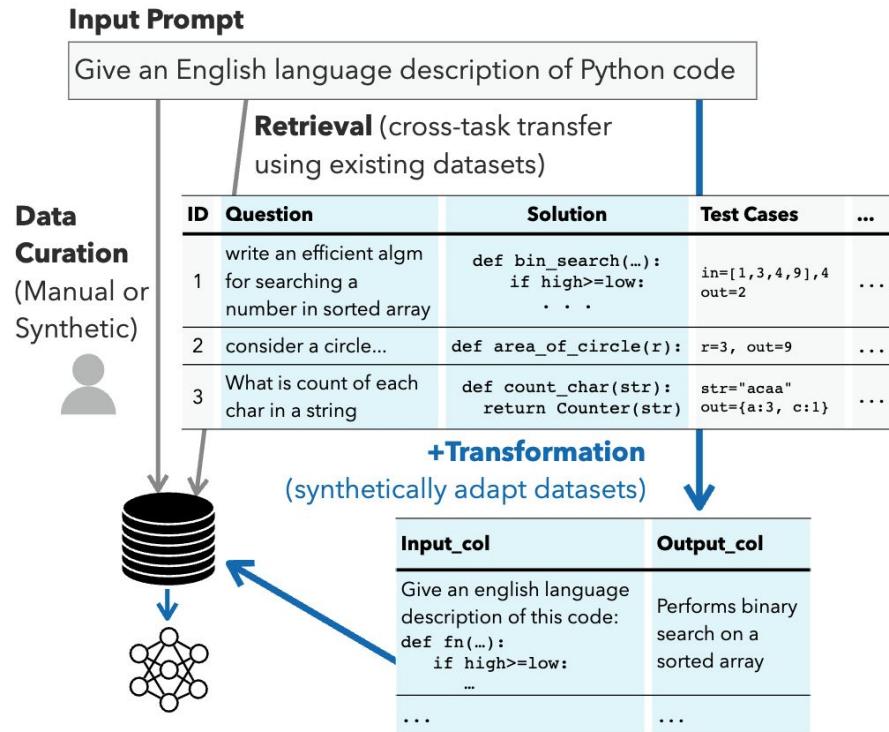
Transformation of existing data

Use or retrieve existing data, then transform it into an example under the desired task

Transform existing data into examples

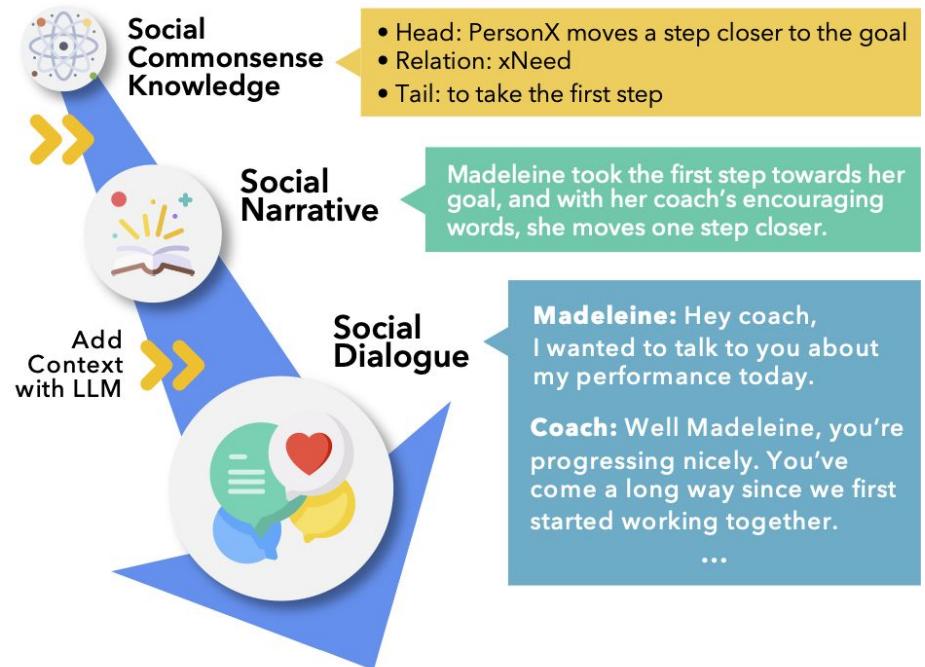
Given a task,

1. Retrieve relevant datasets or documents
2. Transform them into data under the desired task



Ground generation in knowledge graphs

Condition on data from a large commonsense knowledge graph to produce diverse dialogues



Extract instruction data from the web

Identify pages that may contain questions & answers, then extract and refine them!

 Raw Docs Unformatted Text, Site Information, Ads

Topics Science\nAnatomy&Physiology\nAstronomy\nAstrophysics
\nBiology\nChemistry \n...Socratic Meta...Featured Answers

How do you simplify $((u^4 v^3)/(u^2 v^{-1})^4)^0$ and write it using only positive exponents?

Answer by NickTheTurtle (Apr 1, 2017)

Explanation: Anything raised to the 0^{th} power is simply 1.

Related Questions What is the quotient of powers property?

How do you simplify expressions using the quotient rule?...Impact of this question 1274 views around the world

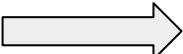
#Apps\niOS\nAndroid\nLinks\n[Privacy](#)\n[Terms](#)\n[Help](#)

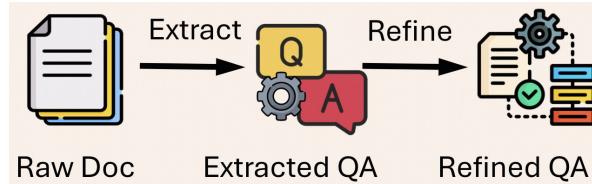
 Extract

 Extracted QA Formatted QA but lacking detailed solutions

Question: How do you simplify $(u^4 v^3/(u^2 v^{-1})^4)^0$ and write it using only positive exponents?

Answer: Explanation: Anything to the 0th power is just simply 1.

Rewrite 



 Rewritten QA Formatted QA augmented with detailed solutions

Question:
How do you simplify $(u^4 v^3/(u^2 v^{-1})^4)^0$ and write it using only positive exponents?

Answer:
To simplify the expression $(u^4 v^3/(u^2 v^{-1})^4)^0$ and rewrite it using only positive exponents, we start by evaluating the expression from the innermost operation outward.
First, consider the exponent of zero on the entire fraction: Any expression raised to the power of zero is equal to 1 (provided the base is not zero). Therefore, $(u^4 v^3/(u^2 v^{-1})^4)^0 = 1$
This simplification makes the other calculations unnecessary because raising any non-zero expression to the power of zero will always result in 1.
So, the simplified expression is simply: 1

Approaches to synthetic data creation

Sampling-based generation

Back-translation

Transformation of existing data

Human-AI collaboration

Symbolic generation

Approaches to synthetic data creation

Sampling-based generation

Back-translation

Transformation of existing data

Human-AI collaboration

Symbolic generation

Human-AI collaboration

LMs are creative & diverse, but not reliably correct

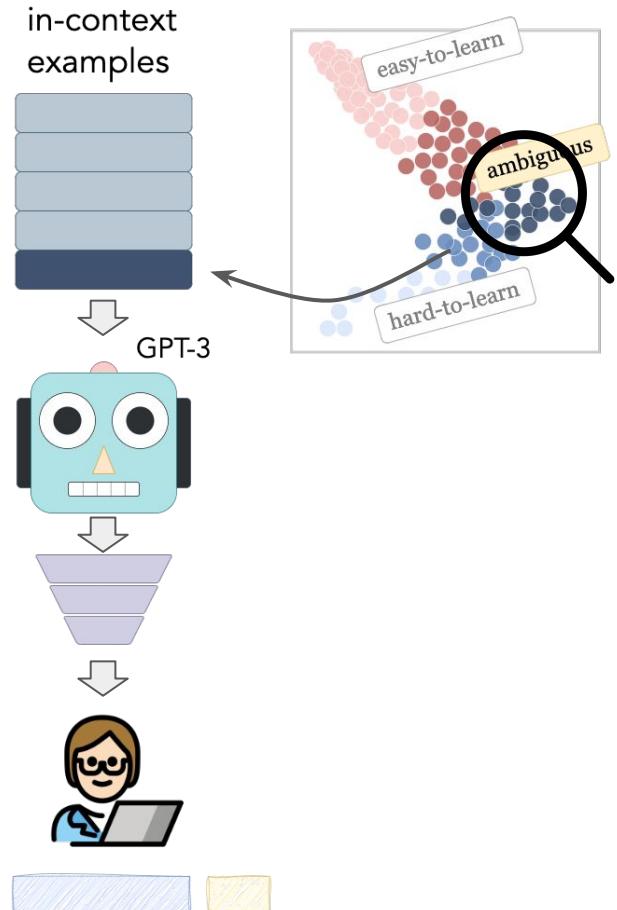
Humans can verify & improve correctness, but are not good at enumerating what they know

Combine the best of both worlds for data creation!

Human-AI collaboration for NLI

Crowdworkers revise & label generated data

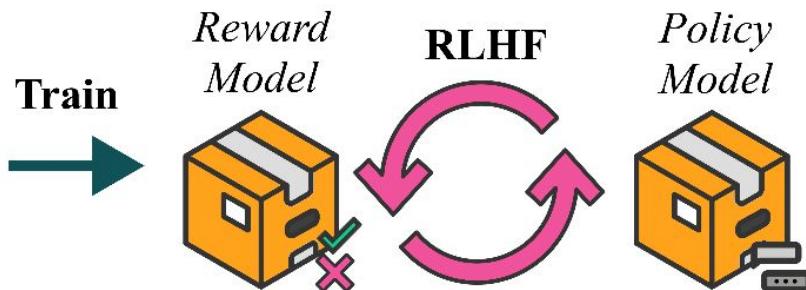
Turns writing task into editing task!



Route instances for human vs. AI feedback

Use router to solicit preference annotation from either human or LM

Prompt	A	B	Annotator
$x^{(1)}$	$y_1^{(1)}$	$y_2^{(1)}$	
$x^{(2)}$	$y_1^{(2)}$	$y_2^{(2)}$	
\vdots	\vdots	\vdots	\vdots
$x^{(n)}$	$y_1^{(n)}$	$y_2^{(n)}$	



Approaches to synthetic data creation

Sampling-based generation

Back-translation

Transformation of existing data

Human-AI collaboration

Symbolic generation

Approaches to synthetic data creation

Sampling-based generation

Back-translation

Transformation of existing data

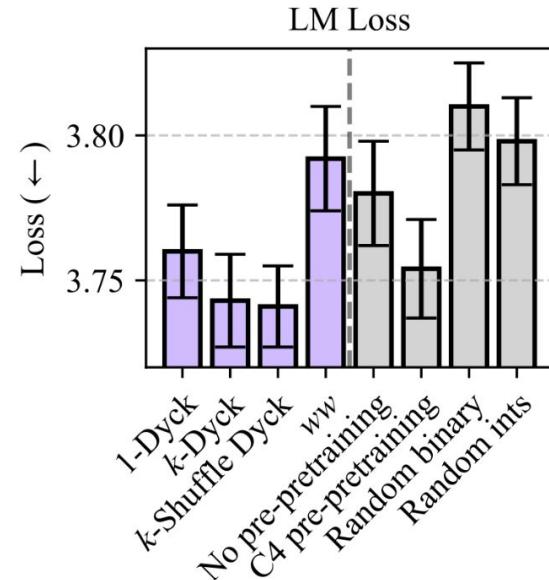
Human-AI collaboration

Symbolic generation

Symbolic generation

Doing initial pretraining on *formal languages* can lead to faster LM training and better generalization

Language	Example
1-Dyck	((()))
k -Dyck	([{}])
k -Shuffle Dyck	([{}])
ww	1 2 3 1 2 3



Summary

Sampling-based generation: Generate examples from scratch from LMs

Back-translation: Given an output, generate an input

Transformation of existing data: Transform existing data into examples of the desired task

Human-AI collaboration: Mix LM generation & human annotation

Symbolic generation: Rule-based generation

Approaches to data filtering

Diversity filtering

Difficulty filtering

Quality filtering

Correctness filtering

Diversity filtering: n-grams

Diversity filtering: gradients

Difficulty filtering: dataset cartography

Quality filtering: reward models

Correctness filtering: verifiable rewards

Given math & code problems and their answers, determine reward based on whether the answer is correct

How can we use synthetic data: Algorithms

How is synthetic data used?

Supporting fundamental language modeling algorithms

Supporting scenario-specific, end-user applications

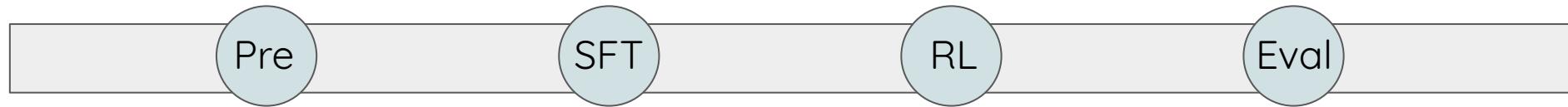
How is synthetic data used?

Supporting fundamental language modeling algorithms

Supporting scenario-specific, end-user applications

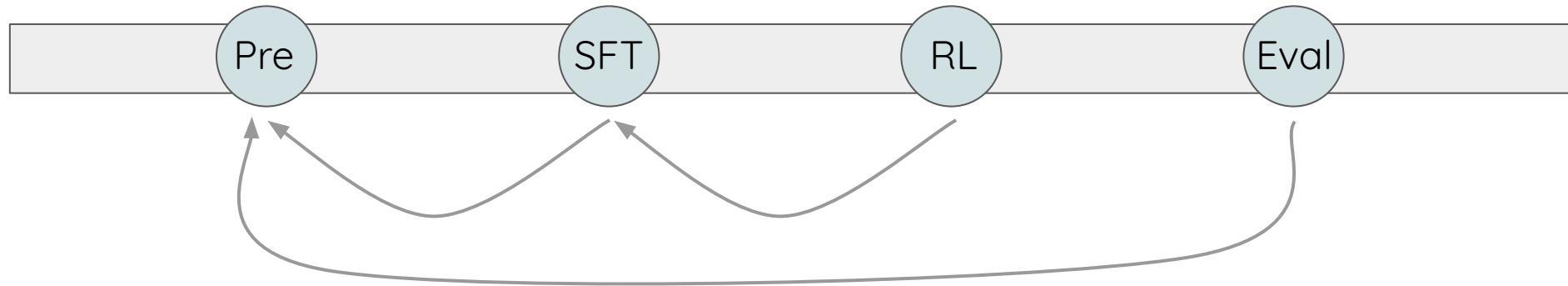
The language modeling pipeline (simplified)

Pretraining Supervised Finetuning RL Training Evaluation & Analysis

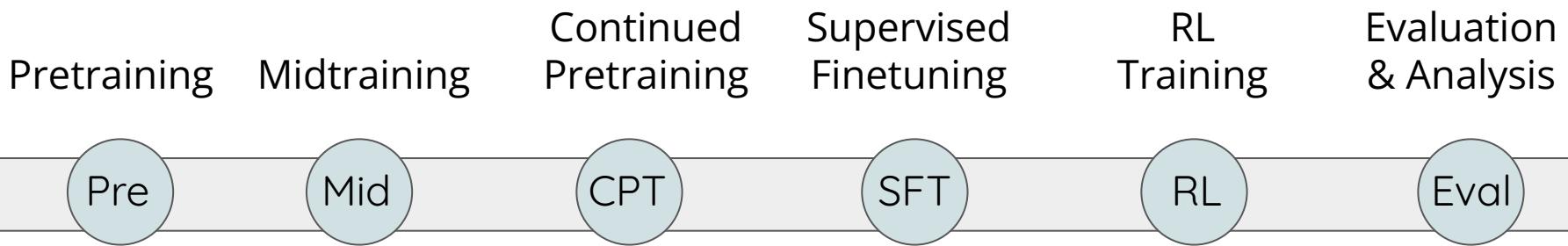


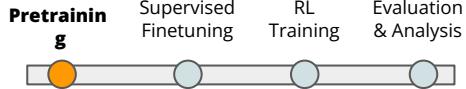
The language modeling pipeline (complex)

Pretraining Supervised Finetuning RL Training Evaluation & Analysis



The language modeling pipeline (complex)





Synthetic Data for Pretraining



From Ilya Sustkover's talk at NeurIPS 2024

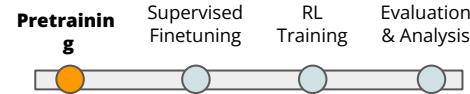
Pre-training as we know it will end

Compute is growing:

- Better hardware
- Better algorithms
- Larger clusters

Data is not growing:

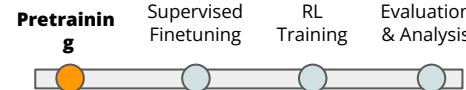
- We have but one internet
- **The fossil fuel of AI**



Does synthetic pretraining make sense?

Problem:

- Pretraining lets models learn linguistic patterns and facts
- Synthetic data generators can't invent *new linguistic patterns* or *real facts* they weren't trained on



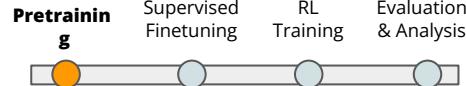
Does synthetic pretraining make sense?

Problem:

- Pretraining lets models learn linguistic patterns and facts
- Synthetic data generators can't invent *new linguistic patterns* or *real facts* they weren't trained on

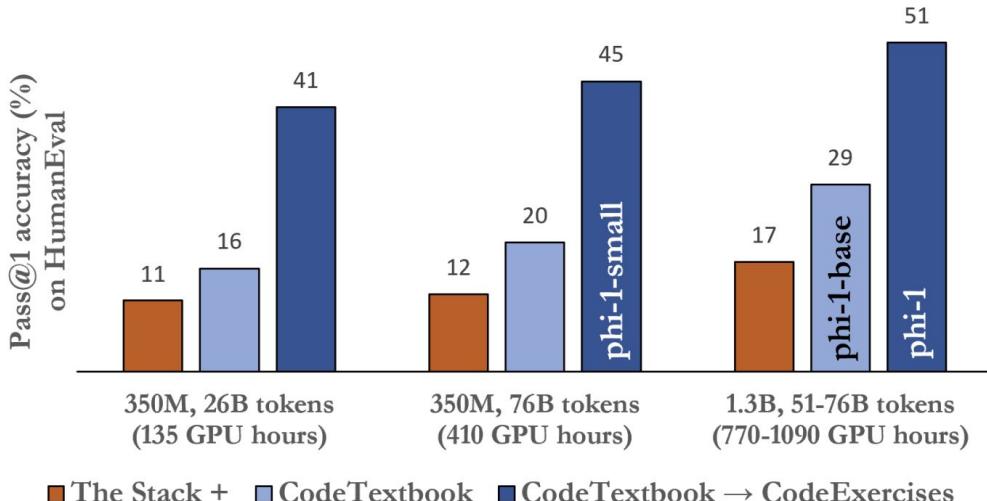
Solutions:

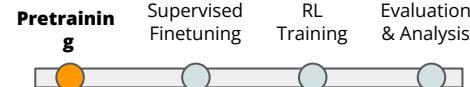
- Rephrase existing text
- Verbalize knowledge bases using LMs
- Generate text without using LMs (e.g. formal languages)



Synthetic Data for Pretraining: formatting

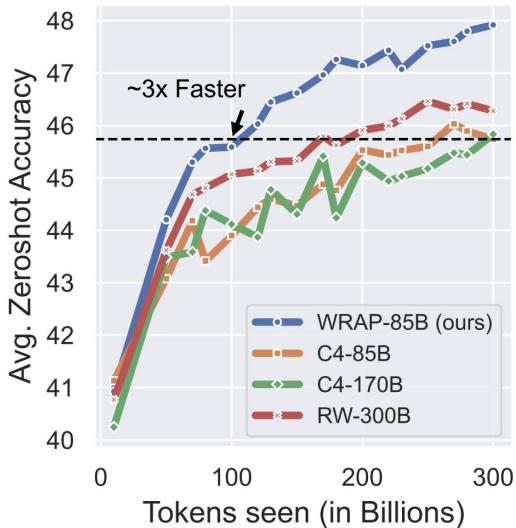
- Synthetically generated pretraining data (1B) can improve over a large amount of scraped data (6B)

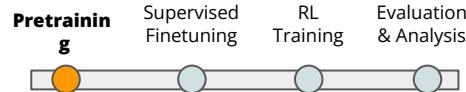




Synthetic Data for Pretraining: formatting

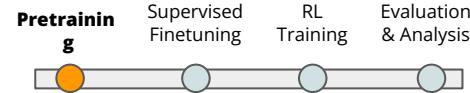
- It seems to be easier to learn from scraped data reworded and cleaned by an LLM than on the same scraped data





Synthetic Data for Pretraining: inductive biases

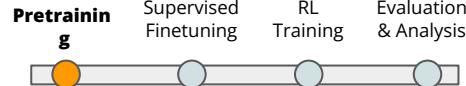
- Doing initial pretraining on *formal languages* can lead to faster LM training and better generalization
 - Learning *language* without learning *knowledge*



Synthetic Data for Pretraining: inductive biases

- Doing initial pretraining on *formal languages* can lead to faster LM training and better generalization
 - Learning *language* without learning *knowledge*

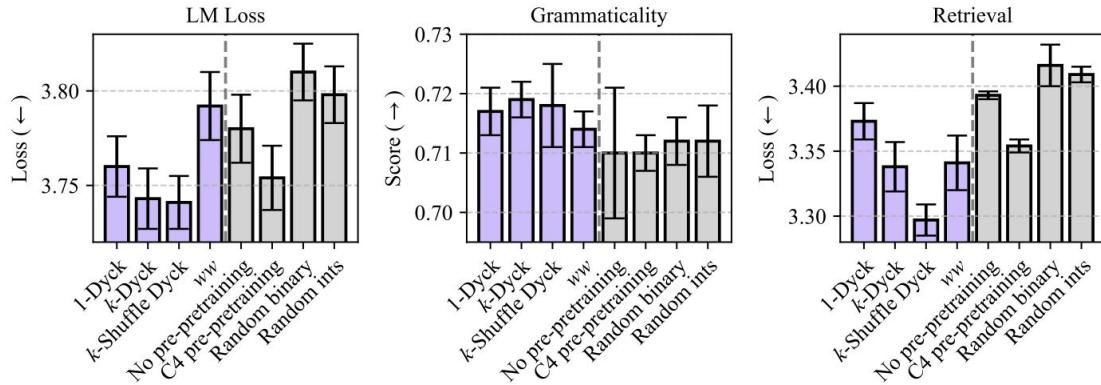
Language	Example
1-Dyck	((()))
k -Dyck	([{}])
k -Shuffle Dyck	([{}])
ww	123123

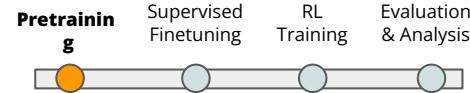


Synthetic Data for Pretraining: inductive biases

- Doing initial pretraining on *formal languages* can lead to faster LM training and better generalization
 - Learning *language* without learning *knowledge*

Language	Example
1-Dyck	((()))
k -Dyck	([{}])
k -Shuffle Dyck	([[{}]])
ww	1 2 3 1 2 3



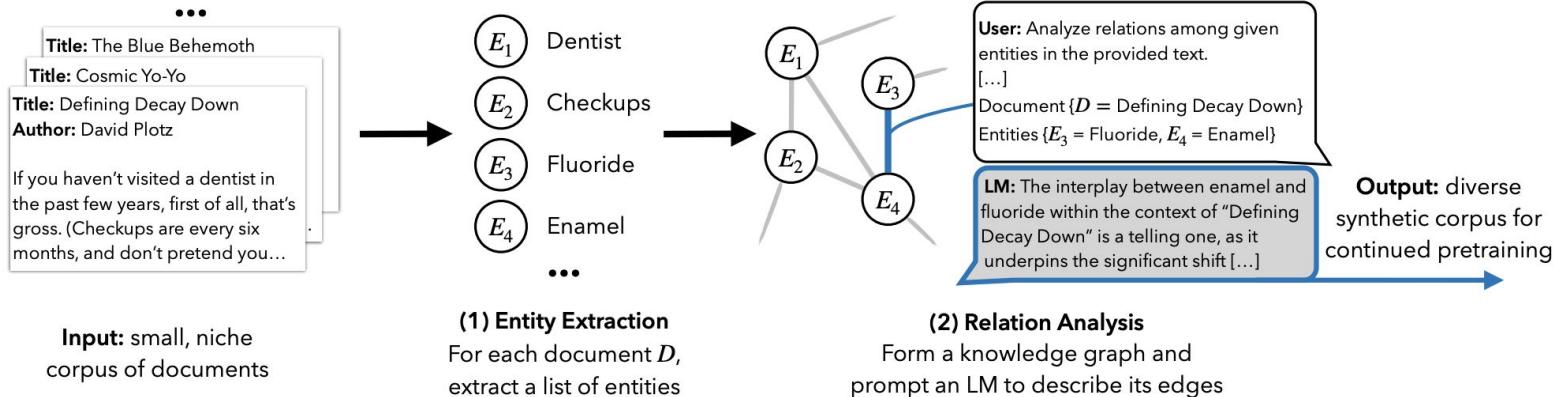


Data for Pretraining: domain adaptation

- Continuing pretraining of an LM on in-domain data is known to improve performance in the target domain
- This requires having abundant in-domain data

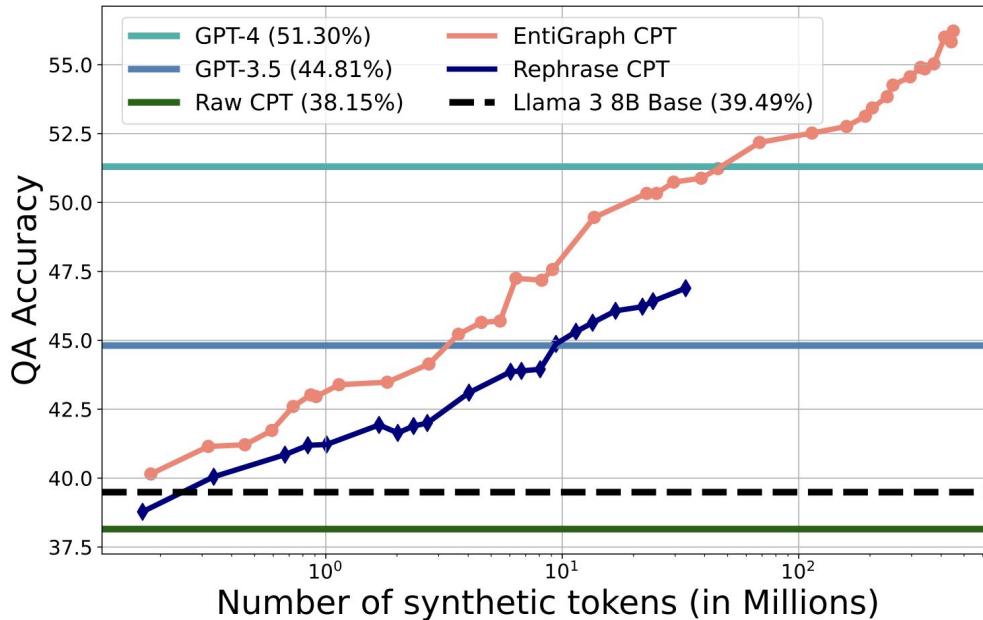


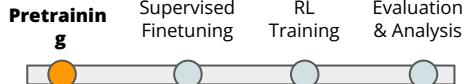
Synthetic Data for Pretraining: domain adaptation



