

SELF-INSTRUCT: Aligning Language Models with Self-Generated Instructions

Yizhong Wang[♣] Yeganeh Kordi[◇] Swaroop Mishra[♡] Alisa Liu[♣]
Noah A. Smith^{♣+} Daniel Khashabi[♣] Hannaneh Hajishirzi^{♣+}
[♣]University of Washington [◇]Tehran Polytechnic [♡]Arizona State University
[♣]Johns Hopkins University ⁺Allen Institute for AI
yizhongw@cs.washington.edu

WizardLM: Empowering Large Language Models to Follow Complex Instructions

Can Xu^{1*} Qingfeng Sun^{1*} Kai Zheng^{1*} Xiubo Geng¹
Pu Zhao¹ Jiazhan Feng^{2†} Chongyang Tao¹ Qingwei Lin¹ Daxin Jiang^{1‡}
¹Microsoft
²Peking University
{caxu,qins,zhengkai,xigeng,puzhao,chongyang.tao,qlin,djiang}@microsoft.com
{fengjiazhan}@pku.edu.cn

LARGE LANGUAGE MODELS CAN SELF-IMPROVE

Jiaxin Huang^{1*} Shixiang Shane Gu² Le Hou^{2†} Yuexin Wu² Xuezhi Wang²
Hongkun Yu² Jiawei Han¹
¹University of Illinois at Urbana-Champaign ²Google
¹{jiaxinh3, hanj}@illinois.edu ²{shanegu, lehou, crickwu,
xuezhiw, hongkunyu}@google.com

Presenter: Jeongwoo Lee

SELF-INSTRUCT: Aligning Language Models with Self-Generated Instructions

Yizhong Wang♣ Yeganeh Kordi◇ Swaroop Mishra♡ Alisa Liu♣
Noah A. Smith♣+ Daniel Khashabi♣ Hannaneh Hajishirzi♣+

♣University of Washington ◇Tehran Polytechnic ♡Arizona State University