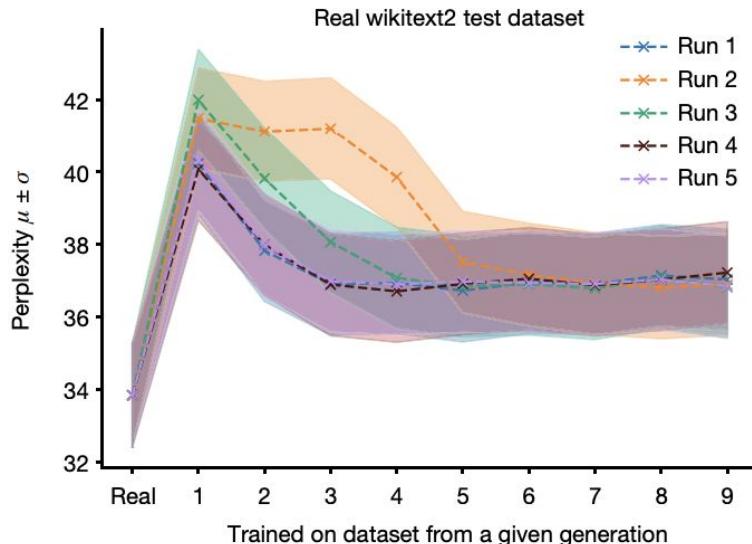
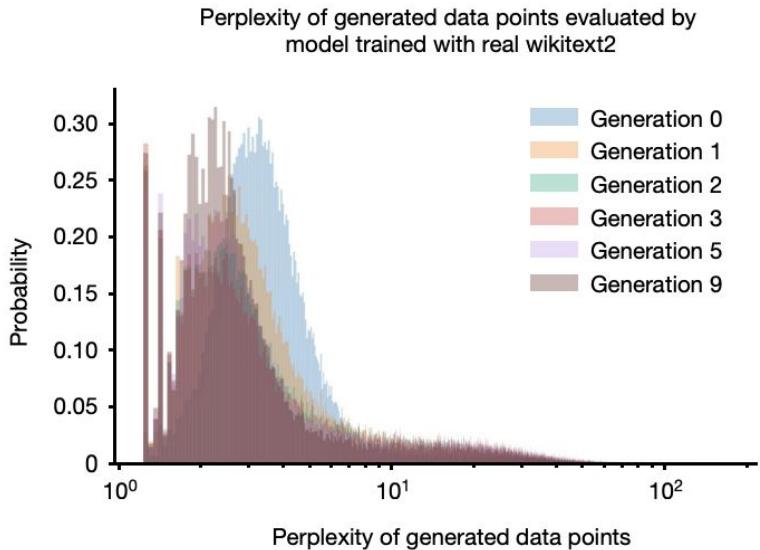


Synthetic Data for Pretraining: domain adaptation





Synthetic Data for Pretraining: risks





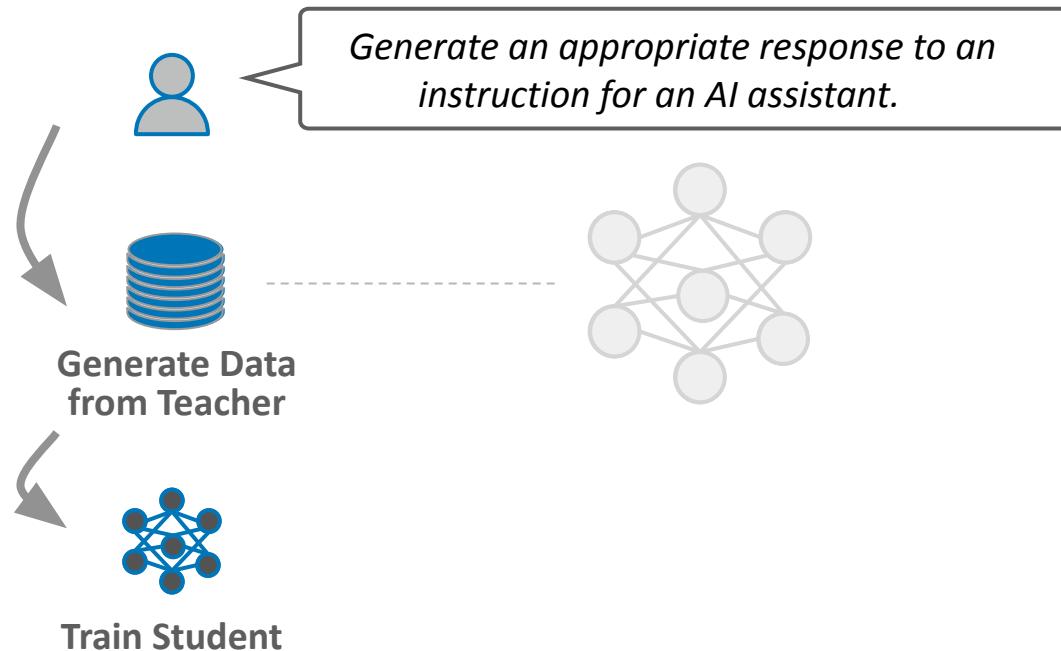
Supervised Finetuning

Goals of SFT:

- Control the style of the model's output
- Specialize behavior for a particular use-case
- Feed new information to the model

Supervised Finetuning: Distillation

If a better model is available, you can train a *student* model to imitate the *teacher*



Supervised Finetuning: Distillation

If a better model is available, you can train a *student* model to imitate the *teacher*

Pros

- Requires less human effort than manual annotation

Cons

- The student's performance is bounded by the teacher
- Legal issues, e.g. no-distillation clauses in terms of service

Supervised Finetuning: Limitations of Distillation

Goals of SFT:

- Control the style of the model's output
- Specialize behavior for a particular use-case
- Feed new information to the model

Observation: much of the benefit of distillation comes from adopting the teacher's style

- This does not require teaching the model new information

Supervised Finetuning: Self-Guide

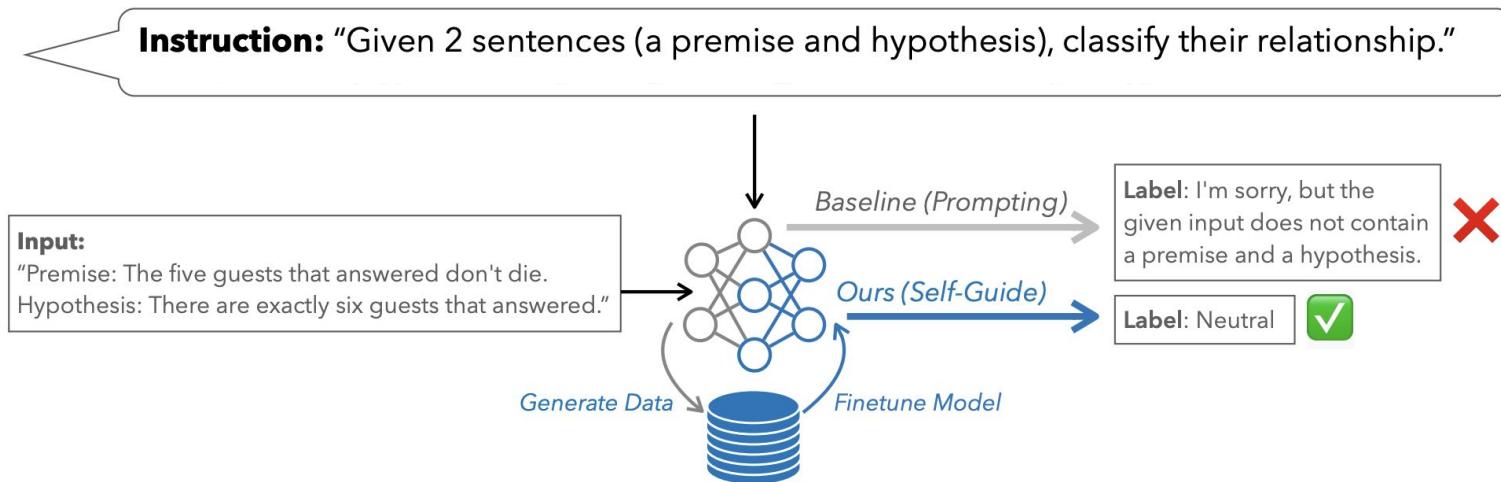


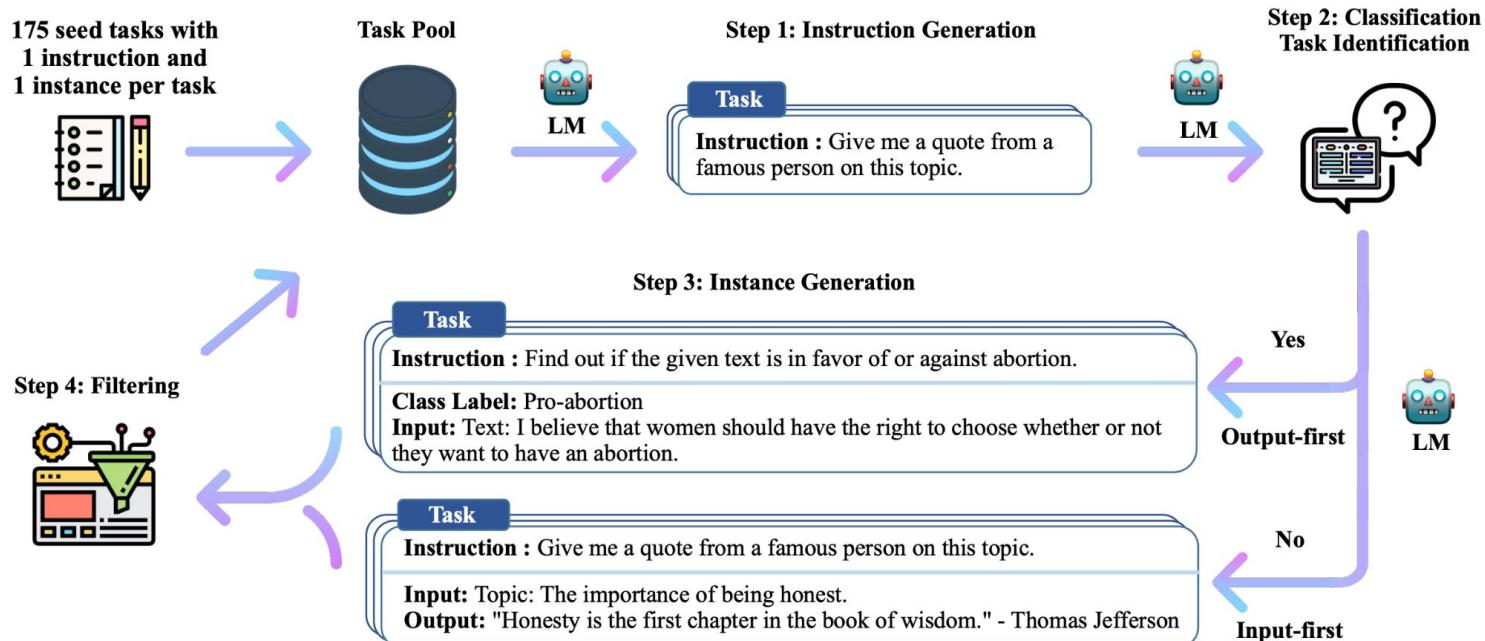
Figure 1: SELF-GUIDE uses a model’s ability to generate synthetic data as a vehicle to improve the model’s ability to execute a task as specified by an instruction.



Supervised Finetuning: Self-Guide

Task ID	Task Description	Prompting (Baseline)	Few-Shot Finetuning	SELF-GUIDE (Ours)	Δ
Classification Tasks					
task1516	NLI (IMPPRES)	17.6	32.2	35.2	17.6
task1529	NLI (SciTail)	8.5	48.9	54.5	46.0
task1612	Sentiment Class. (SICK)	51.3	33.3	33.3	-18.0
task1615	NLI (SICK)	0.5	33.3	33.1	32.6
task284	Sentiment Class. (IMDB)	90.0	71.9	82.2	-7.8
task329	Coreferent Class.	29.1	45.4	44.7	15.6
task346	Word POS Class.	35.1	49.9	50.7	15.6
Avg	Metric: Exact Match	33.2	45.0	47.7	14.5
Generation Tasks					
task1345	Question Paraphrasing	40.7	36.0	50.5	9.8
task281	Find Common Entity	46.8	40.7	49.3	2.5
task1562	Question Paraphrasing	29.5	48.6	59.3	29.8
task1622	Fluency Correction	49.2	86.2	78.5	29.3
Avg	Metric: ROUGE-L	41.6	52.9	59.4	17.9

Supervised Finetuning: Self-Instruct





Supervised Finetuning: Self-Instruct

- Originally developed to instruction-tune GPT-3 using itself, to nearly match InstructGPT

Model	# Params	ROUGE-L
GPT3 _{SELF-INST} (Ours)	175B	39.9
InstructGPT ₀₀₁	175B	40.8



Supervised Finetuning: *Self-Instruct*

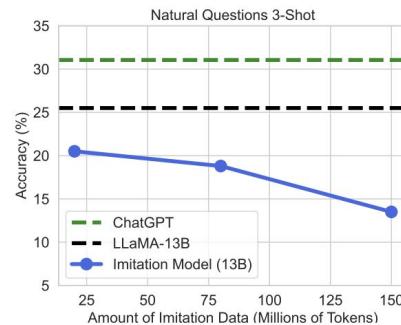
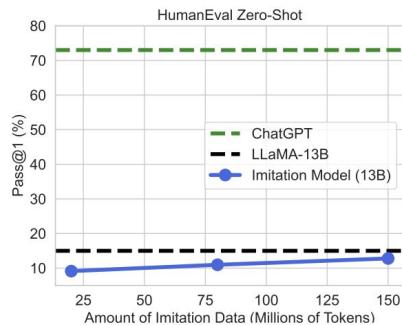
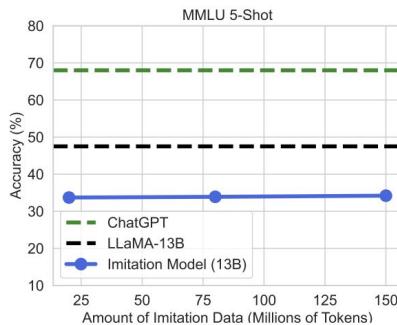
- Unsurprisingly, the same algorithm can also benefit from using a bigger teacher when available
 - *Alpaca* used this algorithm with GPT-3 as the data generator



Supervised Finetuning: *Self-Instruct*

- With sampling-based distillation, students mostly learn only surface-level behaviors (style, toxicity, refusals)

Increasing Amount of Imitation Data





Supervised Finetuning: Distillation

- To learn nontrivial capabilities via *sampling-based generation*, complex sampling strategies are needed

Supervised Finetuning: Distillation

- To learn nontrivial capabilities via *sampling-based generation*, complex sampling strategies are needed
- Example 1: *Evol-Instruct*
 - Start with *Alpaca* data
 - Randomly sample constraint types to add these instructions to add complexity
 - For each base instruction, generate analogies in other domains
 - Filter out problematic instructions



Supervised Finetuning: Distillation

- To learn nontrivial capabilities via *sampling-based generation*, complex sampling strategies are needed
- Example 2: *Orca*
 - Generate 5 million SFT examples from GPT 3.5 and GPT 4 with synthetic reasoning traces
 - Use curriculum learning
 - Randomly sample different system prompts

Supervised Finetuning: Beyond Distillation

- Simpler approach: *transformation of existing data* (e.g. “MAmmoTH2”)

Raw Docs Unformatted Text, Site Information, Ads

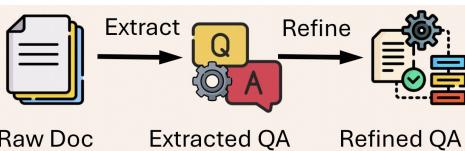
Topics Science\nAnatomy&Physiology\nAstronomy\nAstrophysics
\nBiology\nChemistry\n...Socratic Meta...Featured Answers
How do you simplify $((u^4 v^3)/(u^2 v^{-1})^4)^0$ and write it using only positive exponents?
Answer by NickTheTurtle (Apr 1, 2017)
Explanation:\nAnything raised to the 0^{th} power is simply 1.
\nRelated Questions\nWhat is the quotient of powers property?
\n\nHow do you simplify expressions using the quotient rule?...Impact of this question\n1274 views around the world
#Apps\niOS\nAndroid\nLinks\n[Privacy](#)\n[Terms](#)\n[Help](#)

Extract

Extracted QA Formatted QA but lacking detailed solutions

Question: How do you simplify $(u^4 v^3/(u^2 v^{-1})^4)^0$ and write it using only positive exponents?
Answer: Explanation: Anything to the 0^{th} power is just simply 1.

Rewrite



Rewritten QA Formatted QA augmented with detailed solutions

Question:
How do you simplify $(u^4 v^3/(u^2 v^{-1})^4)^0$ and write it using only positive exponents?

Answer:
To simplify the expression $(u^4 v^3/(u^2 v^{-1})^4)^0$ and rewrite it using only positive exponents, we start by evaluating the expression from the innermost operation outward.

First, consider the exponent of zero on the entire fraction: Any expression raised to the power of zero is equal to 1 (provided the base is not zero). Therefore, $(u^4 v^3/(u^2 v^{-1})^4)^0 = 1$

This simplification makes the other calculations unnecessary because raising any non-zero expression to the power of zero will always result in 1.

So, the simplified expression is simply: 1



Supervised Finetuning: Beyond Distillation

- Simpler approach: *transformation of existing data* (e.g. “MAmmoTH2”)

Supervised Finetuning: Beyond Distillation

 **Raw Docs** *Unformatted Text, Site Information, Ads*

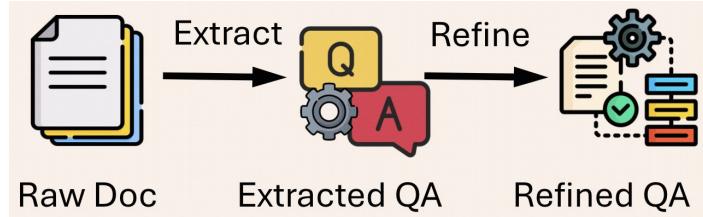
Topics Science\nAnatomy&Physiology\nAstronomy\nAstrophysics
\nBiology\nChemistry \n...Socratic Meta...Featured Answers
How do you simplify $((u^4 v^3)/(u^2 v^{-1})^4)^0$ and write it using only positive exponents?
Answer by NickTheTurtle (Apr 1, 2017)
Explanation: Anything raised to the 0^{th} power is simply 1.
\nRelated Questions\nWhat is the quotient of powers property?
\n\nHow do you simplify expressions using the quotient rule?...Impact of this question\n1274 views around the world
#Apps\niOS\nAndroid\nLinks\n[Privacy](#)\n[Terms](#)\n[Help](#)

 **Extract**

 **Extracted QA** *Formatted QA but lacking detailed solutions*

Question: How do you simplify $((u^4 v^3)/(u^2 v^{-1})^4)^0$ and write it using only positive exponents?
Answer: Explanation: Anything to the 0th power is just simply 1.

 **Rewrite**

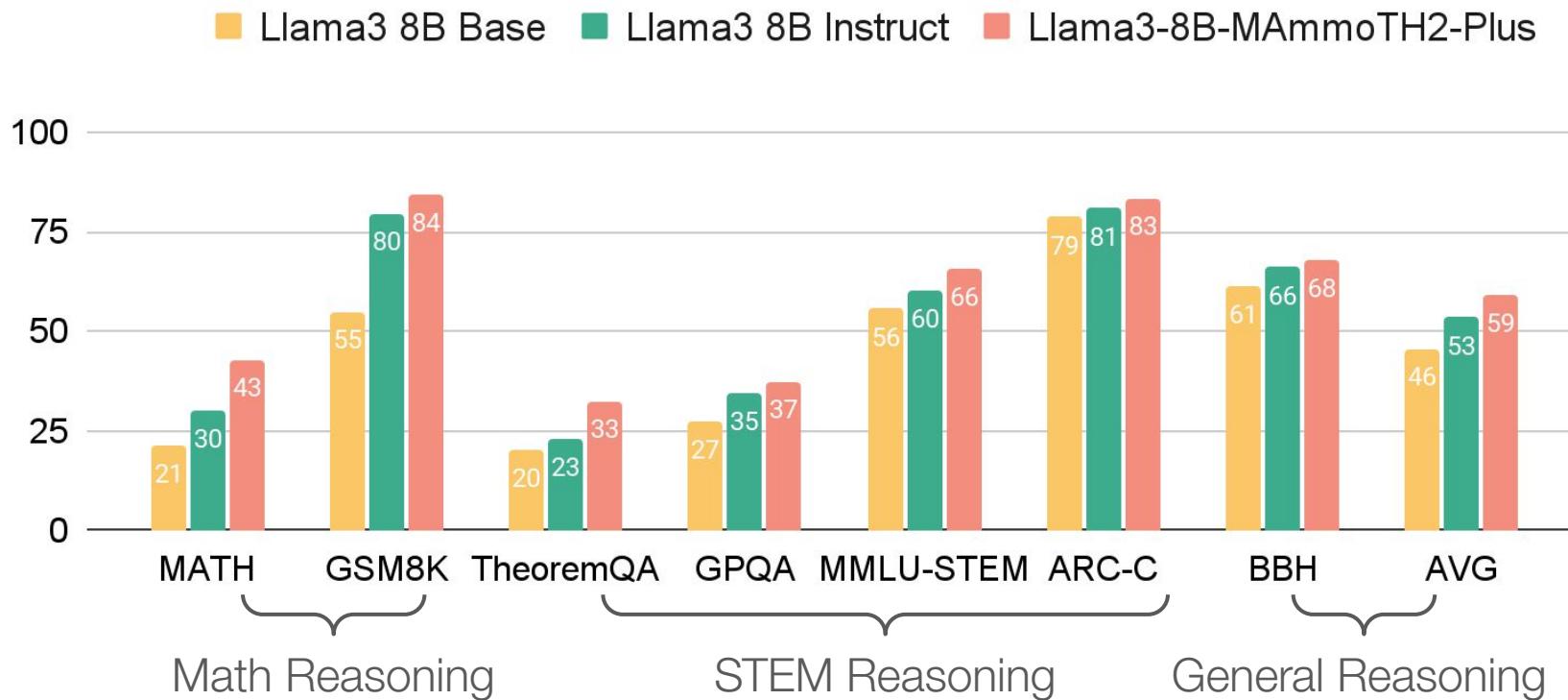


 **Rewritten QA** *Formatted QA augmented with detailed solutions*

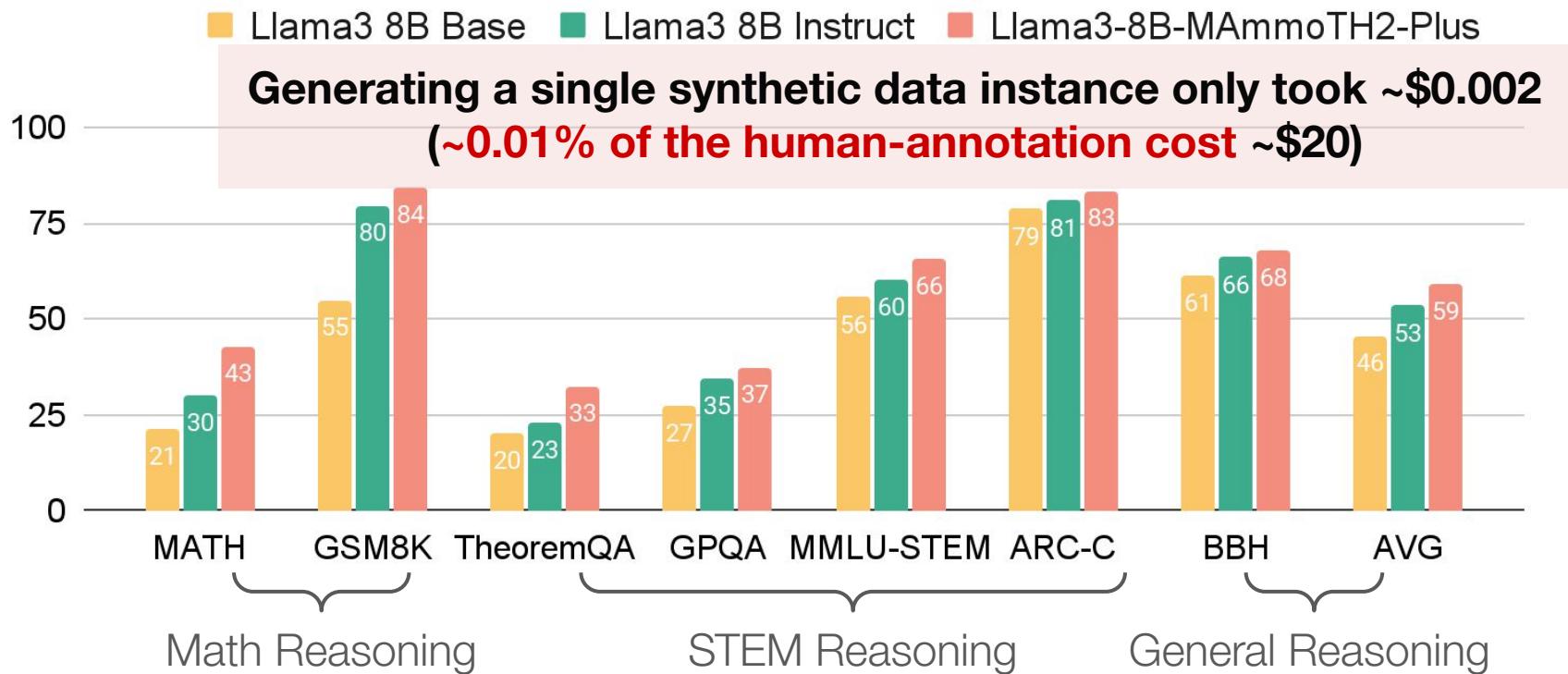
Question:
How do you simplify $(u^4 v^3/(u^2 v^{-1})^4)^0$ and write it using only positive exponents?

Answer:
To simplify the expression $(u^4 v^3/(u^2 v^{-1})^4)^0$ and rewrite it using only positive exponents, we start by evaluating the expression from the innermost operation outward.
First, consider the exponent of zero on the entire fraction: Any expression raised to the power of zero is equal to 1 (provided the base is not zero). Therefore, $(u^4 v^3/(u^2 v^{-1})^4)^0 = 1$
This simplification makes the other calculations unnecessary because raising any non-zero expression to the power of zero will always result in 1.
So, the simplified expression is simply: 1

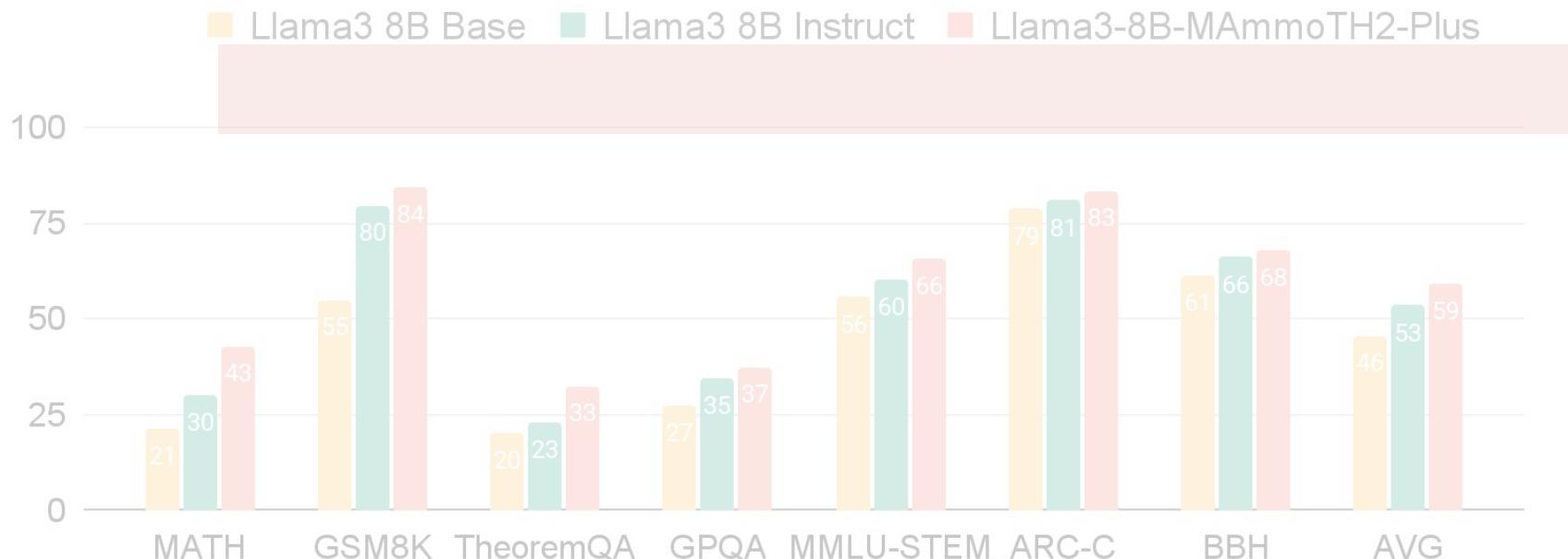
MAMMOTH2 outperforms Llama3-Instruct on science/math/reasoning



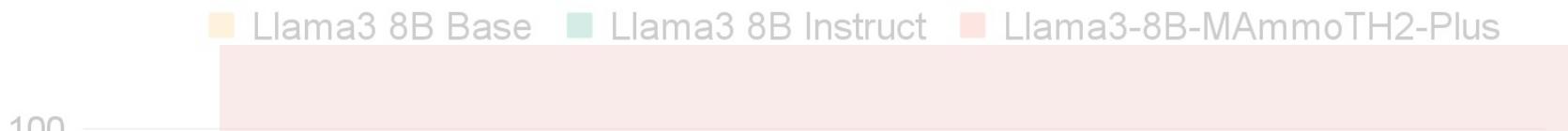
MAMMOTH2 outperforms Llama3-Instruct on science/math/reasoning



MAMMOTH2 outperforms Llama3-Instruct on science/math/reasoning



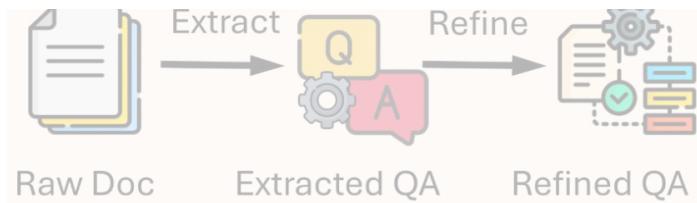
Supervised Finetuning: Beyond Distillation



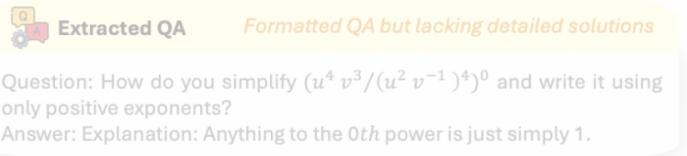
- *Takeaway:* Want to train on diverse tasks and facts?
 - Retrieve these directly instead of using hacks to coax them out of a proprietary LM



Supervised Finetuning: Beyond Distillation



- *Takeaway:* Want to train on diverse tasks and facts?
 - Retrieve these directly instead of using hacks to coax them out of a proprietary LM



the base is not zero). Therefore, $(u^4 v^3/(u^2 v^{-1})^4)^0 = 1$. This simplification makes the other calculations unnecessary because raising any non-zero expression to the power of zero will always result in 1. So, the simplified expression is simply: 1

Reinforcement Learning

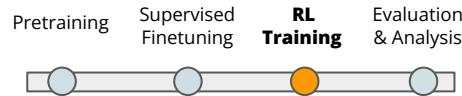
Goals of RL (non-exhaustive):

- Learn from minimal supervision
- Learn from negative examples (e.g. harmful behavior)
- Adapt models to their own token distribution rather than text written by others (“exposure bias”)

RL from Synthetic Feedback

Intuition: it is often easier to *verify* that an utterance is good than to generate that *utterance*

- “The Generator-Verifier Gap”



RL from Synthetic Feedback

- Synthetic feedback is standard in RL
- For example, *reward models* are literally generators of synthetic rewards



RL from Synthetic Feedback

- Synthetic feedback is standard in RL
- For example, *reward models* are literally generators of synthetic rewards
 - Train a model to imitate human preferences
 - Then, use the model to grade sampled responses during RL

RL from Synthetic Feedback

- Synthetic feedback is standard in RL
- For example, *reward models* are literally generators of synthetic rewards
 - Train a model to imitate human preferences
 - Typically using the *Bradley-Terry model* to learn to generate continuous scores from preference pairs
 - Then, use the model to grade sampled responses during RL

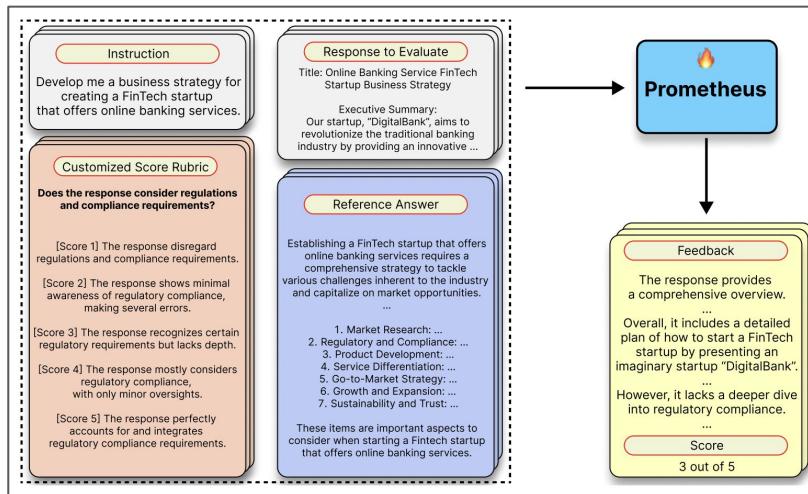


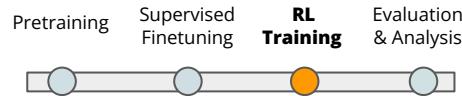
RL from Synthetic Feedback

- Synthetic feedback for RL need not come from Bradley-Terry models

RL from Synthetic Feedback

- Synthetic feedback for RL need not come from Bradley-Terry models, e.g. LLM-as-a-judge
 - A trained judge that, given a rubric, outputs scores and critiques





RL from Synthetic Feedback

- Synthetic feedback doesn't require training data at all
 - You can just prompt a judge model to rate a response from 1 to 5 on a criterion (e.g. UltraFeedback)

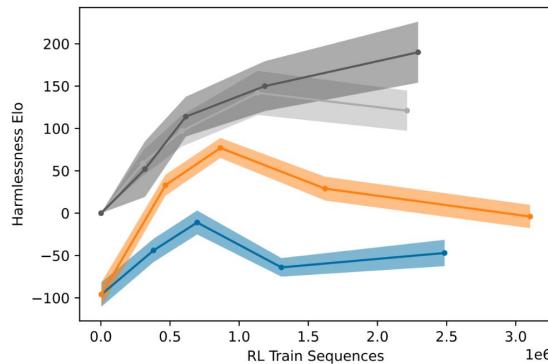
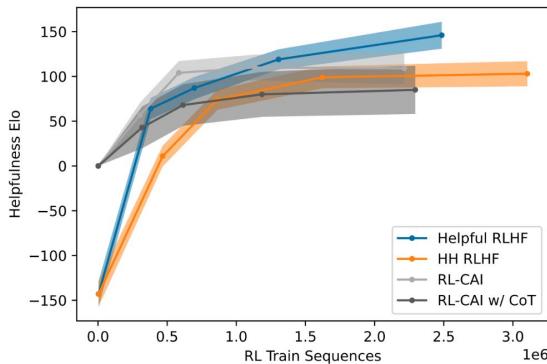
RL from Synthetic Feedback

- Synthetic feedback doesn't require training data at all
 - Or, with preference optimization, ask the judge to choose which of a pair of responses is better, according to a criterion

Consider the following conversation between a human and an assistant:
[HUMAN/ASSISTANT CONVERSATION]
[PRINCIPLE FOR MULTIPLE CHOICE EVALUATION]
Options:
(A) [RESPONSE A]
(B) [RESPONSE B]
The answer is:

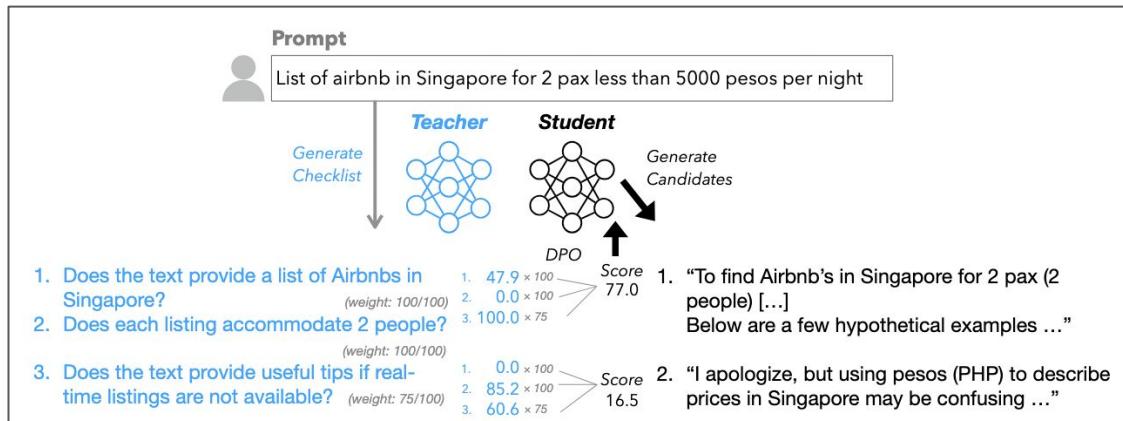
RL from Synthetic Feedback

- Synthetic feedback doesn't require training data at all
 - Or, with preference optimization, ask the judge to choose which of a pair of responses is better, according to a criterion
 - The set of criteria can become complex (e.g. 16 criteria in “Constitutional AI”)



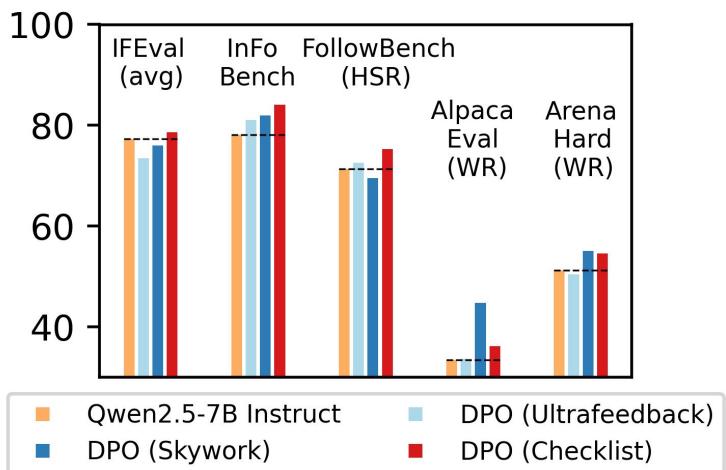
RL with Feedback from Flexible Criteria

- Can we list all possible aspects of response quality?
- Solution: generate *checklists* for each prompt



RL with Feedback from Flexible Criteria

- Solution: generate *checklists* for each prompt
- Better than reward models or judges at teaching instruction-following



Evaluation of Synthetic Feedback

How do you know what kind of judge to use?
Does it even matter?

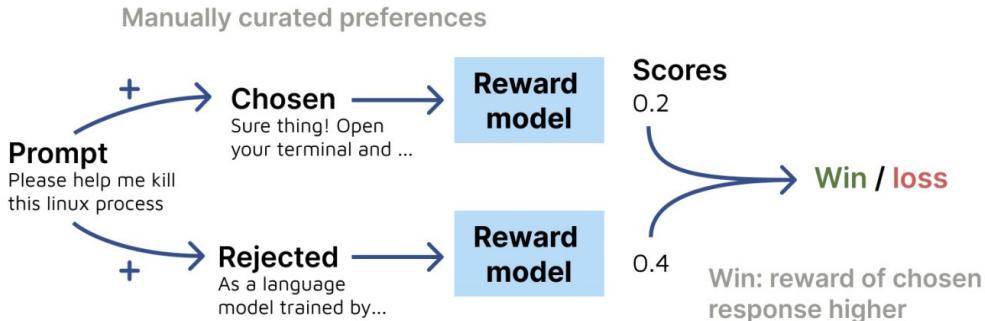
Evaluation of Synthetic Feedback

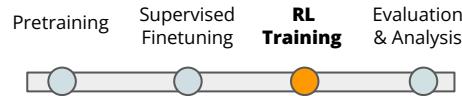
Three ways to evaluate a “judge”:

1. Agreement with human preferences (e.g. RewardBench)

2.

3.





Evaluation of Synthetic Feedback

Three ways to evaluate a “judge”:

1. Agreement with human preferences (e.g. RewardBench)
2. Agreement with generic benchmarks (via reranking)
- 3.



Evaluation of Synthetic Feedback

Three ways to evaluate a “judge”:

1. Agreement with human preferences (e.g. RewardBench)
2. Agreement with generic benchmarks (via reranking)
3.
 - Choose a benchmark you care about
 - For every question, sample 16 responses
 - Use your judge to choose the best response
 - This should improve your score on the benchmark

Evaluation of Synthetic Feedback

Three ways to evaluate a “judge”:

1. Agreement with human preferences (e.g. RewardBench)
2. Agreement with generic benchmarks (via reranking)
- 3.** Effectiveness in RL pipelines

Evaluation of Synthetic Feedback

Three ways to evaluate a “judge”:

1. Agreement with human preferences (e.g. RewardBench)
2. Agreement with generic benchmarks (via reranking)
- 3.** Effectiveness in RL pipelines
 - Choose an RL algorithm, a model, and some prompts
 - Train the model to maximize your reward
 - Evaluate it on benchmarks of your choosing

Evaluation of Synthetic Feedback

Problem: these things are not always correlated

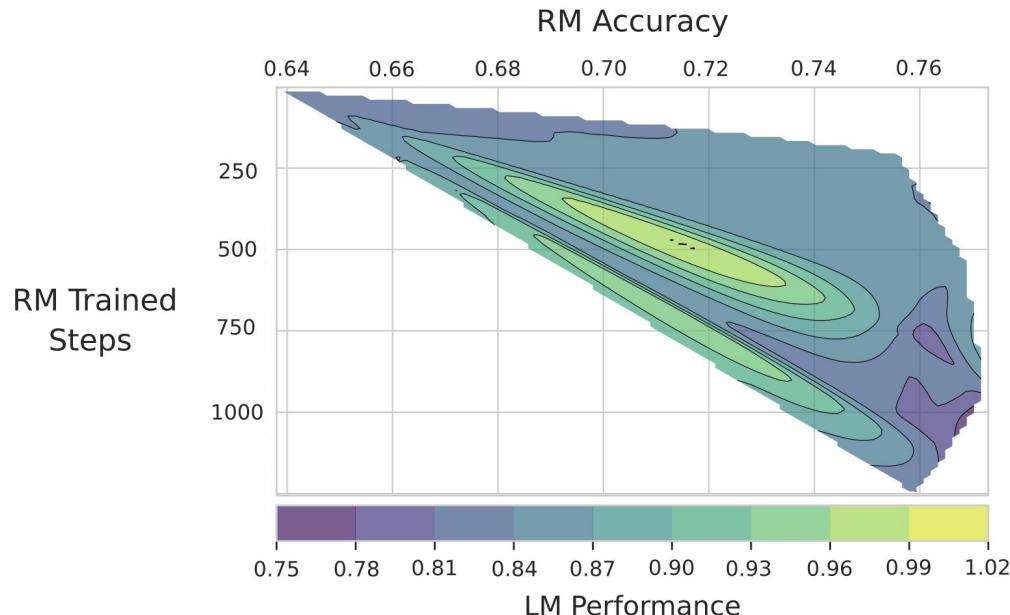
1. Agreement with human preferences (e.g. RewardBench)
2. Agreement with generic benchmarks (via reranking)
3. Effectiveness in RL pipelines

Open Problem in RL from Synthetic Feedback

- What makes a judge a good teacher for RL?

Open Problem in RL from Synthetic Feedback

- What makes a judge a good teacher for RL? Accuracy?

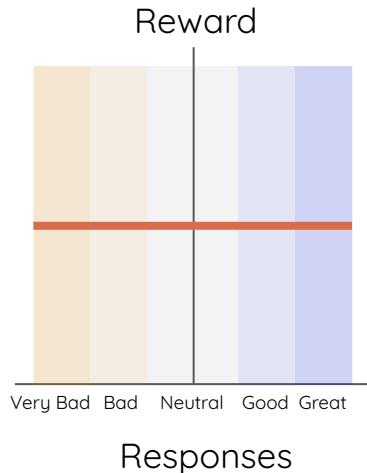




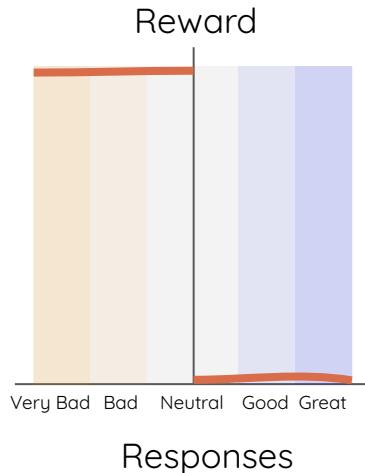
Open Problem in RL from Synthetic Feedback

- What makes a judge a good teacher for RL? Variance?

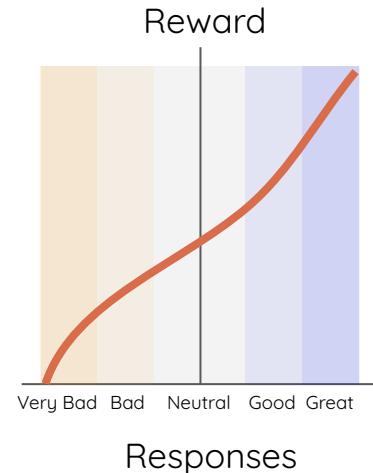
Low Acc, Low Var



Low Acc, High Var



High Acc, Low Var



High Acc, High Var





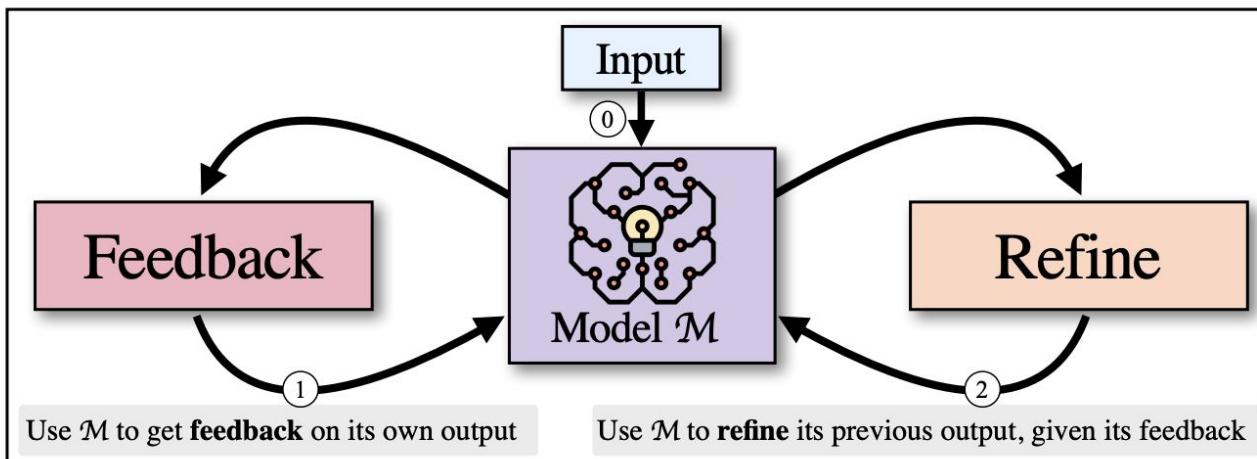
Open Problem in RL from Synthetic Feedback

- What makes a judge a good teacher for RL? ***Unclear***



Non-RL Feedback (Critiques)

- Treating synthetic feedback as a turn in a conversation



Non-RL Feedback (Critiques)

- Treating synthetic feedback as a turn in a conversation
 - This can be effective when prompting with big models

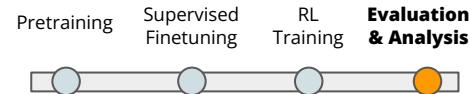
Task	GPT-3.5		GPT-4	
	Base	+SELF-REFINE	Base	+SELF-REFINE
Sentiment Reversal	8.8	30.4 (\uparrow 21.6)	3.8	36.2 (\uparrow 32.4)
Dialogue Response	36.4	63.6 (\uparrow 27.2)	25.4	74.6 (\uparrow 49.2)
Code Optimization	14.8	23.0 (\uparrow 8.2)	27.3	36.0 (\uparrow 8.7)
Code Readability	37.4	51.3 (\uparrow 13.9)	27.4	56.2 (\uparrow 28.8)
Math Reasoning	64.1	64.1 (0)	92.9	93.1 (\uparrow 0.2)
Acronym Generation	41.6	56.4 (\uparrow 14.8)	30.4	56.0 (\uparrow 25.6)
Constrained Generation	16.0	39.7 (\uparrow 23.7)	4.4	61.3 (\uparrow 56.9)



Non-RL Feedback (Critiques)

- Treating synthetic feedback as a turn in a conversation
 - This can be effective when prompting with **big** models

Task	GPT-3.5		GPT-4	
	Base	+SELF-REFINE	Base	+SELF-REFINE
Sentiment Reversal	8.8	30.4 (\uparrow 21.6)	3.8	36.2 (\uparrow 32.4)
Dialogue Response	36.4	63.6 (\uparrow 27.2)	25.4	74.6 (\uparrow 49.2)
Code Optimization	14.8	23.0 (\uparrow 8.2)	27.3	36.0 (\uparrow 8.7)
Code Readability	37.4	51.3 (\uparrow 13.9)	27.4	56.2 (\uparrow 28.8)
Math Reasoning	64.1	64.1 (0)	92.9	93.1 (\uparrow 0.2)
Acronym Generation	41.6	56.4 (\uparrow 14.8)	30.4	56.0 (\uparrow 25.6)
Constrained Generation	16.0	39.7 (\uparrow 23.7)	4.4	61.3 (\uparrow 56.9)



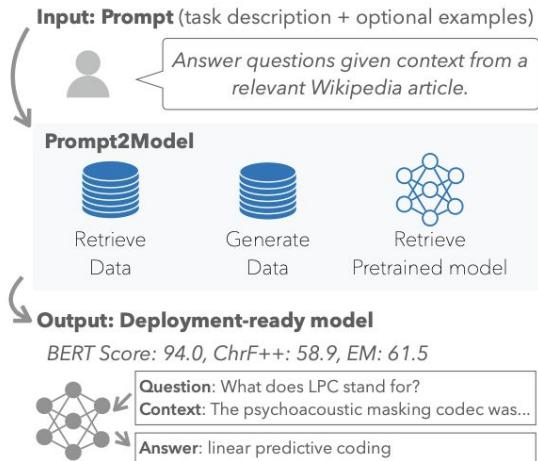
Evaluation and Analysis

Goals of Evaluation:

- Test readiness of a model for deployment
- Discover areas of improvement to make a model better
- Learn fundamental insights about language or machine learning

Generating Synthetic Eval Data

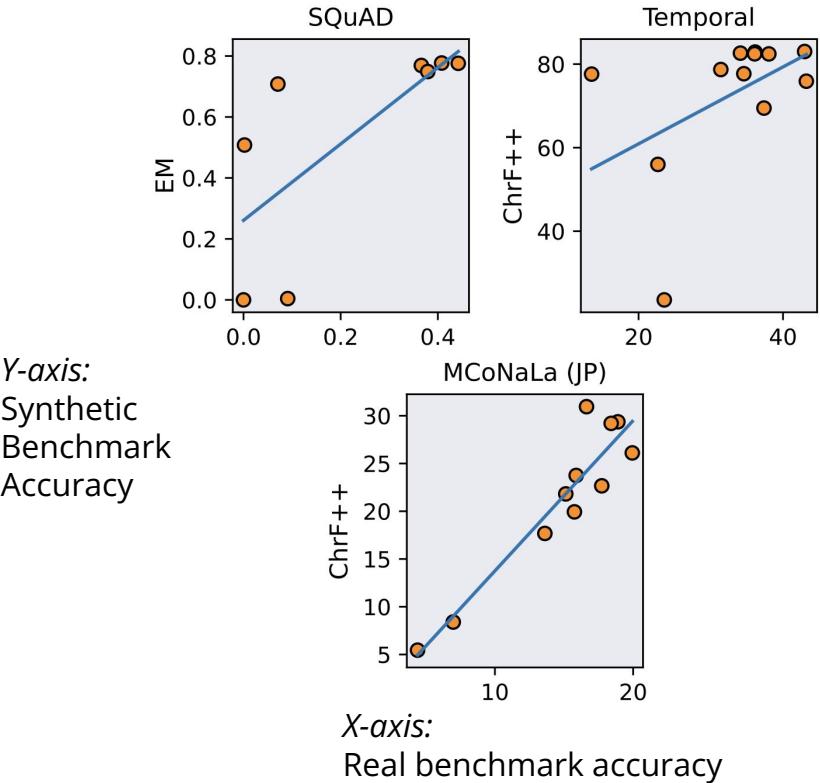
- The Prompt2Model library synthesizes training and *evaluation* data



<https://github.com/neulab/prompt2model>

Generating Synthetic Eval Data

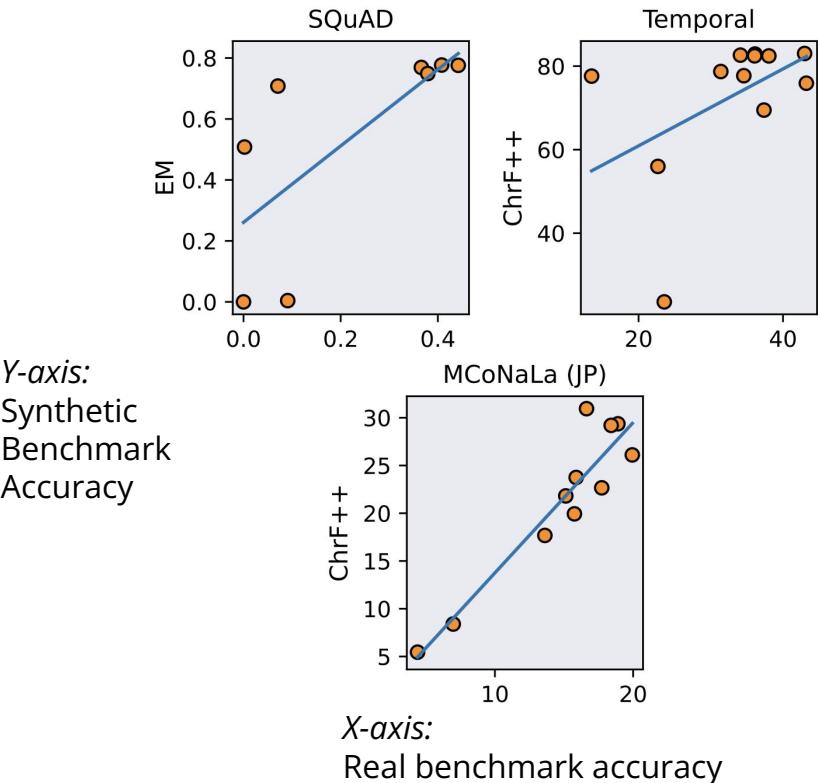
- Synthetic evaluation data consistently overestimates true performance



<https://github.com/neulab/prompt2model>

Generating Synthetic Eval Data

- Synthetic evaluation data consistently overestimates true performance



<https://github.com/neulab/prompt2model>

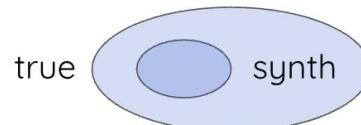
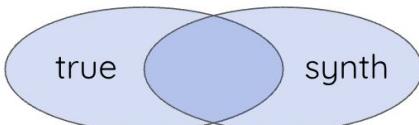
Generating Synthetic Eval Data

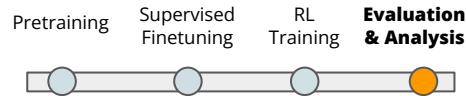
- Synthetic evaluation data consistently overestimates true performance

Recall from earlier slides

The input distribution may be **off**, or **not diverse enough**

$$P_{\text{true}}(x) \neq P_{\text{synth}}(x)$$





Generating Adversarial Eval Data

- *If you have an accurate mistake classifier:*
 1. Train a generator to generate prompts to elicit mistakes via RL

Generating Adversarial Eval Data

- *If you have an accurate mistake classifier:*
 1. Train a generator to generate prompts to elicit mistakes via RL
 2. Find inputs that lead your *target model* to make mistakes by rejection sampling via the *red model*

Generating Adversarial Eval Data

- *If you have an accurate mistake classifier:*
 1. Train a generator to generate prompts to elicit mistakes via RL
 2. Find inputs that lead your *target model* to make mistakes by rejection sampling via the *red model*
- Synthetic data generation helps cheaply obtain adversarial inputs in non-critical situations

Synthetic Data for Analyzing LMs

- Rule-based synthetic data has been used for probing and understanding models since the advent of NLP

Synthetic Data for Analyzing LMs

- Modern example: Kassner, Krojer, and Schutze (2020) generated a synthetic pretraining corpus generated by *synthetic facts* (e.g. “jupiter is big”) and logical rules (e.g. “jupiter is not small”)

Synthetic Data for Analyzing LMs

- Modern example: Kassner, Krojer, and Schutze (2020) generated a synthetic pretraining corpus generated by *synthetic facts* (e.g. “jupiter is big”) and logical rules (e.g. “jupiter is not small”)
- Pretrained BERT from scratch on this corpus
- Discovered that BERT struggles with two-hop reasoning

How is synthetic data used?

Supporting fundamental language modeling algorithms

Supporting scenario-specific, end-user applications

How is synthetic data used?

Supporting fundamental language modeling algorithms

Supporting scenario-specific, end-user applications

How can we use synthetic data: Applications

Part IV

Scenario-specific Applications



Xiang Yue

Postdoc Researcher

Carnegie Mellon University

<https://xiangyue9607.github.io/>



Table of Content

Scenario-specific Applications

- Reasoning
- Code Generation
- Tool use and Agents
- Multilingual and Multimodal
- Conclusion and Future Directions

Reasoning

Solving Complex Reasoning Problems with LLMs

OPENAI / ARTIFICIAL INTELLIGENCE / TECH

OpenAI releases o1, its first
'rea

OpenAI o1 Model Sets New Math and Understanding R1 and DeepSeek

R1 belongs to a new category of AI models known as "reasoning models," with OpenAI's o1 being the most well-known example. What makes reasoning models special is their approach to problem-solving. Rather than generating immediate responses, they employ an internal reasoning process that mirrors human trains of thought.

Image: The Verge

OpenAI is releasing a new model called o1, the first in a planned series of "reasoning" models that have been trained to answer more complex questions, faster than a human can. It's being released alongside o1 mini, a smaller, cheaper version. And yes, if you're steeped in AI rumors:

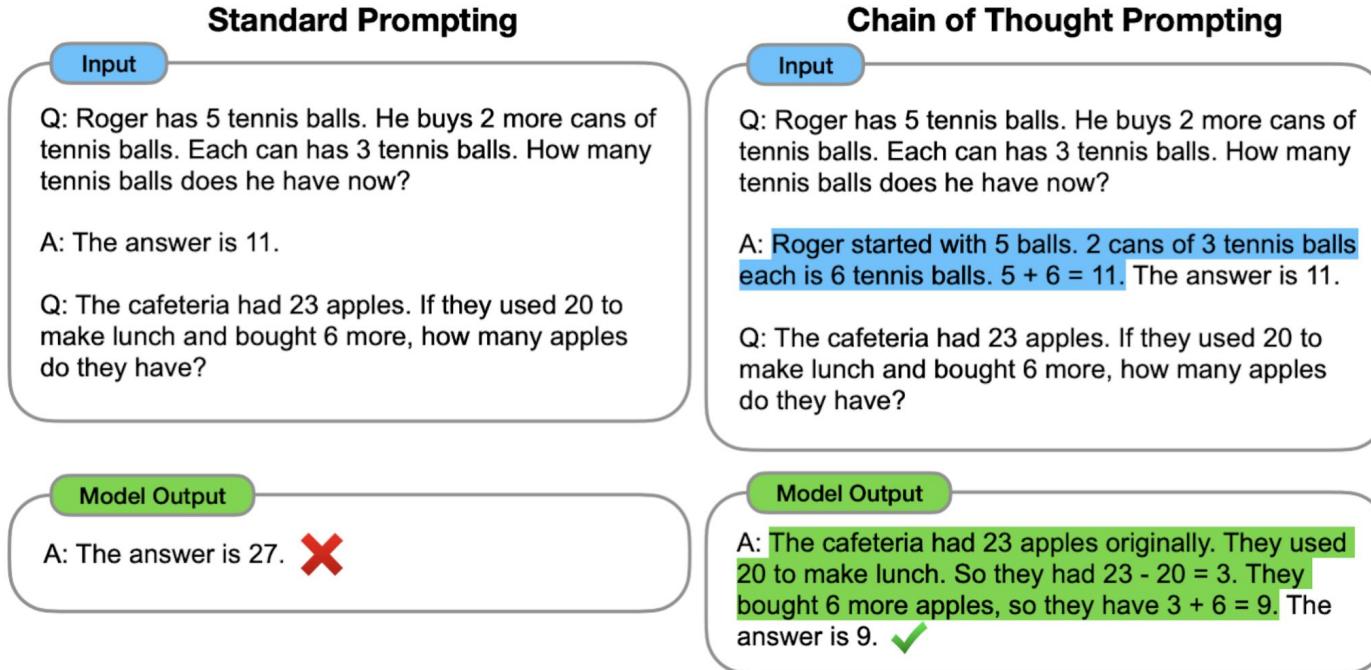


IMAGE CREDITS: PERESMEH / GETTY IMAGES

OpenAI's o1 on certain
benchmarks

Kyle Wiggers — 2:27 PM PST · January 27, 2025

Reason with Rationales: Chain-of-thought (CoT)



Wei, J. (2022). Chain-of-thought prompting elicits reasoning in large language models. NeurIPS 2022



Scaling up *Inference* Compute with Long CoT

~200
Tokens

GPT-4o

such that $1 \leq |k| \leq n$.

Find all other real numbers x for which
 $p\left(\frac{1}{x}\right) = x^2$.

We are given a monic real polynomial $p(x)$ of degree $2n$, which has the form

$$p(x) = x^{2n} + a_{2n-1}x^{2n-1} + \dots + a_1x + a_0$$

and we know that for integers k such that $1 \leq |k| \leq n$, the polynomial satisfies the condition

$$p\left(\frac{1}{k}\right) = k^2.$$

We are tasked with finding all other real

~2000
Tokens

OpenAI o1-preview

such that $1 \leq |k| \leq n$.

Find all other real numbers x for which
 $p\left(\frac{1}{x}\right) = x^2$.

Thought for 5 seconds ^

We are told that for all integer values of k satisfying $1 \leq |k| \leq n$,

$$p\left(\frac{1}{k}\right) = k^2$$

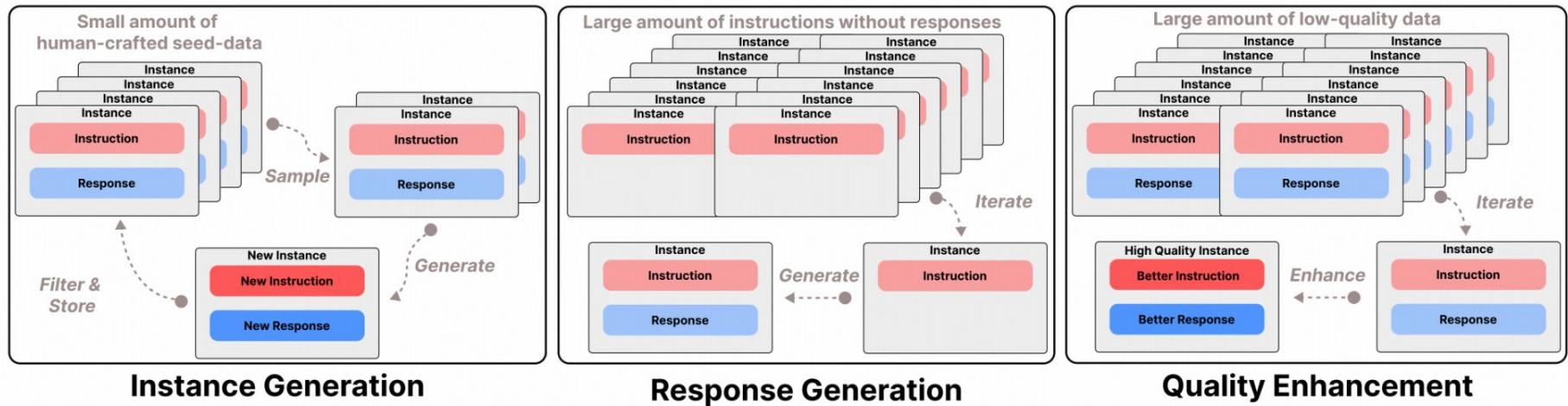
Our problem is: Given that p is monic of degree $2n$, and that $p\left(\frac{1}{k}\right) = k^2$ for $k = \pm 1, \pm 2, \dots, \pm n$, find all other real numbers x such that $p\left(\frac{1}{x}\right) = x^2$.

In other words, we are to find all real numbers x such that $p\left(\frac{1}{x}\right) = x^2$.

OpenAI o1 generates **longer CoTs** for reasoning

How do we generate synthetic reasoning data?

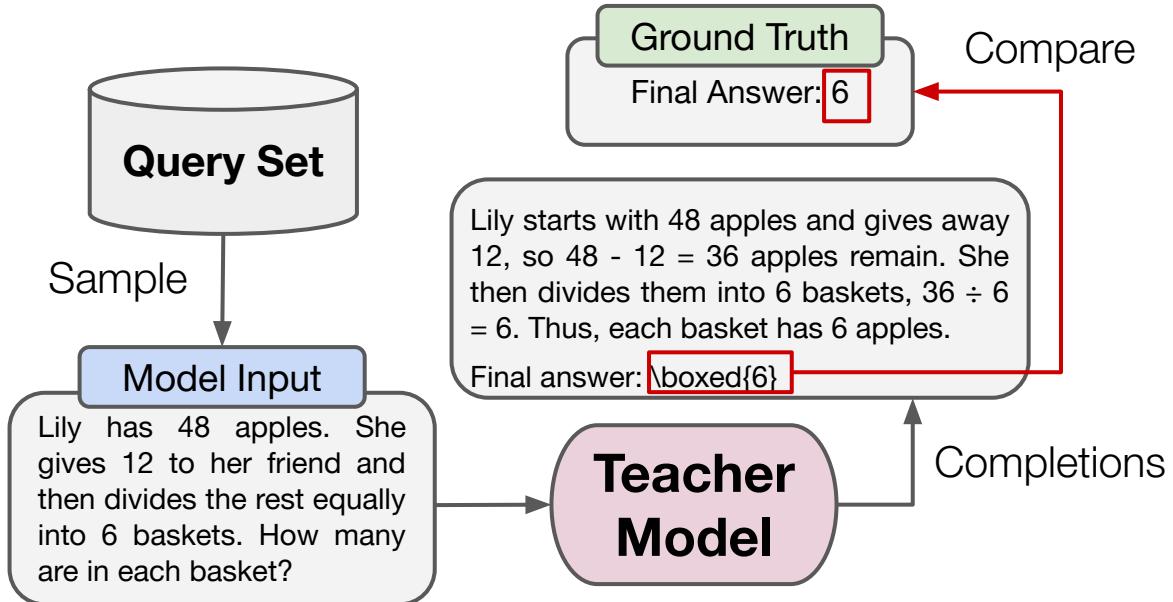
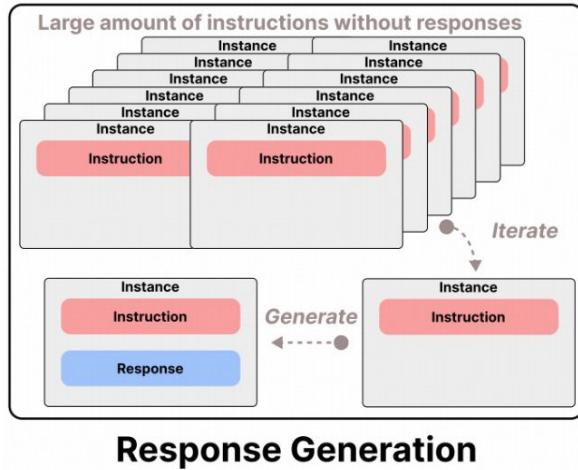
A Unified View of Synthetic Text Generation



Apply an existing LLM θ to the prompt and seed data D to produce text

$$D_s = \underbrace{\mathcal{O}(\text{Prompt}_s(D); \theta)}_{\text{LLM output}}$$

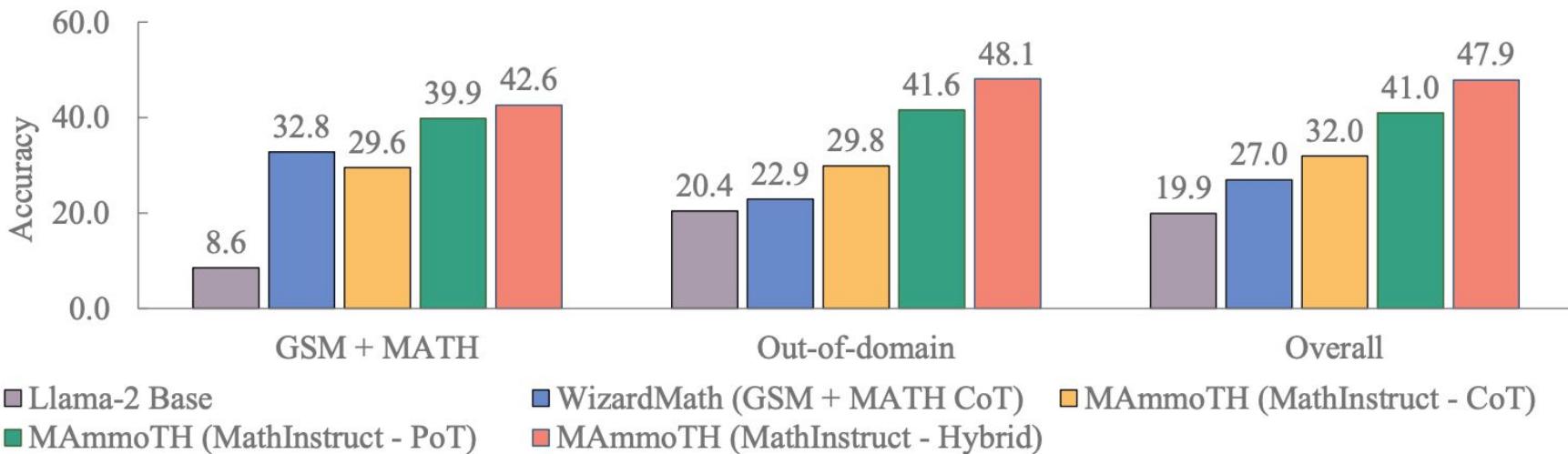
Reasoning Data: Response Generation



MAMmoTH: MathInstruct

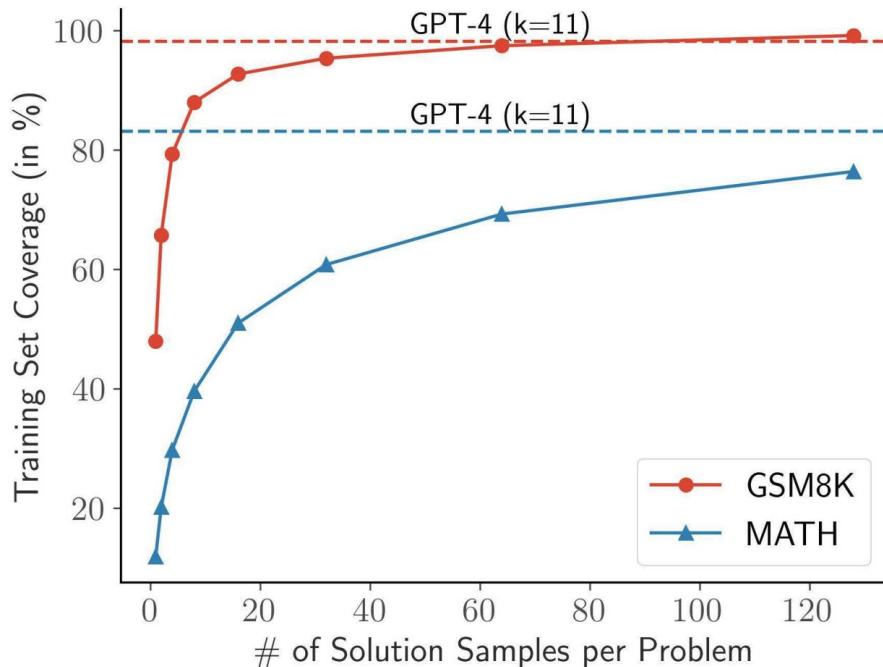
Training Dataset	Type	Annotation	# Samples	Characteristics	Fields
GSM8K (Cobbe et al., 2021)	CoT	Human	7K	Grade Schol Exam	■
GSM8K-RFT (Yuan et al., 2023)	CoT	Llama	28K	Llama + Validated	■
AQuA-RAT (Ling et al., 2017)	CoT	Human	90K	GRE/GMAT Exam	■
MATH (Hendrycks et al., 2021b)	CoT	Human	7K	Math Competition	■ ■ ■ ■ ■ ■ ■ ■
TheoremQA (Chen et al., 2023) ★	CoT	GPT-4	600	GPT4 + Validated	■ ■ ■ ■ ■ ■
Camel-Math (Li et al., 2023a)	CoT	GPT-4	50K	GPT4 (Unvalidated)	■ ■ ■ ■ ■ ■
College-Math ★	CoT	GPT-4	1.8K	GPT4 (Unvalidated)	■
GSM8K ★	PoT	GPT4	14K	GPT4 + Validated	■
AQuA-RAT ★	PoT	GPT4	9.7K	GPT4 + Validated	■
MATH ★	PoT	GPT4	7K	GPT4 + Validated	■ ■ ■ ■ ■
TheoremQA ★	PoT	GPT4	700	GPT4 + Validated	■ ■ ■ ■ ■ ■
MathQA (Amini et al., 2019)	PoT	Human	25K	AQuA-RAT Subset	■
NumGLUE (Mishra et al., 2022a)	PoT	Human	13K	Lila Annotated	■
MathInstruct			260K		■ ■ ■ ■ ■ ■ ■ ■

MAMmoTH: MathInstruct



Hybrid of thoughts improves the reasoning performance

OpenMathInstruct-1



Prompt	MATH		
	# Samples	# Unique Solns.	Coverage (in %)
Default	224	177K	80.1
+ Subj	224	191K	80.1
Mask-Text	224	192K	85.9
+ Subj	224	227K	87.5
Total	896	787K	93.0
Prompt	GSM8K		
	# Samples	# Unique Solns.	Coverage (in %)
Default	128	434K	99.1
+ Subj	-	-	-
Mask-Text	128	602K	99.9
+ Subj	-	-	-
Total	256	1036K	99.9

MetaMath

Meta-Question: James buys 5 packs of beef that are 4 pounds each. The price of beef is \$5.50 per pound. How much did he pay?

Answer: He bought $5 \times 4 = 20$ pounds of beef. So he paid $20 \times 5.5 = \$110$. The answer is: 110

Original Data



Question Bootstrapping

Rephrasing Question: What is the total amount that James paid when he purchased 5 packs of beef, each weighing 4 pounds, at a price of \$5.50 per pound? **Answer:**

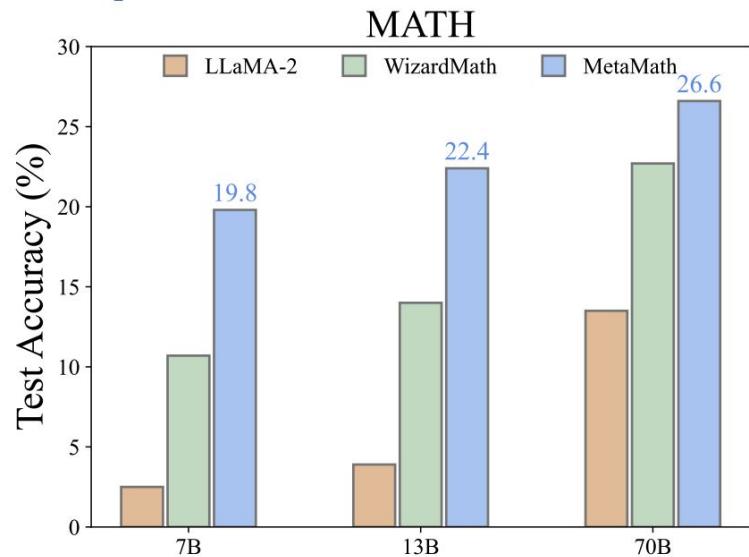
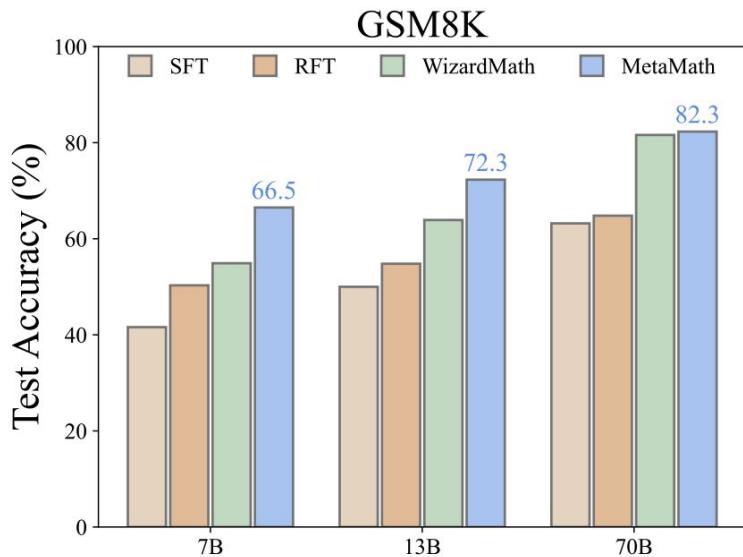
Self-Verification Question: James buys x packs of beef that are 4 pounds each. The price of beef is \$5.50 per pound. He paid 110. *What is the value of unknown variable x ?* **Answer:**

FOBAR Question: James buys x packs of beef that are 4 pounds each. The price of beef is \$5.50 per pound. How much did he pay? *If we know the answer to the above question is 110, what is the value of unknown variable x ?* **Answer:**

Answer Augment: James buys 5 packs of beef that are 4 pounds each, so he buys a total of $5 \times 4 = 20$ pounds of beef. The price of beef is \$5.50 per pound, so he pays $20 \times \$5.50 = \110 . The answer is: 110

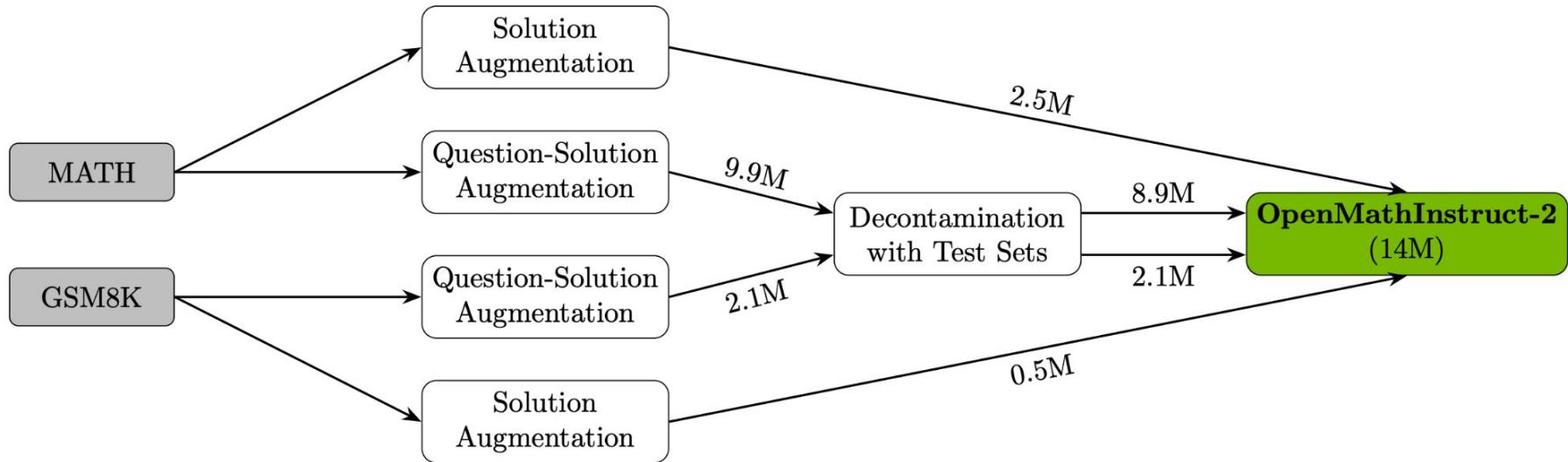
MetaMathQA

MetaMath

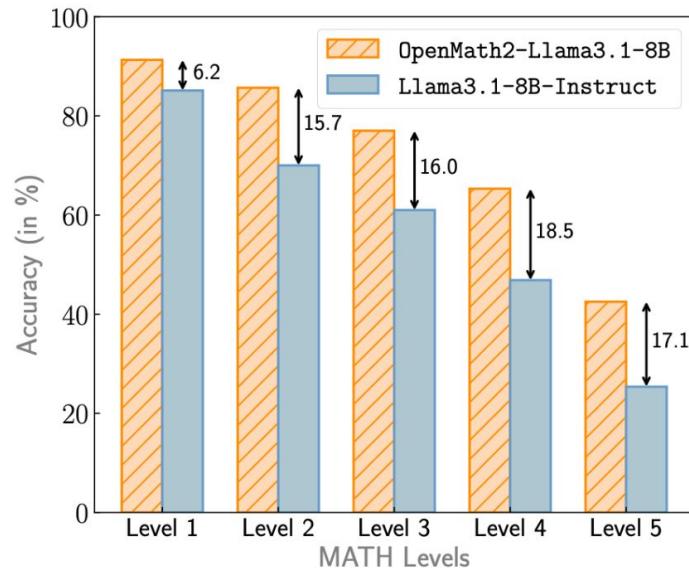
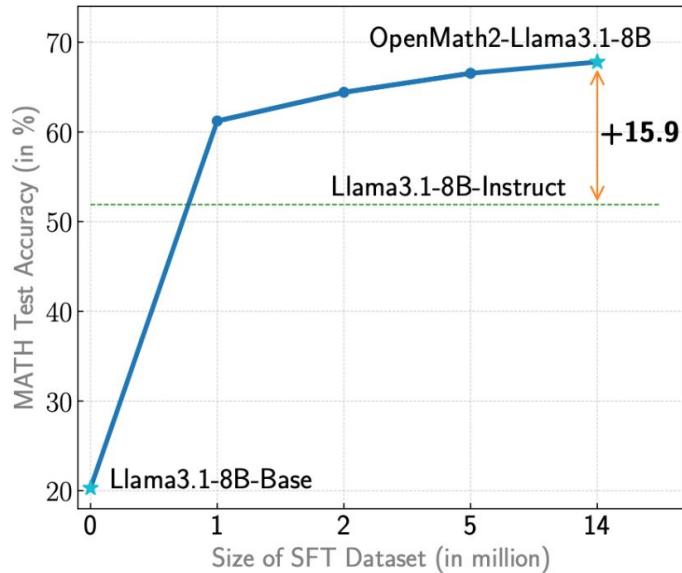


Yu, Longhui, et al. "Metamath: Bootstrap your own mathematical questions for large language models." *ICLR 2024*

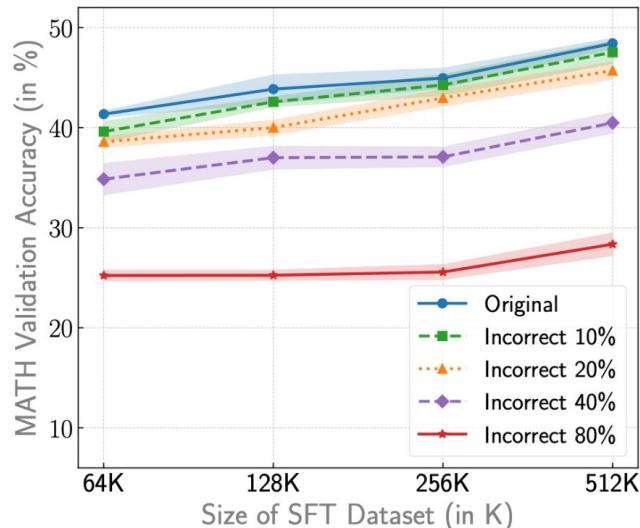
OpenMathInstruct-2



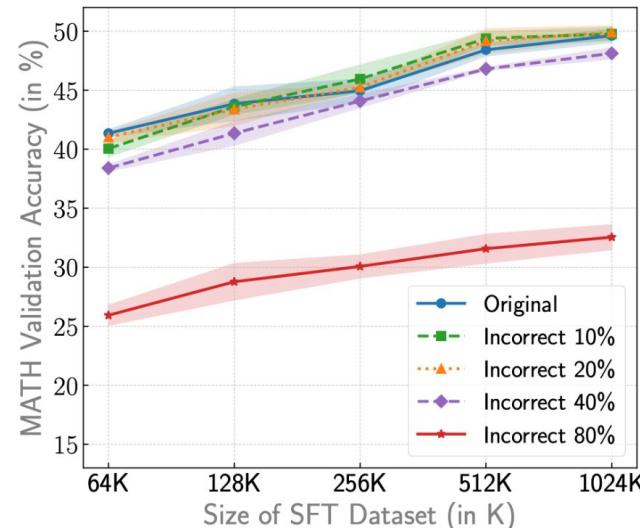
OpenMathInstruct-2



OpenMathInstruct-2



(a) Adding wrong-answer solutions.



(b) Correct solutions mismatched with questions

Figure 5: Impact of low-quality solutions on the SFT performance.

OpenMathInstruct-2

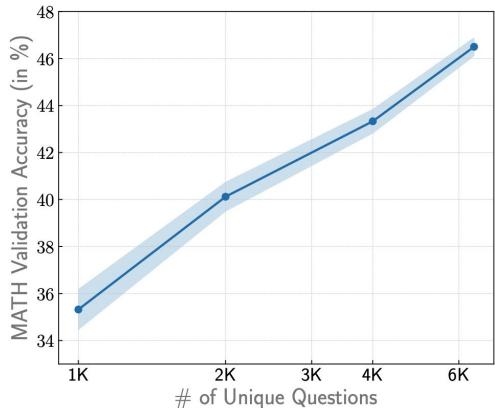
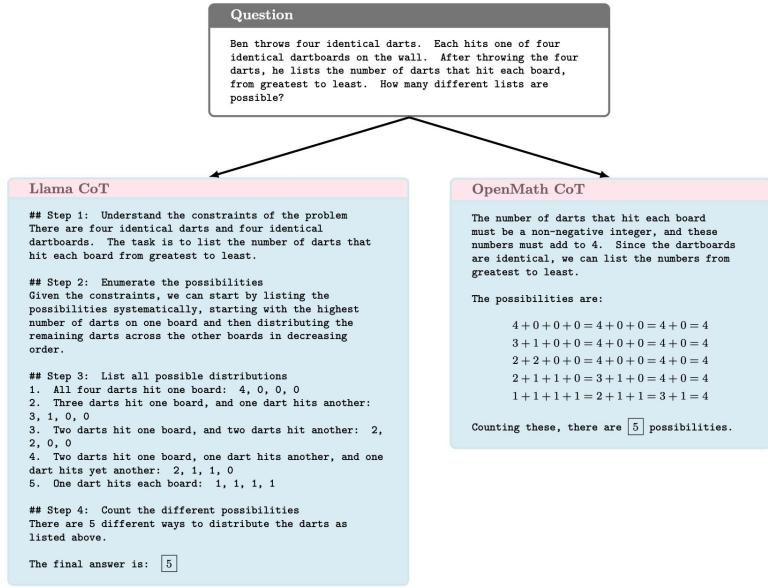
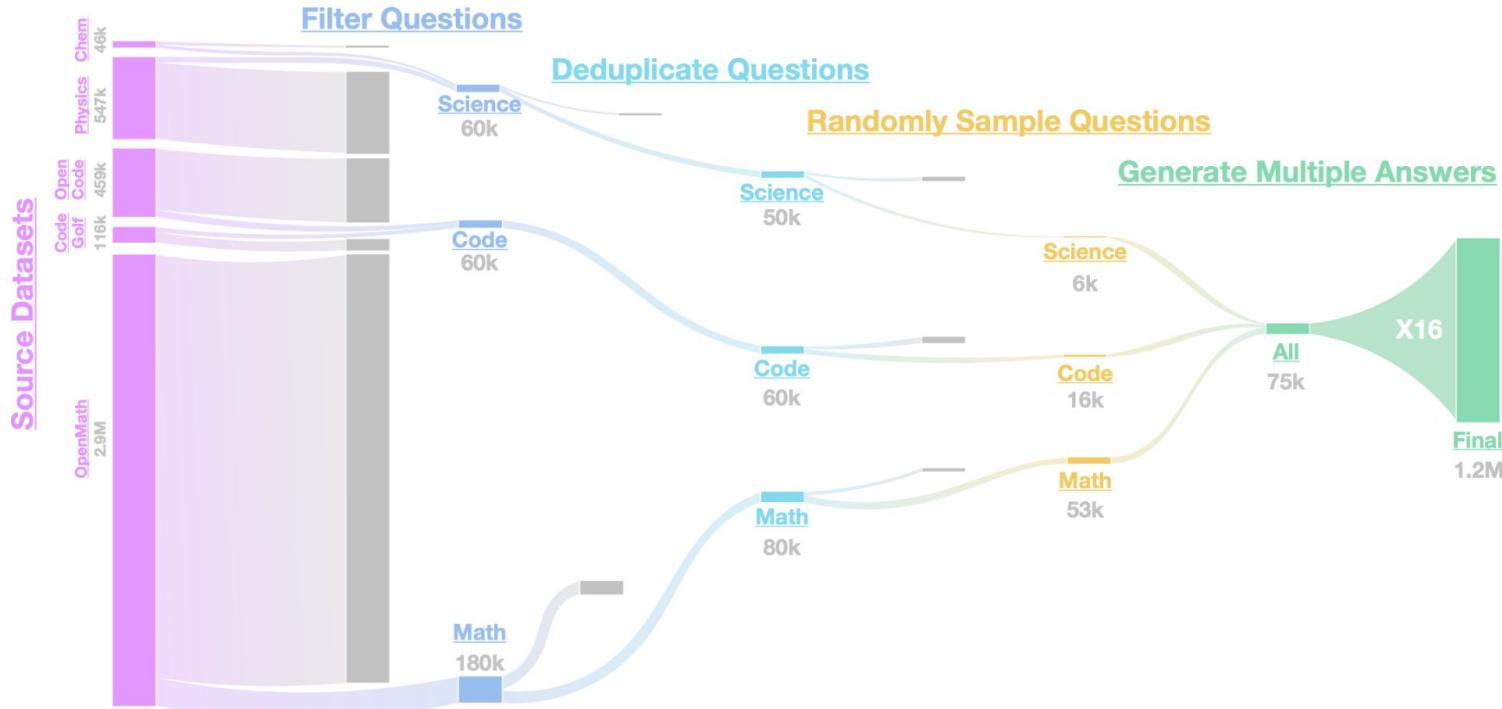


Figure 6: Impact of question diversity on MATH validation accuracy.

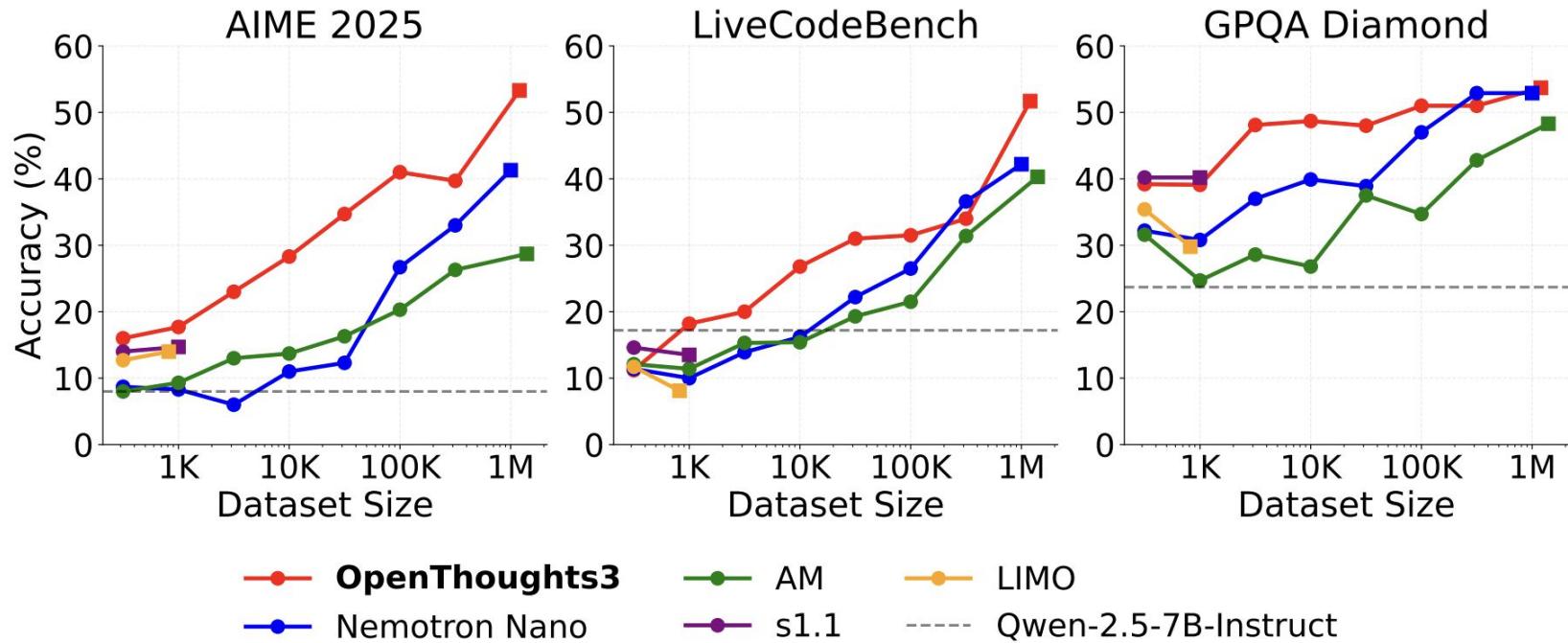


	MATH Validation Accuracy	Mean Solution Length
Llama CoT	40.6 ± 0.6	331.3
OpenMath CoT	44.5 ± 0.8	237.0

OpenThoughts



OpenThoughts



WebInstruct: 10M Synthetic Instruction Data from the Web

Physics Forums INSI ← Health & Medicine / Diseases in Health & Medicine / Inflammation

Forums > Homework Help > Biology a

Are there infinite combinations of equilibrium constant?

Chemistry · zenterix · Wednesday, 3:26 PM

Wednesday, 3:26 PM

zenterix

WOLFRAM COMMUNITY Dashboard

Is there a way to show the factors of a Perfect Number expressed as a sum rather than a product

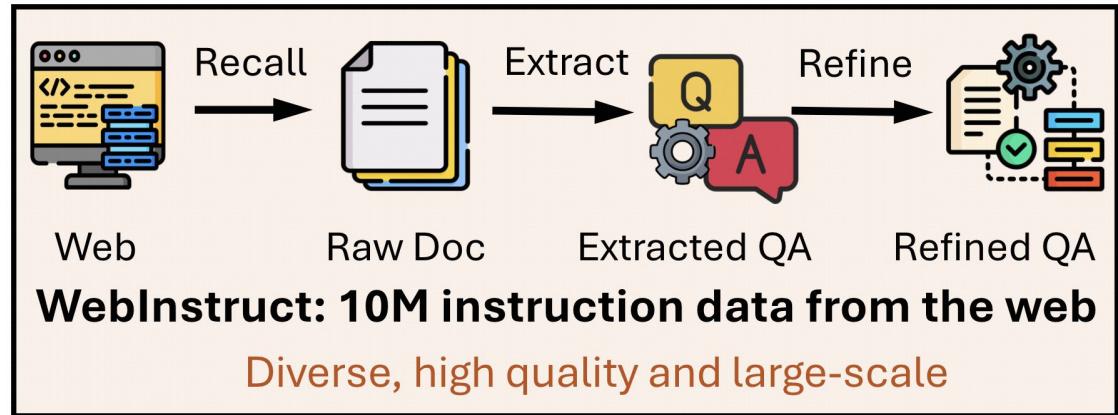
Bob Freeman, replied 21 hours ago

In 1890, Louisiana enacted a "black people and for white people only" train. He refused to move when arrested. The case went to the Supreme Court, which ruled to uphold the Louisiana law.

PO: CON-6 (EU), CON-6-A (LO), C

Which statement accurately describes the Ferguson (1896) decision?

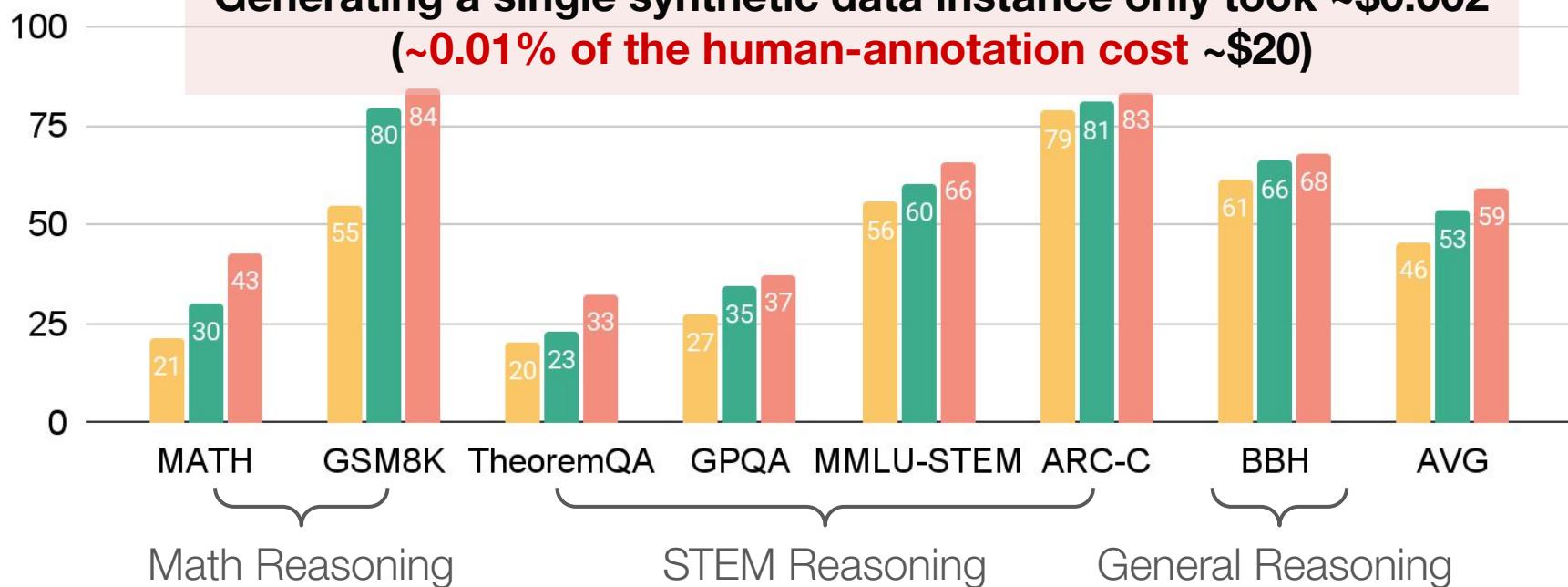
Do 4 problems



10M Synthetic Data vs 10M Human Annotated Data

■ Llama3 8B Base ■ Llama3 8B Instruct ■ Llama3-8B-MAmmoTH2-Plus

**Generating a single synthetic data instance only took ~\$0.002
(~0.01% of the human-annotation cost ~\$20)**



Code Generation

Code Instructions

Datasets: sahil2801/**CodeAlpaca-20k** like 139

Tasks: Text Generation Languages: English Size Categories: 10K<n<100K

Dataset card Viewer Files and versions Community 6

WAVECODER: WIDESPREAD AND VERSATILE ENHANCED INSTRUCTION TUNING WITH REFINED DATA GENERATION.

Zhaojian Yu, Xin Zhang, Ning Shang, Yangyu Huang, Can Xu, Yishujie Zhao, Wenxiang Hu, Qifeng Yin
Microsoft
{v-zhaojianyu,xinzhang3,nishang,yanghuan,caxu,v-yiszhaowenxh,qfyin}@microsoft.com

WizardCoder: Empowering Code Large Language Models with Evol-Instruct

Ziyang Luo^{2*} Can Xu^{1*} Pu Zhao¹ Qingfeng Sun¹ Xiubo Geng¹
Wenxiang Hu¹ Chongyang Tao¹ Jing Ma² Qingwei Lin¹ Daxin Jiang^{1†}
¹Microsoft

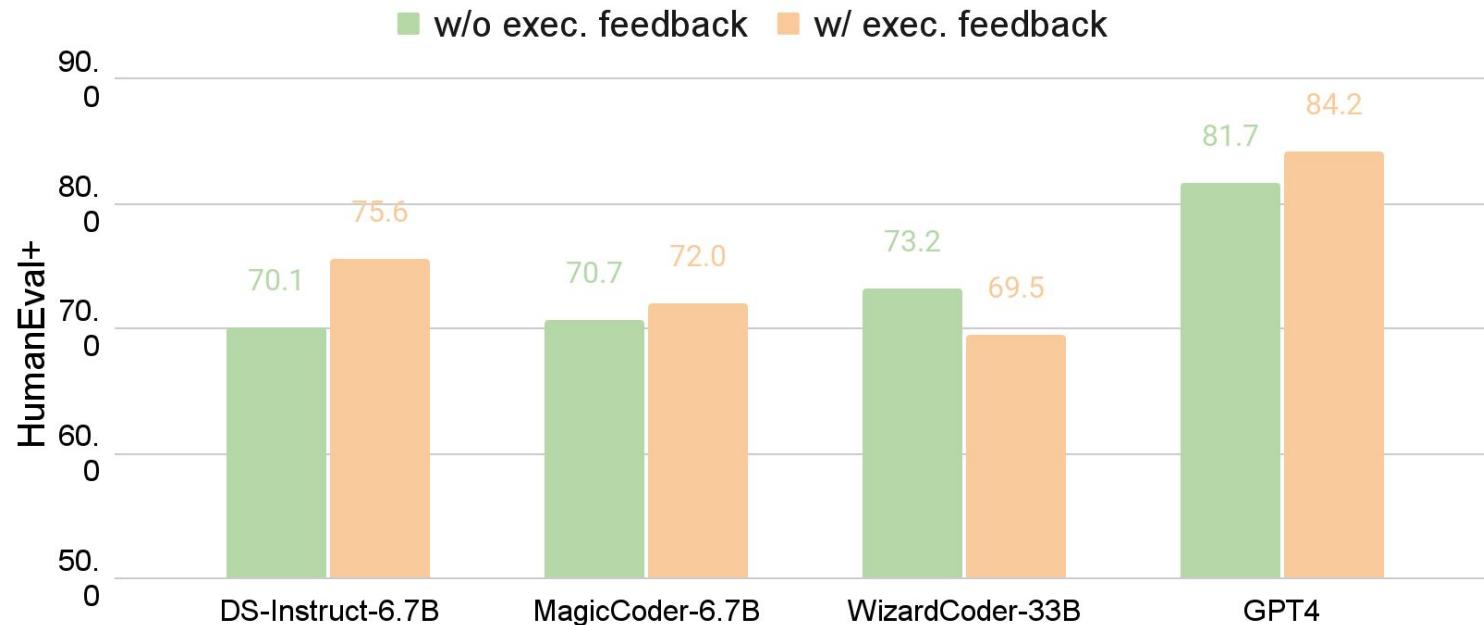
²Hong Kong Baptist University
{caxu,puzhao,qins,xigeng,wenxh,chongyang.tao,qlin,djiang}@microsoft.com
{cszyluo, majing}@comp.hkbu.edu.hk

Magicoder: Source Code Is All You Need

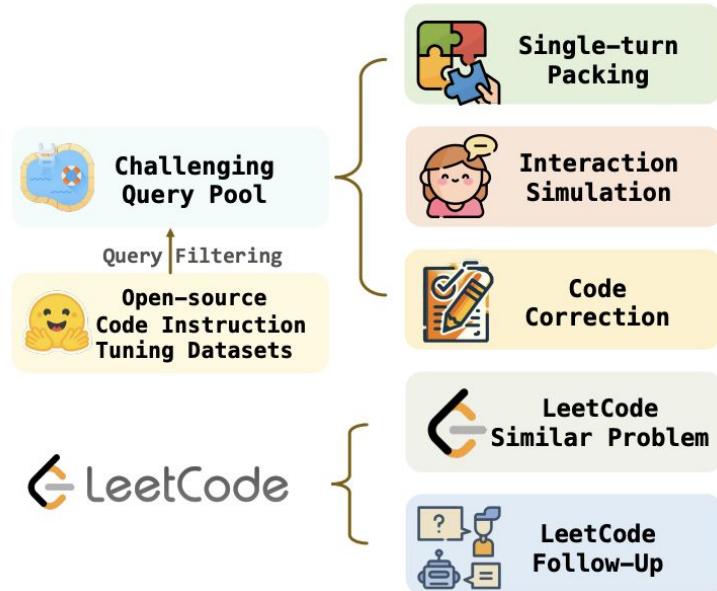
Yuxiang Wei¹ Zhe Wang² Jiawei Liu¹ Yifeng Ding¹ Lingming Zhang¹

¹University of Illinois at Urbana-Champaign ²Tsinghua University
{ywei40,jiawei6,yifeng6,lingming}@illinois.edu zhewang20@mails.tsinghua.edu.cn
 <https://github.com/ise-uiuc/magicoder>

Can models understand the execution feedback?



Code-Feedback



Dataset	#Sample	#Turn	M.T	E.F	H.F
CodeAlpaca ⁰	20k	20K	✗	✗	✗
Magicoder-OSS-Instruct ¹	75K	75K	✗	✗	✗
Python-Code-ShareGPT ²	23K	23K	✗	✗	✗
Magicoder-Evol-Instruct ³	111K	111K	✗	✗	✗
EvolInstruct-Code ⁴	80k	80K	✗	✗	✗
Code-Feedback (Ours)	68K	192K	✓	✓	✓
Single-turn Packing	16K	33.5K	✓	✗	✓
Interaction Simulation	51K	155.5K	✓	✓	✓
Code Correction	0.5K	1.2K	✓	✓	✗
LeetCode Similar Problem	0.3K	0.65K	✓	✗	✓
LeetCode Follow-Up	0.2K	0.76K	✓	✗	✓

Figure 2: Summary of our proposed dataset Code-Feedback construction and comparison with existing code instruction tuning datasets. M.T: Multi Turn, E.F: Execute Feedback, H.F: Human Feedback.

OpenCodeInterpreter



Implement a Python function to check if a string is a valid IPv6 address with regular expressions.



Sure, here is a Python function using Regular Expression (regex) to check if a string is a valid IPv6 address:
```python  
def is\_valid\_ipv6(ip):  
 ...  
 print(is\_valid\_ipv6("2404:6800:4003:c02::8a"))#True



```
error Traceback (most recent call last)
Input In [1], in <cell line: 11>()
error: nothing to repeat at position 21
```



I apologize for the confusion. It seems like there was a mistake in the regular expression pattern. Let's correct that. Here is the corrected function:  
```python  
def is_valid_ipv6(ip):
 ...
 print(is_valid_ipv6("2404:6800:4003:c02::8a"))#True



Execution Result: True



Great! The function appears to be working as expected now. It correctly identifies the strings "2404:6800:4003:c02::8a"..." Let me know if there's anything else I can assist you with!



Can you explain the generated regular expression in detail and make this function more maintainable?



Sure, let me break down the regular expression for you:
- `([0-9a-fA-F]{1,4}:[){7,7}[0-9a-fA-F]{1,4}:` This pattern matches an IPv6 address that consists of ...

OpenCodeInterpreter

Type a message...

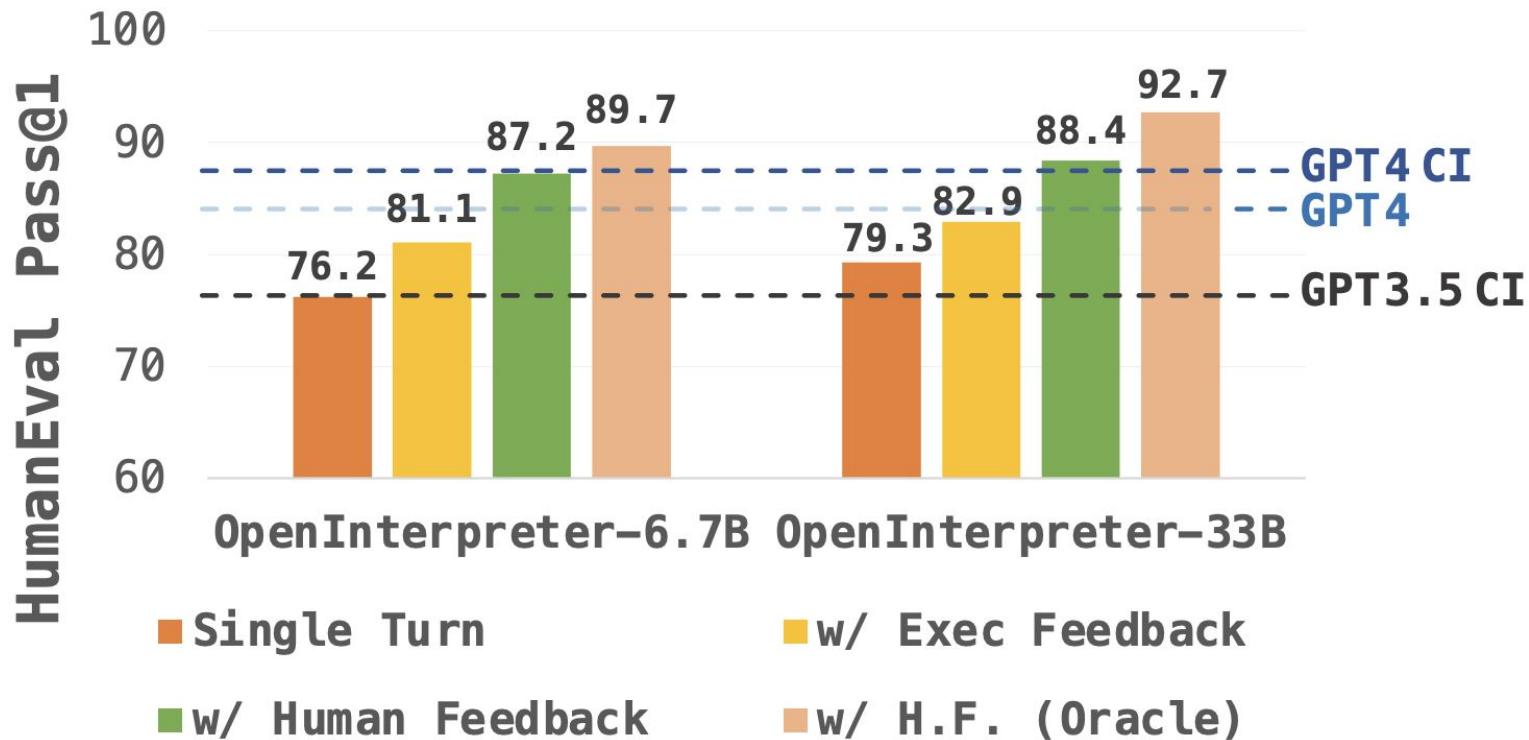
Submit

Retry

Undo

Clear

OpenCodeInterpreter: Results



Tool use and Agents

Toolformer

The New England Journal of Medicine is a registered trademark of [QA("Who is the publisher of The New England Journal of Medicine?") → Massachusetts Medical Society] the MMS.

Out of 1400 participants, 400 (or [Calculator(400 / 1400) → 0.29] 29%) passed the test.

The name derives from "la tortuga", the Spanish word for [MT("tortuga") → turtle] turtle.

The Brown Act is California's law [WikiSearch("Brown Act") → The Ralph M. Brown Act is an act of the California State Legislature that guarantees the public's right to attend and participate in meetings of local legislative bodies.] that requires legislative bodies, like city councils, to hold their meetings open to the public.

Your task is to add calls to a Question Answering API to a piece of text. The questions should help you get information required to complete the text. You can call the API by writing "[QA(question)]" where "question" is the question you want to ask. Here are some examples of API calls:

Input: Joe Biden was born in Scranton, Pennsylvania.

Output: Joe Biden was born in [QA("Where was Joe Biden born?")] Scranton, [QA("In which state is Scranton?")] Pennsylvania.

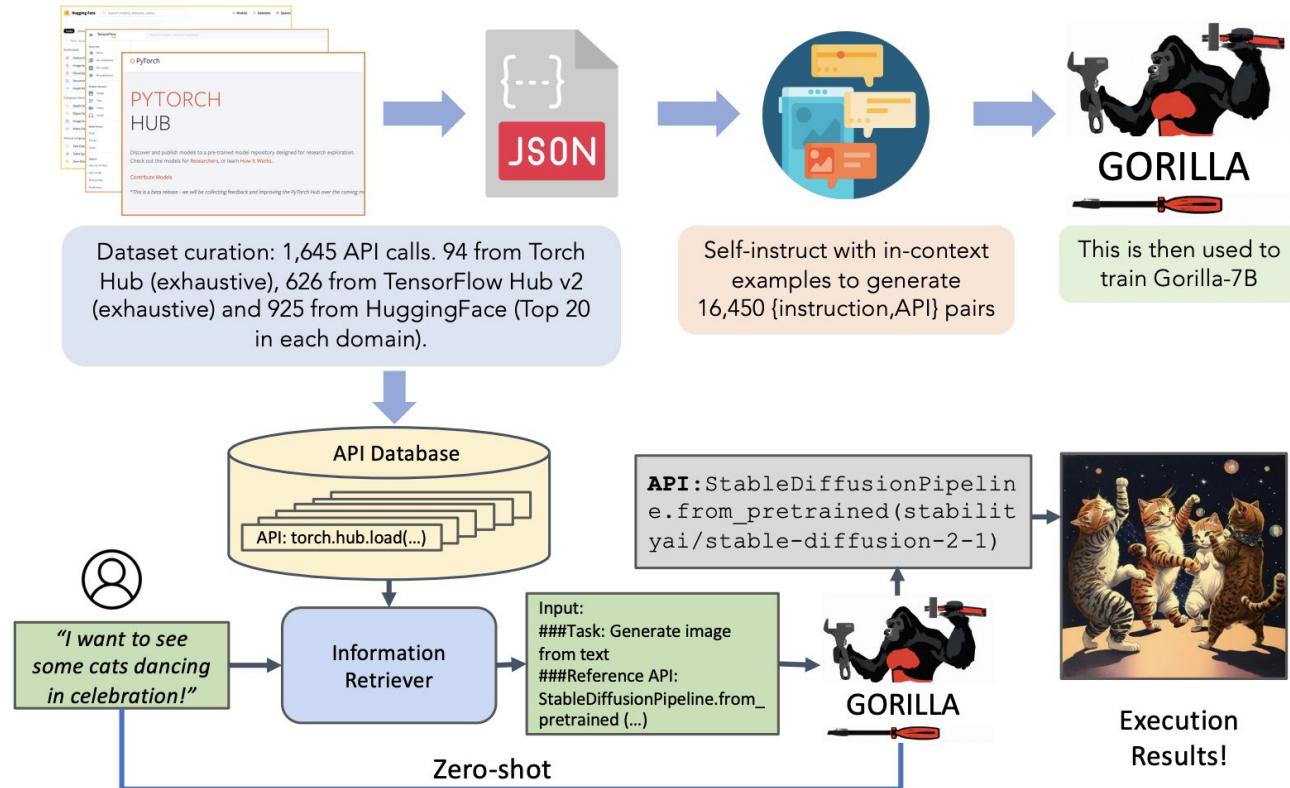
Input: Coca-Cola, or Coke, is a carbonated soft drink manufactured by the Coca-Cola Company.

Output: Coca-Cola, or [QA("What other name is Coca-Cola known by?")] Coke, is a carbonated soft drink manufactured by [QA("Who manufactures Coca-Cola?")] the Coca-Cola Company.

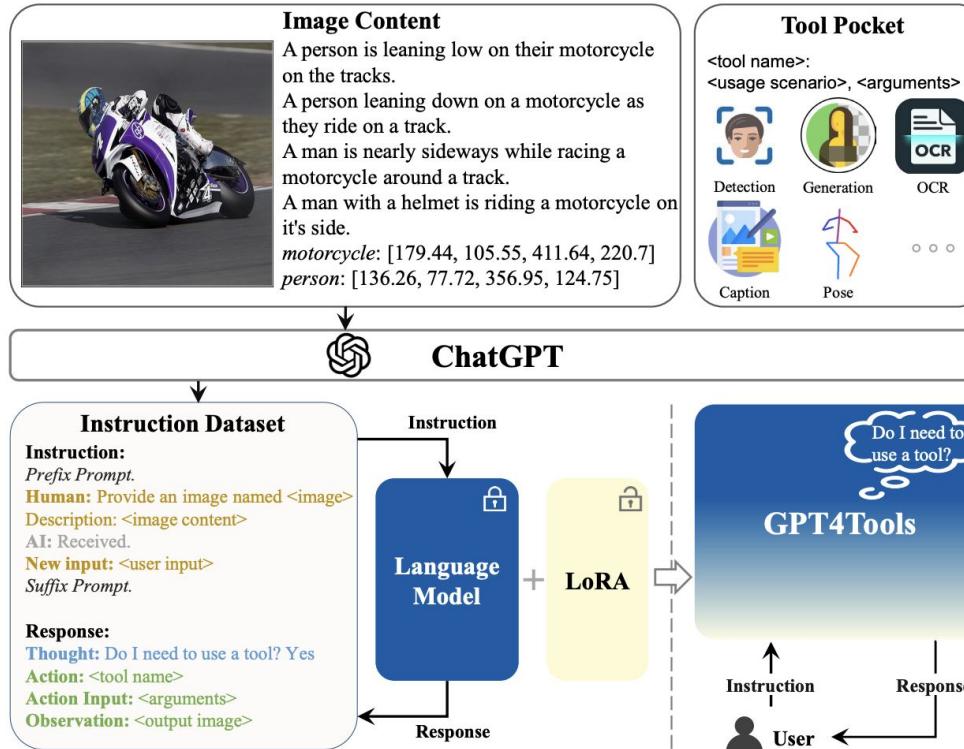
Input: x

Output:

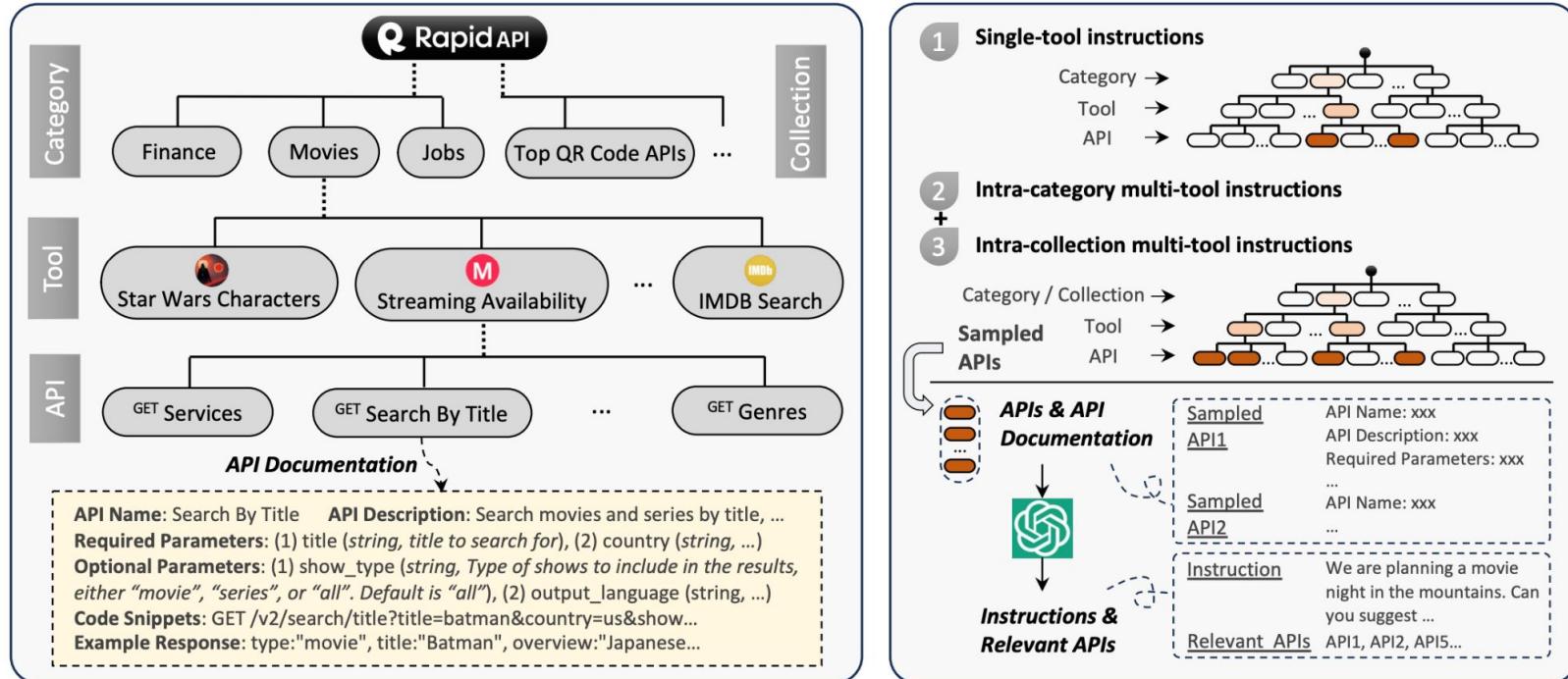
Gorilla



GPT4Tools



TooLLM



ToRA: Tool-Integrated Reasoning

Problem: Suppose that the sum of the squares of two complex numbers x and y is 7 and the sum of their cubes is 10. List all possible values for $x + y$, separated by commas.

We are given that $x^2 + y^2 = 7$ and $x^3 + y^3 = 10$. We can factor $x^3 + y^3$ to get $(x + y)(x^2 - xy + y^2)$.

Thus, we can write $10 = (x + y)(7)$. It follows that $x + y$ must divide 10.

The only possibilities are that $x + y$ is 1, 2, 5 or 10.



```
import math

def solution():
    x = symbols('x'), y = symbols('y')
    x**2 + y**2 = 7
    x**3 + y**3 = 10

    x = (-1 + math.sqrt(17)) / 2
    y = (-1 - math.sqrt(17)) / 2

    # Get all possible values for x+y
    result = [x + y, -x - y]
```



>>> SyntaxError: cannot assign to expression.

Let's use `Sympy` to calculate and print all possible values for $x + y$,

```
from sympy import symbols, Eq, solve, simplify

def possible_values():
    x, y = symbols('x y')
    eq1 = Eq(x**2 + y**2, 7)
    eq2 = Eq(x**3 + y**3, 10)
    solutions = solve((eq1, eq2), (x, y))
    sums = [simplify(sol[0] + sol[1]) for sol in
            solutions]
    return sums
print(possible_values())
```



>>> [-5, -5, 1, 1, 4, 4]

Removing duplicates, the possible values for $x + y$ are \boxed{-5, 1, 4}.

r_1

a_1

O_1

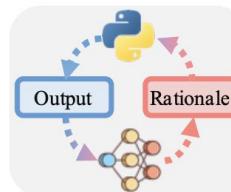
r_2



(a) Rationale-based



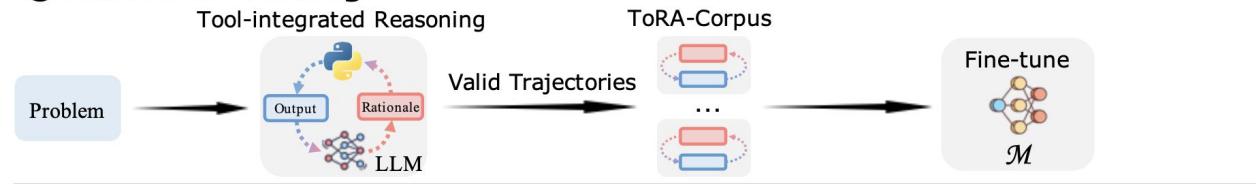
(b) Program-based



(c) Tool-integrated Reasoning
(Format used by ToRA)

ToRA: Tool-Integrated Reasoning

① Imitation Learning



② Output Space Shaping

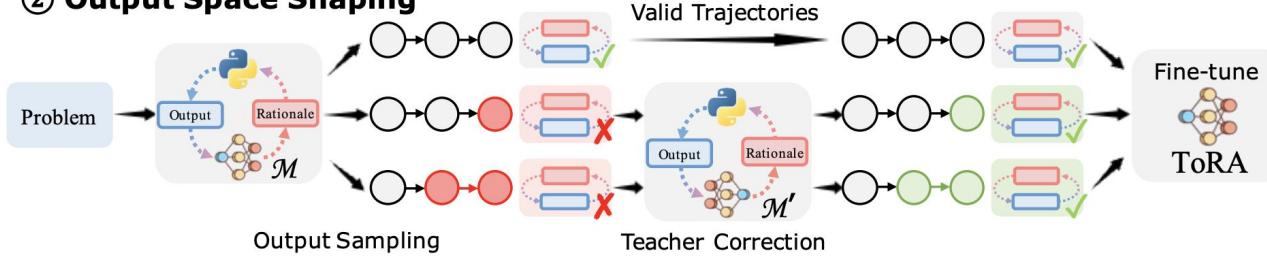
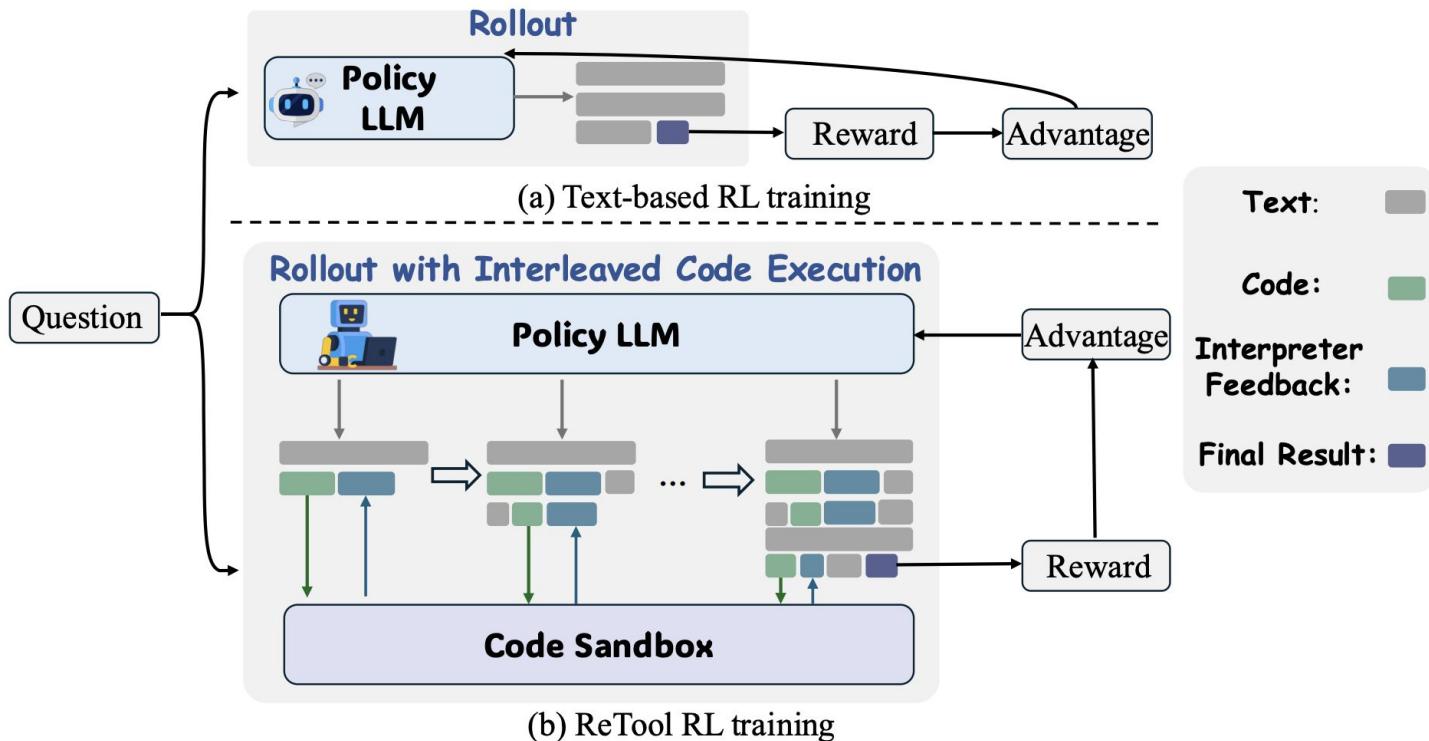


Figure 3: Training ToRA contains two steps. ① **Imitation Learning**: Prompt LLMs like GPT-4 to generate Tool-integrated Reasoning trajectories (ToRA-CORPUS) and use this corpus to fine-tune a model \mathcal{M} ; ② **Output Space Shaping**: Sample diverse tool-use trajectories with \mathcal{M} , keep the valid ones, correct the invalid ones with a teacher model \mathcal{M}' , and retrain \mathcal{M} on the union of sampled valid trajectories, corrected ones, and the initial ToRA-CORPUS to obtain ToRA.

ReTool: RL for Strategic Tool Use



ReTool: RL for Strategic Tool Use

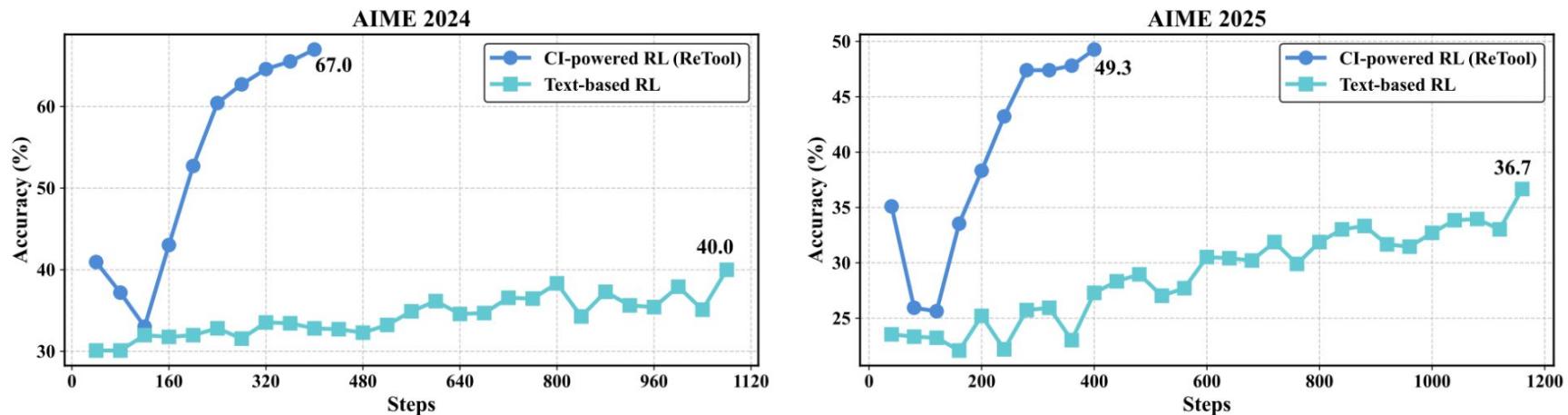
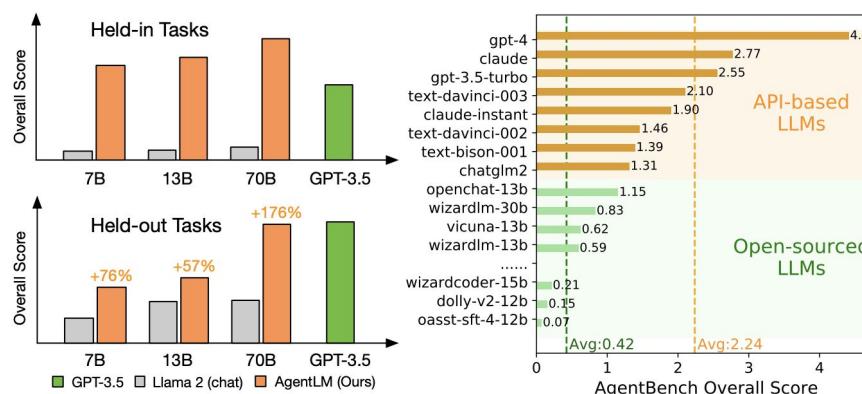
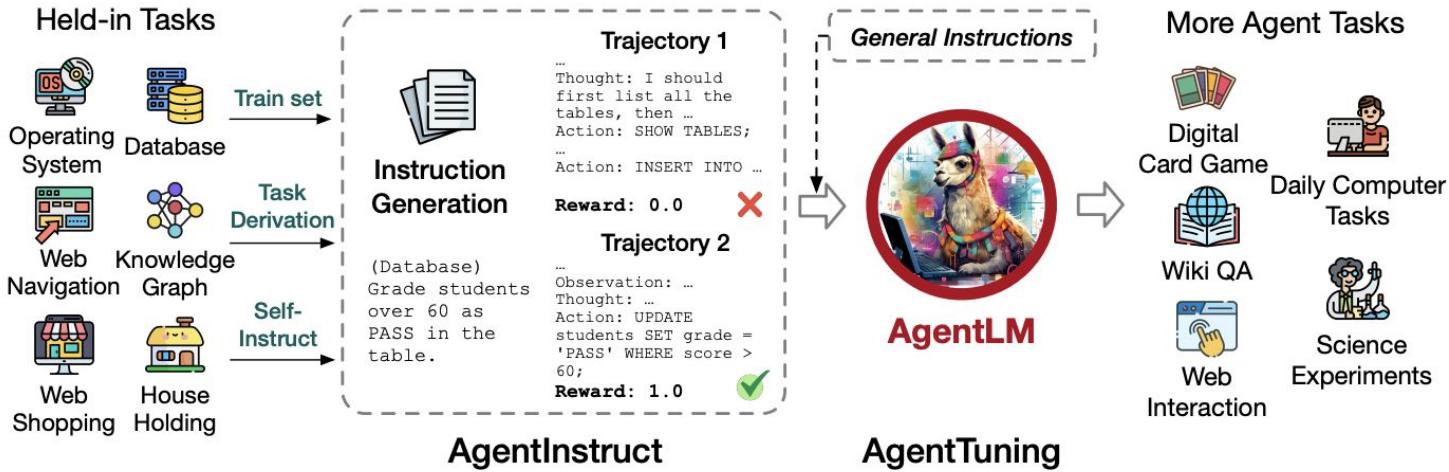
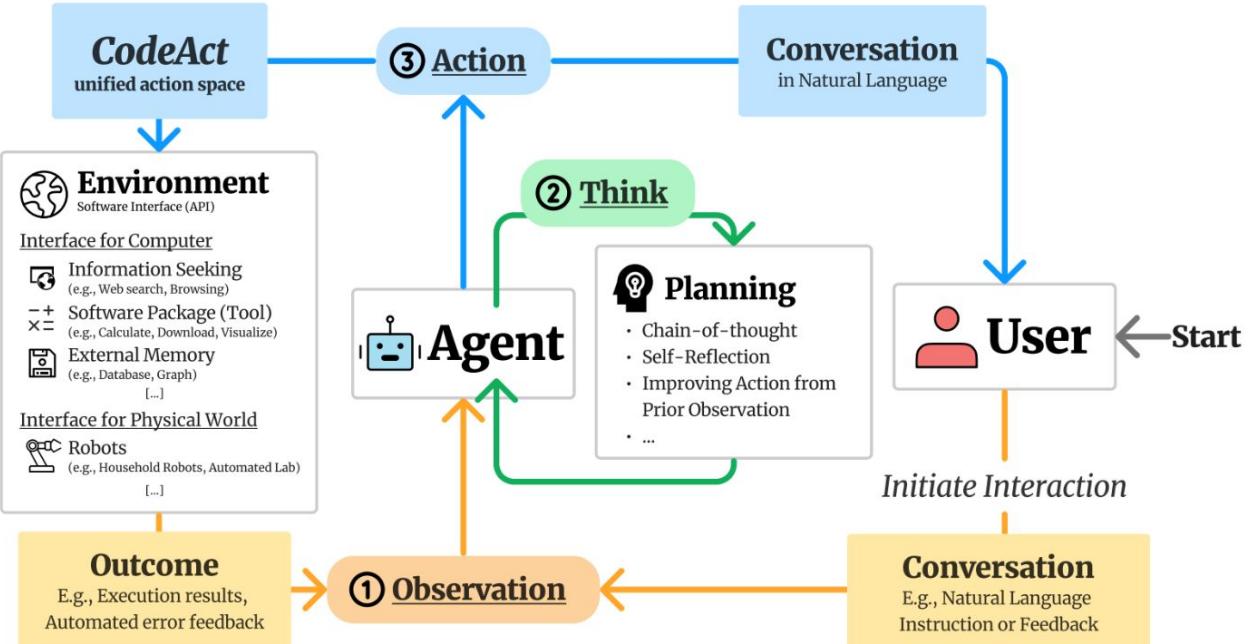
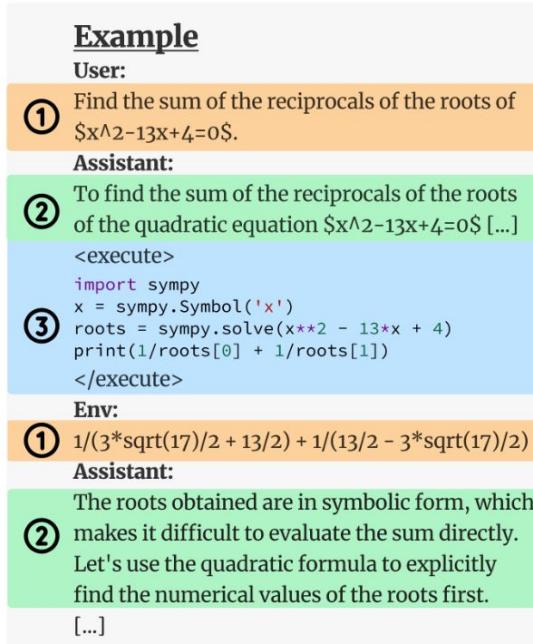


Figure 1 AIME 2024 & 2025 scores of ReTool and text-based RL baseline on the Qwen2.5-32B-Instruct model.

AgentTuning



CodeActInstruct



SWE-Smith

Real Repositories

<> Source Code

- src/ README.rst
- tests/ setup.py

Unit Tests

- ✓ tests/test_api.py
- ✓ tests/test_auth.py
- ✓ tests/test_client.py
- ✓ tests/test_utils.py

Environment Creation

SWE-agent

Attempts to install the repo and execute tests

Developer

Writes Dockerfile based on SWE-agent work

Task Gen. Strategies

Procedural Modification

LM Generated

Combine Bugs

PR Mirroring

New Task Instances

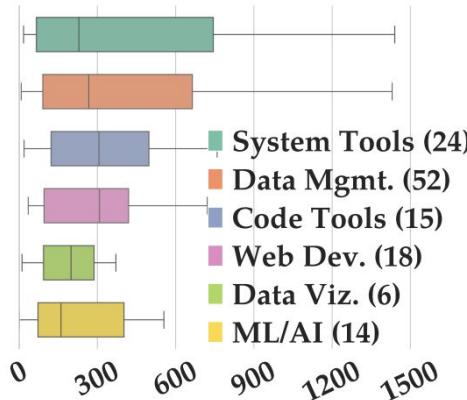
Environment Image

Synthetic Tasks

Generated Issue

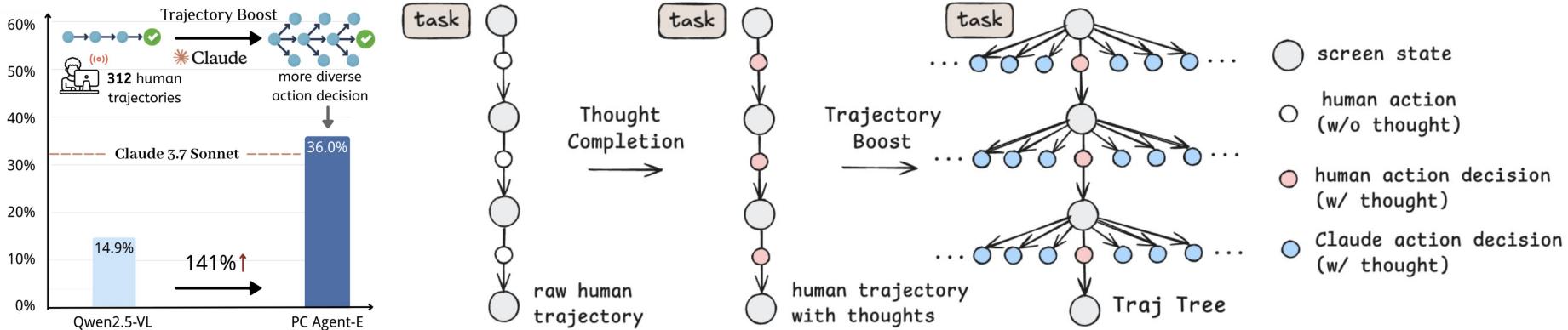
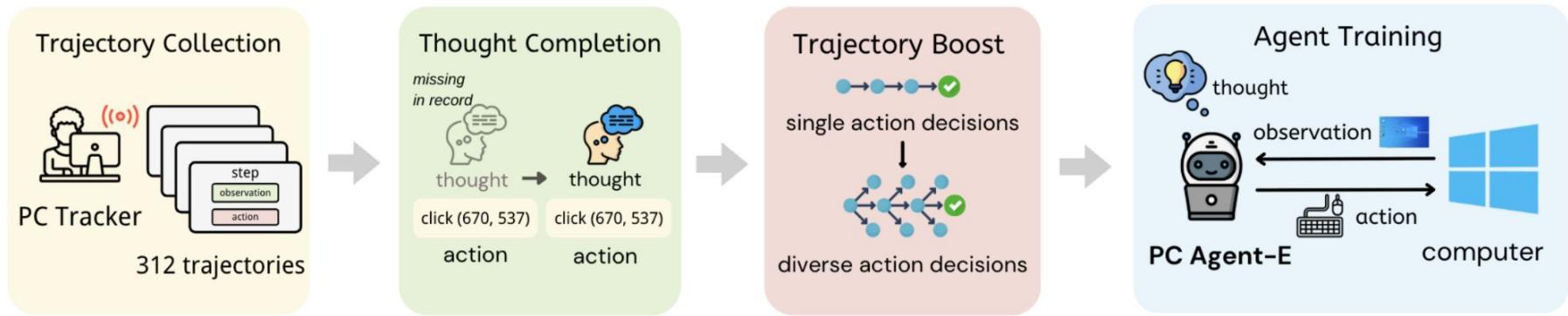
Bugged Patch

Verified Tests



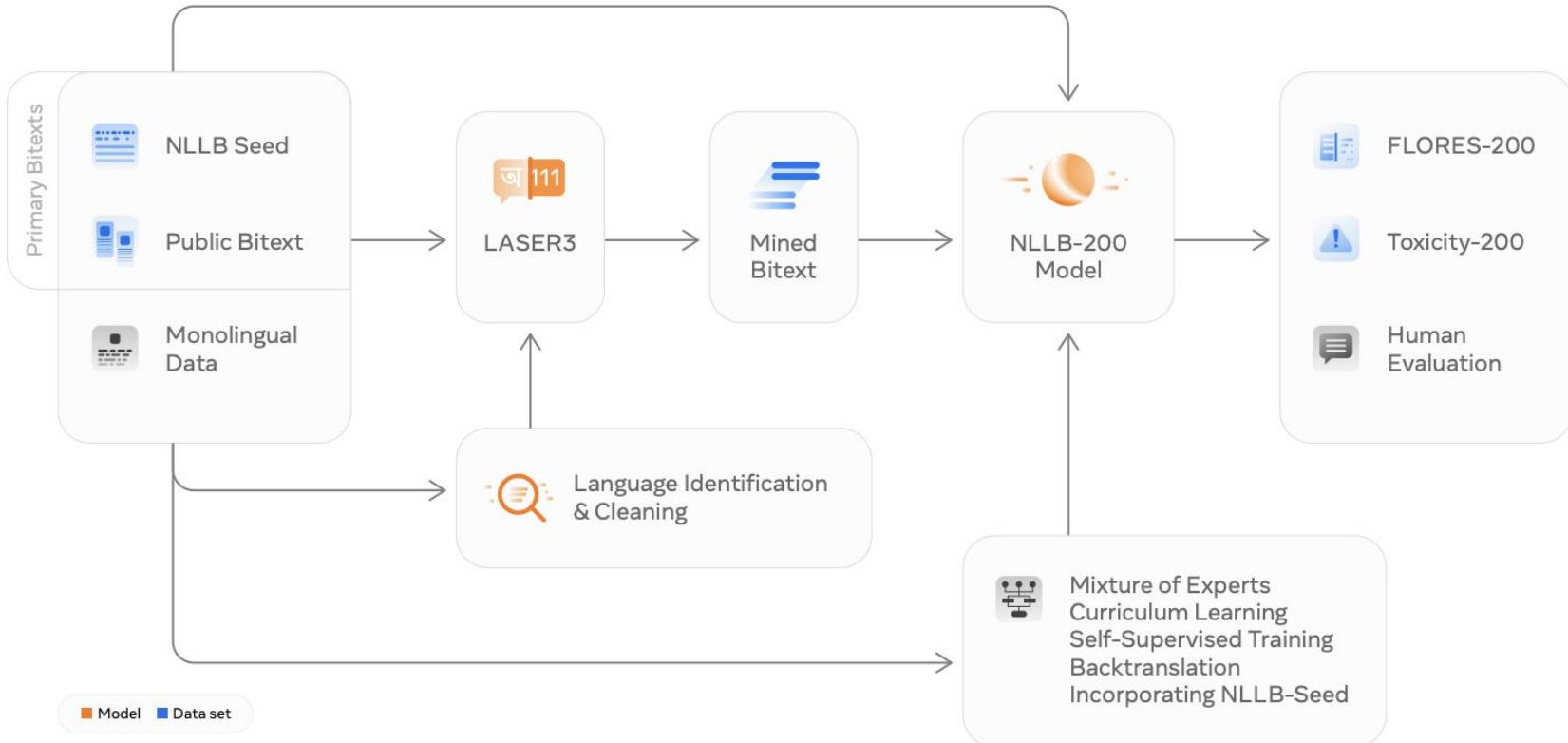
| Bug Type | Yield % | # Insts | Cost | F2P | Lines |
|------------|---------|---------|-------|-----|-------|
| Combine | 96.9% | 10,092 | 0.00¢ | 15 | 11 |
| LM Modify | 56.0% | 17,887 | 0.38¢ | 4 | 3 |
| LM Rewrite | 35.0% | 4,173 | 3.93¢ | 4 | 24 |
| PR Mirror | 33.8% | 2,344 | 5.53¢ | 3 | 14 |
| Procedural | 40.2% | 15,641 | 0.00¢ | 7 | 5 |
| Total | 50.1 | 50,137 | 2.32¢ | 6 | 5 |

Agent-E



Multilingual and Multimodal

NLLB: Scaling Machine Translation



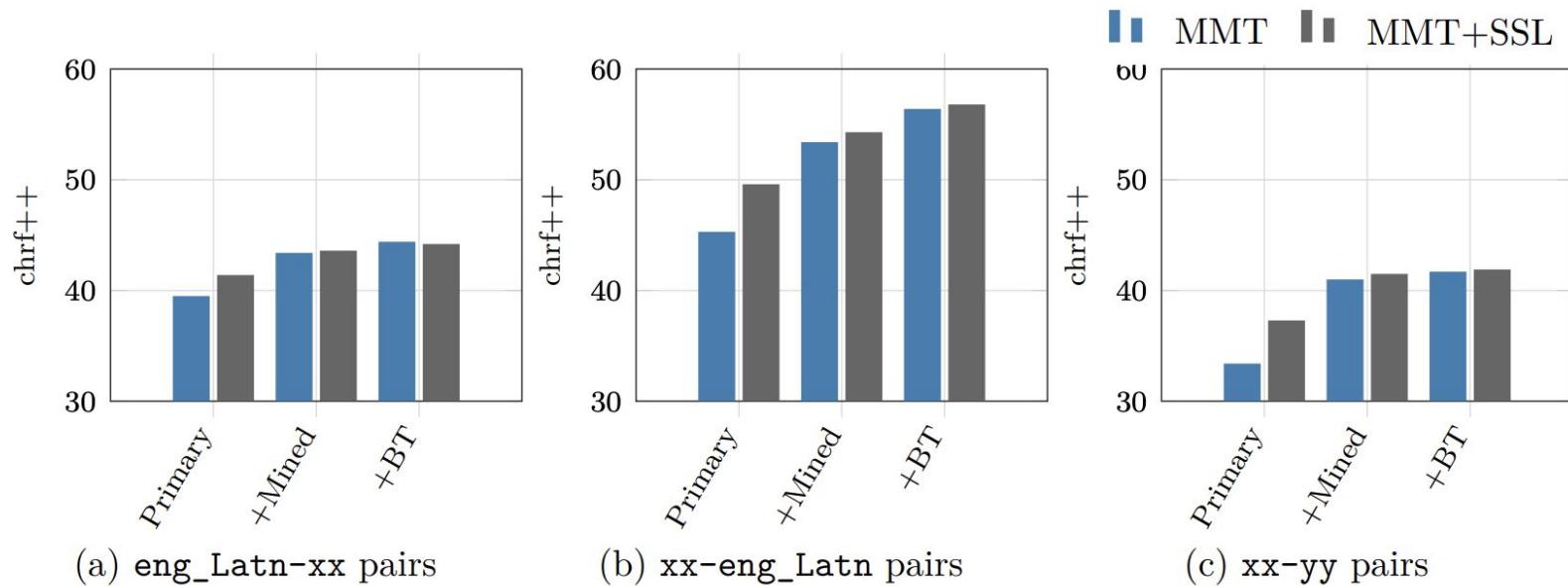


Figure 36: **Comparison of only MMT vs MMT+SSL Multitask Performance** on aggregate over Into English and Out of English directions, when trained on different sources of data. We compare on chrF++ scores. We observe that as we add more and more data from mining and/or backtranslation, we start to see diminishing improvements from the SSL task.

LLaVA: Visual Instruction Tuning

Context type 1: Captions

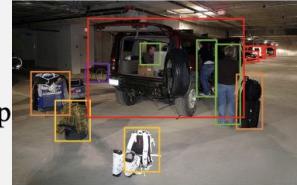
A group of people standing outside of a black vehicle with various luggage.

Luggage surrounds a vehicle in an underground parking area

People try to fit all of their luggage in an SUV.

The sport utility vehicle is parked in the public garage, being packed for a trip

Some people with luggage near a van that is transporting it.



Context type 2: Boxes

person: [0.681, 0.242, 0.774, 0.694], backpack: [0.384, 0.696, 0.485, 0.914], suitcase: ...<omitted>

Response type 1: conversation

Question: What type of vehicle is featured in the image?

Answer: The image features a black sport utility vehicle (SUV) ...<omitted>

Response type 2: detailed description

The image is an underground parking area with a black sport utility vehicle (SUV) parked. There are three people in the scene, with one person standing closer to the left side of the vehicle, another person in the middle, and the third person on the right side. They are all working together to pack their luggage into the SUV for a trip. ...<omitted>

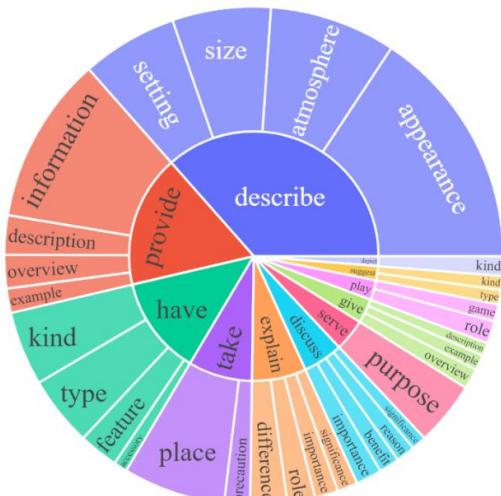
Response type 3: complex reasoning

Question: What challenges do these people face?

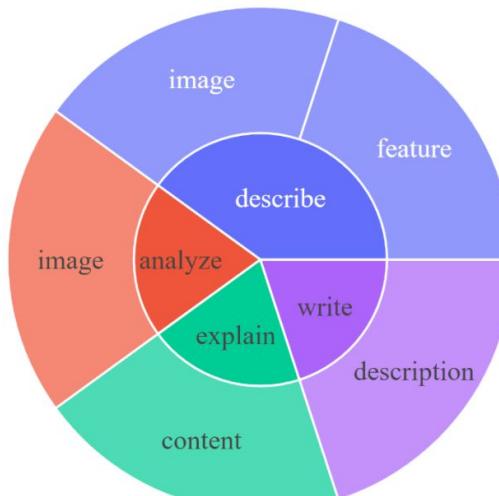
Answer: In the image, a group of people is standing outside a black SUV in a parking area, surrounded by various pieces of luggage, including suitcases and backpacks. They are facing the challenge of fitting all their luggage into the black SUV. There are multiple suitcases and backpacks to be packed, which suggests that the group has a significant amount of belongings ...<omitted>

| Data file name | File Size | Sample Size |
|----------------------------|-----------|-------------|
| conversation_58k.json | 126 MB | 58K |
| detail_23k.json | 20.5 MB | 23K |
| complex_reasoning_77k.json | 79.6 MB | 77K |

For each subset, we visualize the root noun-verb pairs for the instruction and response. For each chart, please click the link for the interactive page to check out the noun-verb pairs whose frequency is higher the given number.



Instruction: Conversation [0, 20, 50]



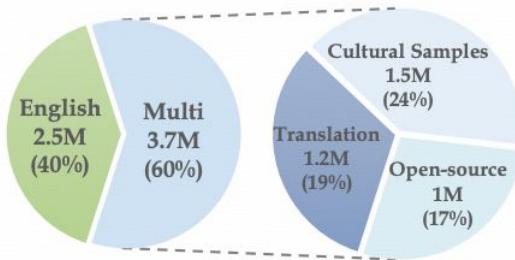
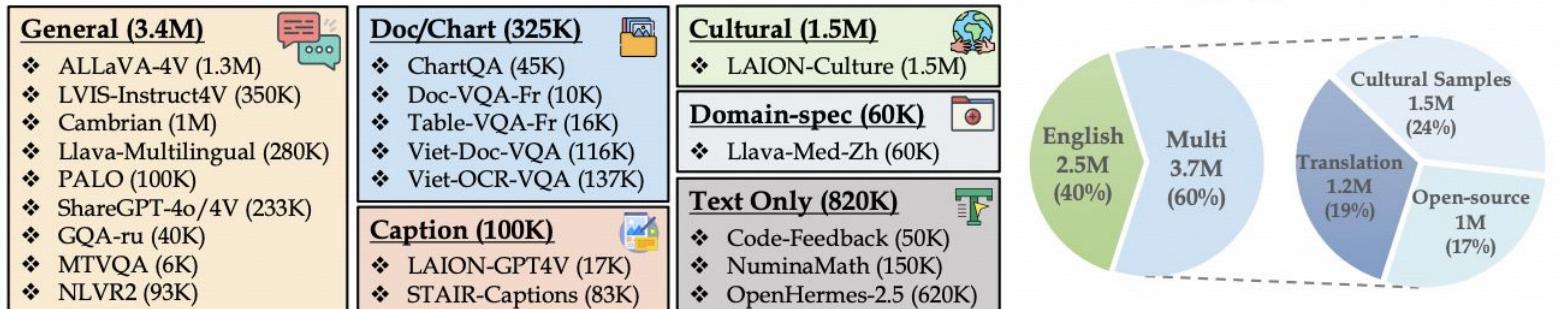
Instruction: Detailed Description [0]



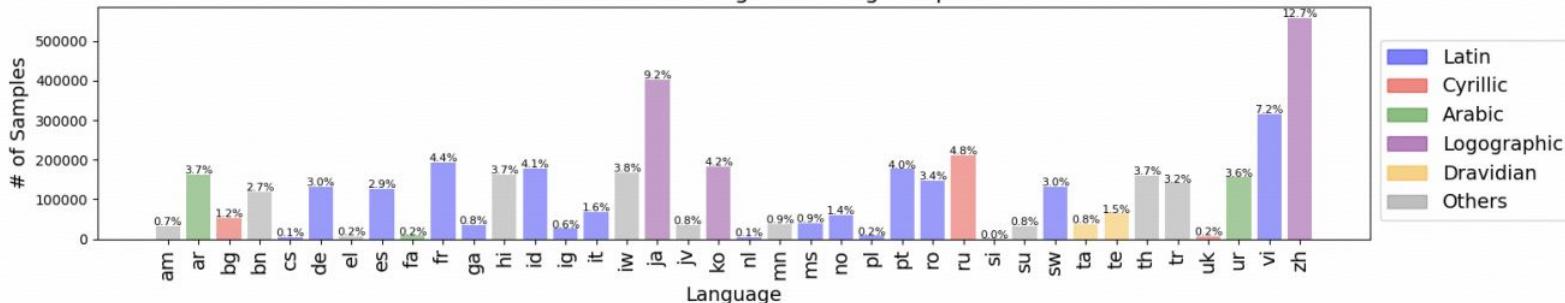
Instruction: Complex Reasoning [0, 20, 50]

Training Data: 6M Synthetic Instructions in 39 Languages

PangeaIns: 6M Multilingual Multimodal Instructions for 39 Languages



Distribution of Multilingual Training Samples



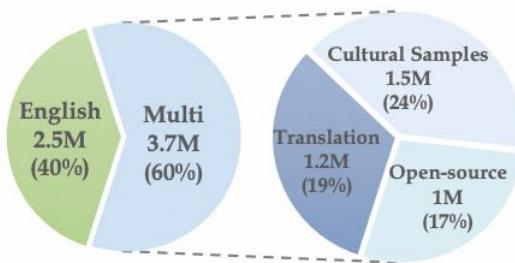
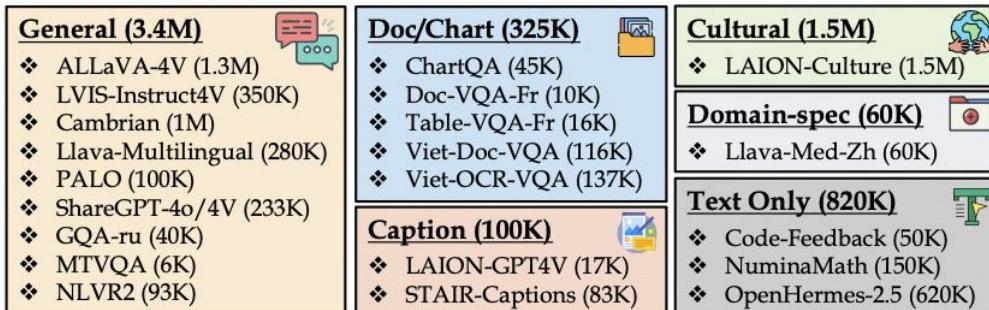
Translation

Culture-aware Synthetic Data

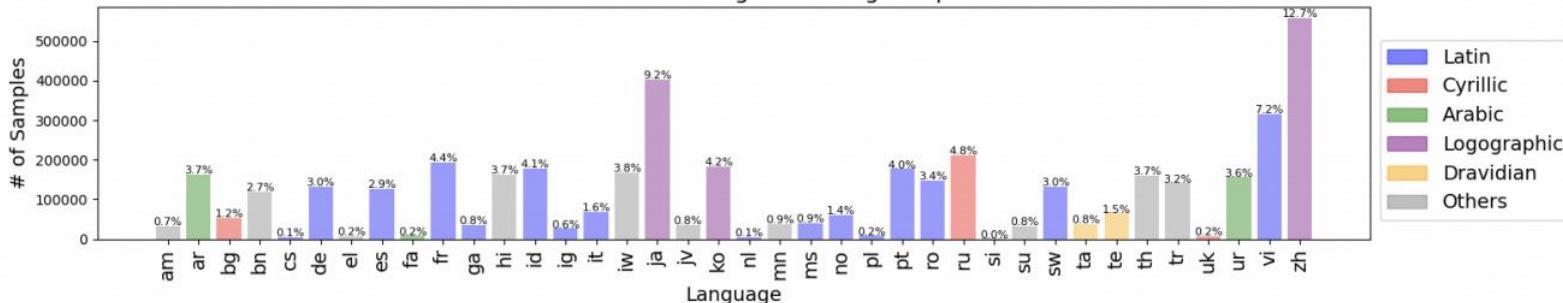
Open Data

Training Data: 6M Synthetic Instructions in 39 Languages

PangeaIns: 6M Multilingual Multimodal Instructions for 39 Languages



Distribution of Multilingual Training Samples



Translation

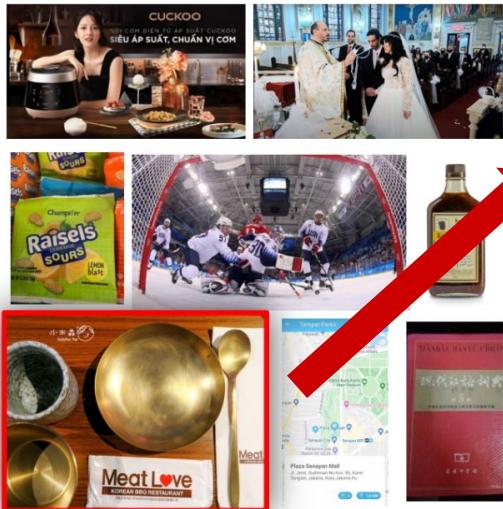
Culture-aware Synthetic Data

Open Data

Step 1: Selecting Culturally Diverse Images with LLMs



LAION-Multi



Informativeness: 4

Select images with informative text



Alt Text: 台北信義燒肉 | Meat Love 橡木炭火燒肉 韓國烤肉最高7折優惠!小菜無限吃到飽! (Taipei Xinyi BBQ | Meat Love Oak Charcoal BBQ – Korean BBQ up to 30% off! Unlimited free side dishes!)

Topic: Cooking and Food

Balance domains and topics of images



LLM Scoring

Informativeness

Rate the following alt text on a scale from 1 to 5 based on its quality in describing the image...

Topic Classify

Assign a category to the alt text based on its content. Choose from the following categories...

Country Classify

Decide if the alt text is related to a specific country's culture...

Country: Korea

Select culturally relevant images

Step 2: Recaptioning Multicultural Images



Alt Text: 上海ディズニーランドの模型のそばにいるウォルト・ディズニー・カンパニーの社長兼CEO

(President and CEO of The Walt Disney Company by a model of Shanghai Disneyland)

Caption w/o Alt Text: 画像には、暗いスーツに淡い青色のシャツを着て、ネクタイをしていない男性が、大きな城の模型の前に立っている様子が映っています。その城は... (*The image features a man in a dark suit, light blue shirt, and no tie, standing in front of a large model of a castle. The castle ...*)

Recaption with Alt Text: 画像には、ウォルト・ディズニー・カンパニーの社長兼CEOが上海ディズニーランドの模型の前に立っている様子が映っています。背景には色鮮やかなフラワー・アレンジメントが広がっています。(*The image features the President and CEO of The Walt Disney Company standing in front of a model of Shanghai Disneyland. In the background, vibrant floral arrangements ...*)

Step 3: Generating Multilingual Instructions



Recaption with Alt Text: 画像には、ウォルト・ディズニー・カンパニーの社長兼CEO が 上海ディズニーランド の模型の前に立っている様子が映っています。背景には色鮮やかなフラワーアレンジメントが広がっています。*(The image features the President and CEO of The Walt Disney Company standing in front of a model of Shanghai Disneyland. In the background, vibrant floral arrangements ...)*

$D_s = F_s(\mathcal{O}(\text{Prompt}_s(D); \theta))$ Prompt a LLM to generate data

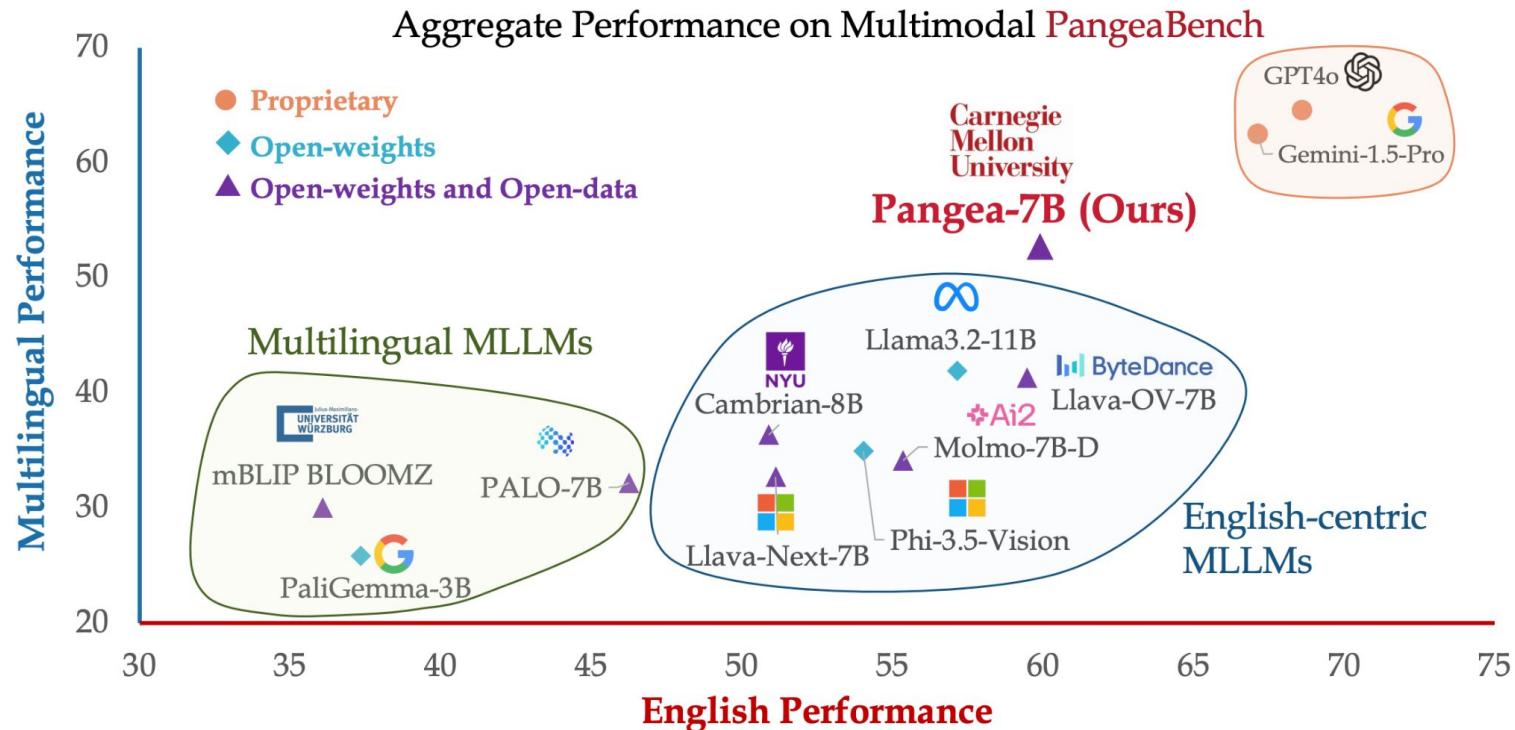
Prompt_s: Generate two **instruction-response pairs** based on the visual content of an image. The language should be in **[Japanese]**, the same as the caption text...

Instruction: 画像の中の男性がイベント中に行っているかもしれないスピーチを想像してください。

(Imagine a speech the man in the image might be giving during the event.)

Response: ここに立ち、上海ディズニーランドの素晴らしいモデルの前で皆さんとこの特別な瞬間を共有できることを、大変光栄に思います。... *(I am truly honored to stand here and share this special moment with you in front of this magnificent model of Shanghai Disneyland...)*

Pangea: SOTA fully-open **multilingual multimodal** LLM



Conclusion and Future Directions

Conclusion and Future Directions

- Quality vs. Quantity
- How to incorporate domain-specific experience?
- More diverse modalities & tasks
- Combining techniques
- Community-driven efforts

Limitations and Open Questions

Tutorial Summary

- Motivation
 - Why do we need synthetic data?
 - What is “high-quality” synthetic data?
- How do we get the data?
 - Sampling-based generation
 - Back-translation
 - Transformation of existing data
 - Human-AI collaboration
 - Symbolic generation

Tutorial Summary (cont'd)

- How do we use the synthetic data?
 - Pretraining
 - Supervised finetuning
 - RL training
 - Evaluation & Analysis
- Scenario-specific applications
 - Reasoning
 - Code generation
 - Tool use & agents
 - Multilingual & multimodal

Synthetic vs Real-world data

- Real-world data refers to data produced by **real users when interacting with real products.**

Synthetic vs Real-world data

- Real-world data refers to data produced by **real users when interacting with real products.**

 OpenAI Newsroom  
@OpenAINewsroom

Fresh numbers shared by [@sama](#) earlier today:

300M weekly active ChatGPT users

1B user messages sent on ChatGPT every day

1.3M devs have built on OpenAI in the US

10:17 AM · Dec 4, 2024 · 345.9K Views

Sundar Pichai  
@sundarpichai

The world is adopting AI faster than ever before.

This time last year we were processing 9.7 trillion tokens a month across our products and APIs.

Today, that number is 480 trillion. That's a 50X increase in just a year. 🐻



Monthly Tokens Processed
Across our surfaces

9.7T

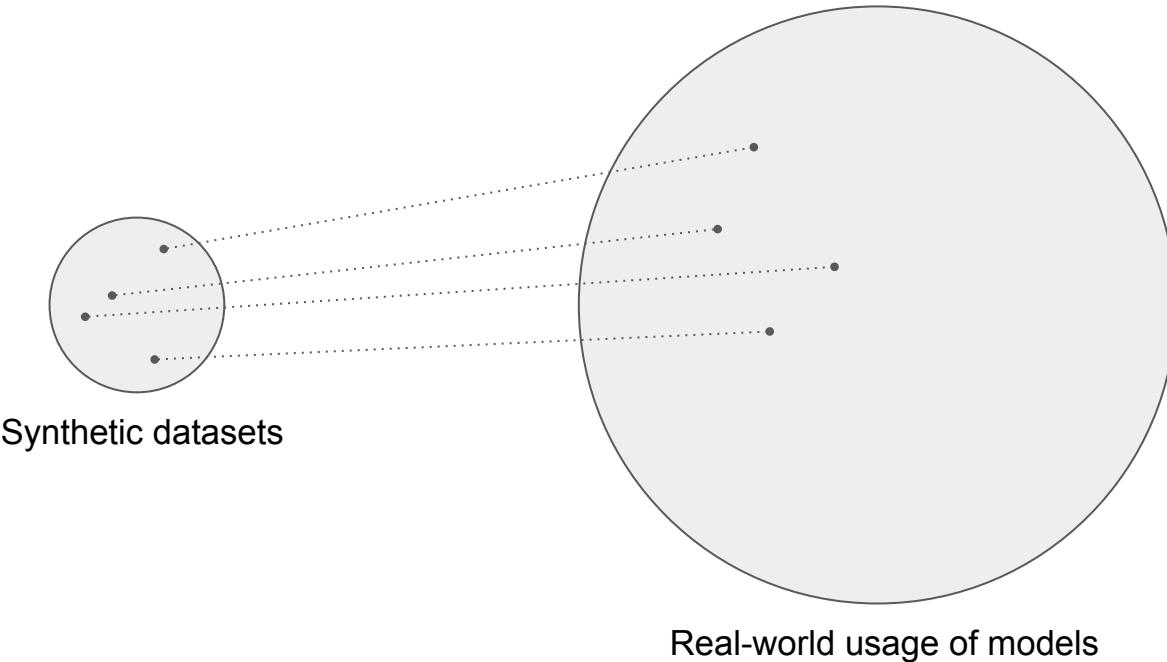
~50x

480T+

12:27 PM · May 20, 2025 · 173.3K Views

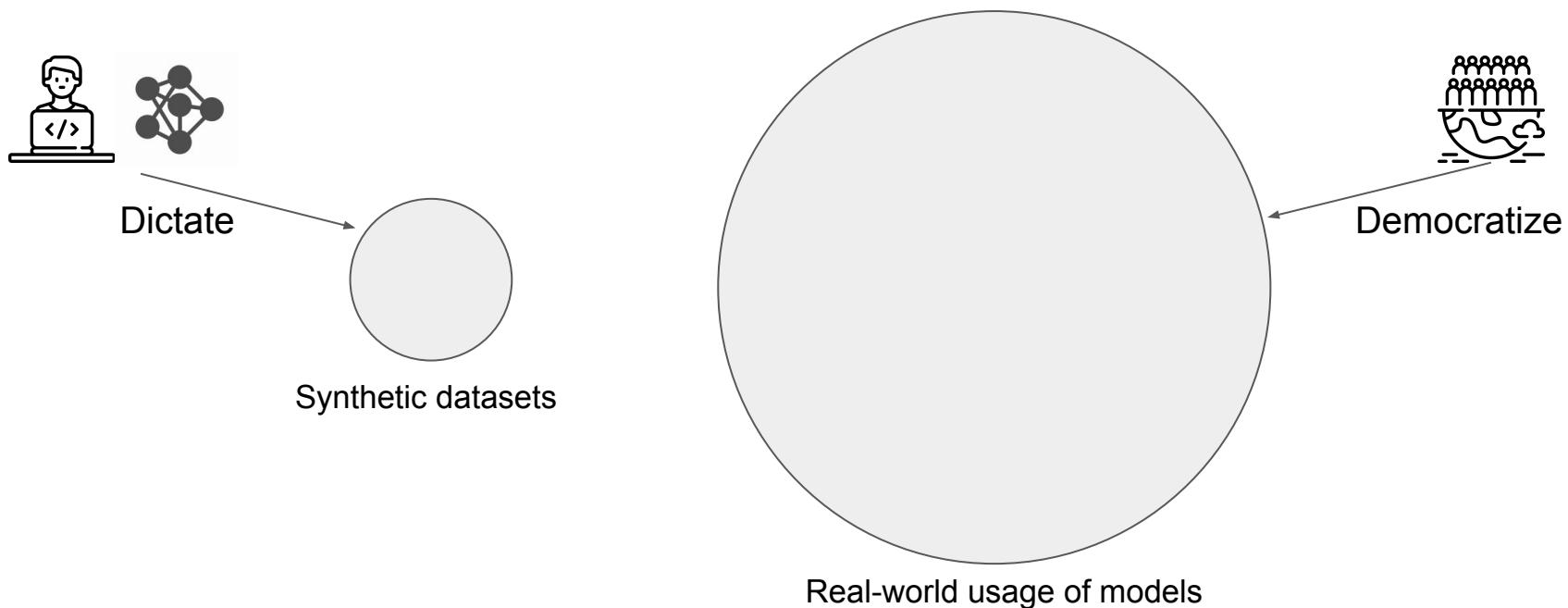
Synthetic vs Real-world data

- There is still a significant **gap in size, diversity, & distribution!**



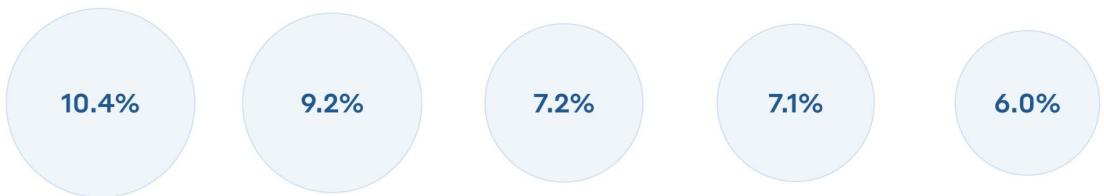
Synthetic vs Real-world data

- Fundamentally, they are produced differently.

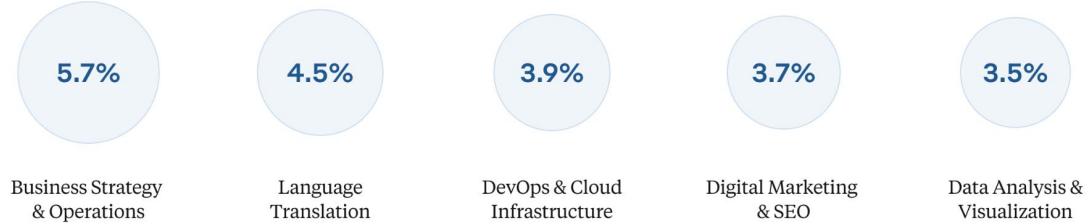


Understanding real-world usage

- Anthropic Clio [[Tamkin et al., 2024](#)]



Top use cases on Claude.ai



Understanding real-world usage

- Anthropic Clio [[Tamkin et al., 2024](#)]
- **Open** data: WildChat [[Zhao et al., 2024](#)]
- **Open** data: LMSys-Chat-1M [[Zheng et al., 2024](#)]
- **Open-source tool** for analysis: EvalTree [[Zeng et al., 2025](#)]
- It's very hard to capture the nuances and richness of real-world data.

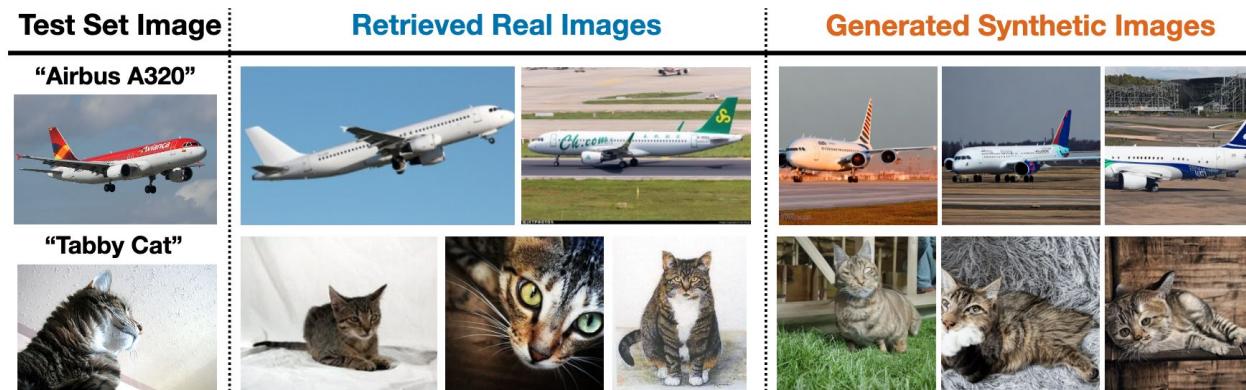
**How does synthetic data compare to real data
in terms of quality?**

How does synthetic data compare to real data in terms of quality?

- In a controlled setup, synthetic data still underperform real data (if available) [[Geng et al., 2024](#)].

How does synthetic data compare to real data in terms of quality?

- In a controlled setup, synthetic data still underperform real data (if available) [Geng et al., 2024].
- Analysis shows synthetic data often contains generator artifacts and distort class-level visual content.



How can we measure the quality of data?

- **Quality control** is a critical step in classic data pipeline
[[Daniel et al., 2018](#)]
- Most synthetic datasets from the community ***do not implement quality checks.***

How can we measure the quality of data?

- There are two common proxies for synthetic data quality:
 - Better **downstream performance** of the trained model implies better quality of the data.
 - Distilling from **stronger generation models** (e.g., GPT4) produces data of better quality.

How can we measure the quality of data?

- There are two common proxies for synthetic data quality:
 - Better **downstream performance** of the trained model implies better quality of the data.
 - Distilling from **stronger generation models** (e.g., GPT4) produces data of better quality.
- Both of them are not always true, as there are many moving factors in the generator, data, and downstream tasks [[Kim et al. 2025](#)].

Ideally, we also want to measure instance-level quality

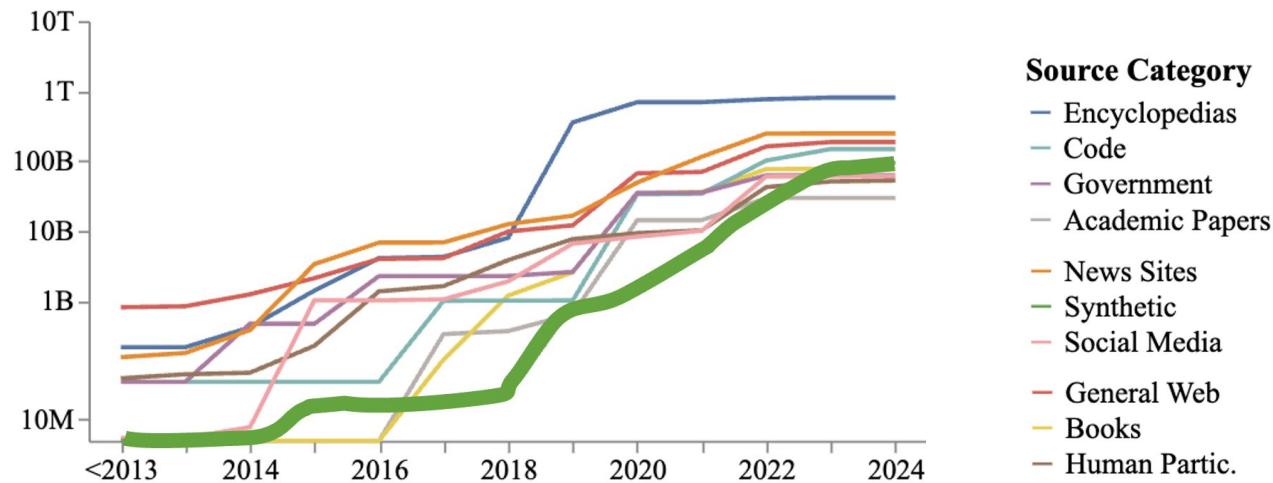
- Reward models
- LLM as a judge
- Rule-based verification
- Decomposition [[Min et al., 2023](#); [Li et al., 2024](#)]
- However, building a **generalist verifier** is hard! [[Sutton, 2001](#)]

What if synthetic data scales up?

- AI models will be trained on **increasing** amount of model-generated data, inevitably.
 - Model builders intentionally add them.
 - Model-generated content populates on the Internet

What if synthetic data scales up?

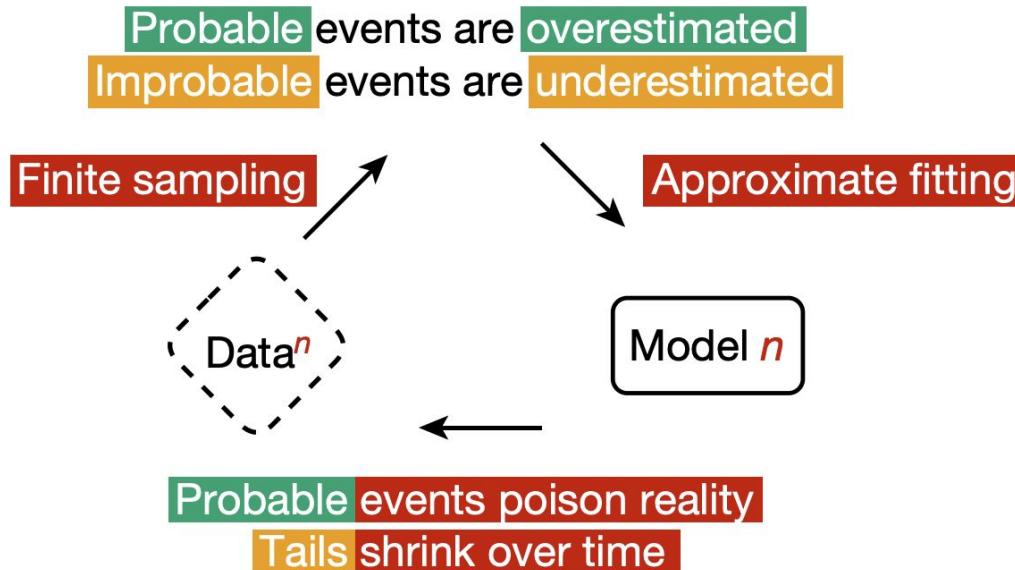
- AI models will be trained on **increasing** amount of model-generated data, inevitably.



The cumulative size of text data from different sources for post-training LMs
[[Longpre et al., 2024](#)]

Model collapse or self-improving

- AI models collapse when trained on recursively generated data [[Shumailov et al., 2023](#)].



Model collapse or self-improving

- AI models collapse when trained on recursively generated data [[Shumailov et al., 2023](#)].
- AI models can **self**-instruct [[Wang et al., 2022](#)], **self**-improve [[Huang et al., 2022](#)], **self**-refine [[Madaan et al., 2023](#)], **self**-reward [[Yuan et al., 2024](#)], etc.

Model collapse or self-improving

- AI models collapse when trained on recursively generated data [Shumailov et al., 2023].



What makes the difference?

- AI models can self-instruct [Wang et al., 2022], self-improve [Huang et al., 2022], self-refine [Madaan et al., 2023], self-reward [Yuan et al., 2024], etc.

Bringing in additional information

- Human selection, editing, & supervision.
- Human prior in the generation pipeline design.
 - E.g., prompts, principles for filtering, etc.
- Grounded documents, retrieved information, tool results.
- Rewards from interacting with environments.
- ...

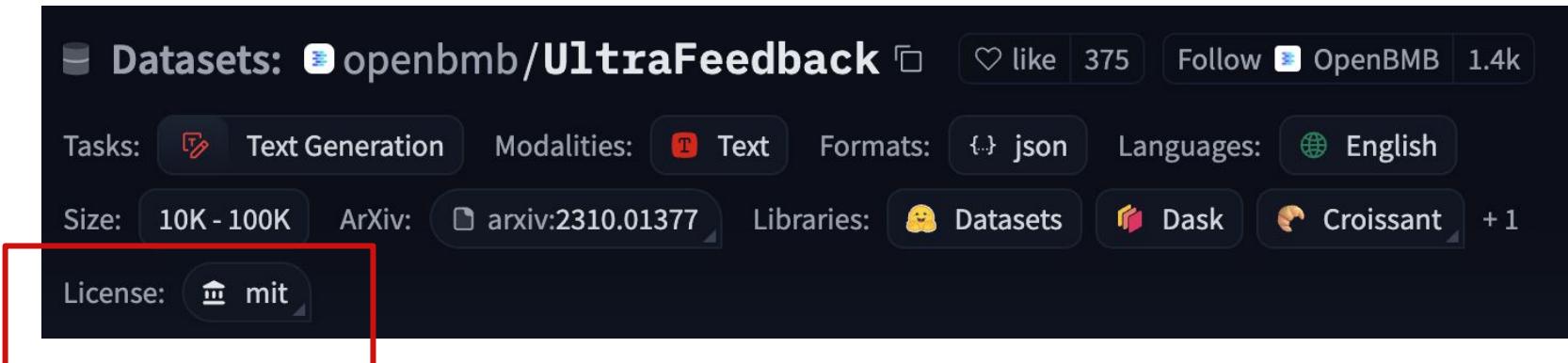
Bringing in additional information

- Human selection, editing, & supervision.
- Human prior in the generation pipeline design.
 - E.g., prompts, principles for filtering, etc.
- Grounded documents, retrieved information, tool results.
- Rewards from interacting with environments.
- ...
- Synthetic data **accelerates the combination & transformation** of useful information.

What licensing is proper for synthetic data?

What licensing is proper for synthetic data?

- Ultrafeedback as an example: A widely-used synthetic preference dataset with MIT license [[Cui et al., 2024](#)]



What licensing is proper for synthetic data?

- The licensing of synthetic datasets is tricky:
 - The dataset might be transformed from **other less permissive datasets**.

| Source datasets of Ultrafeedback | License |
|----------------------------------|-------------------------------------|
| evol_instruct | MIT |
| false_qa | Unclear |
| flan | Apache 2.0 only for generation code |
| sharegpt | Unclear |
| truthful_qa | Apache 2.0 |
| ultrachat | MIT |

What licensing is proper for synthetic data?

- The licensing of synthetic datasets is tricky:
 - The dataset might be transformed from other less permissive datasets.
 - A lot of generators rely on **distillation-unfriendly models**.

What licensing is proper for synthetic data?

DeepSeek ToS

“ You may apply the Inputs and Outputs of the Services to a wide range of use cases, including personal use, academic research, derivative product development, **training other models (such as model distillation)**, etc. ”

What licensing is proper for synthetic data?

OpenAI ToS

“ You may not use our Services for any illegal, harmful, or abusive activity. For example, you may not:

...

Use output to develop models that compete with OpenAI. ”

Anthropic ToS

“ You may not access or use our services

...

To develop any products or services that compete with our Services, including to develop or train any AI or ML algorithms or models or resell the Services. ”

Gemini API ToS

“ You may not use the Services **to develop models that compete with the Services** (e.g., Gemini API or Google AI Studio).”

What licensing is proper for synthetic data?

- It's not clear about the ownership/copyright of AI-generated content [[more guidelines from US Copyright Office](#)].
- Technically, how to detect and enforce whether one model's outputs are used to train another model?

Summary of open problems

- **Diversity & distribution:** understand the richness of real-world data and close the gap.
- **Quality:** develop quality checks/validation methods to promote high-quality data.
- **New information:** bring in additional information in model self-improving, and avoid model collapse.
- **Licensing & copyright:** call for lawful guidance on the use of synthetic data, and the community for technical innovation and ethical practice.
- **Many more ...**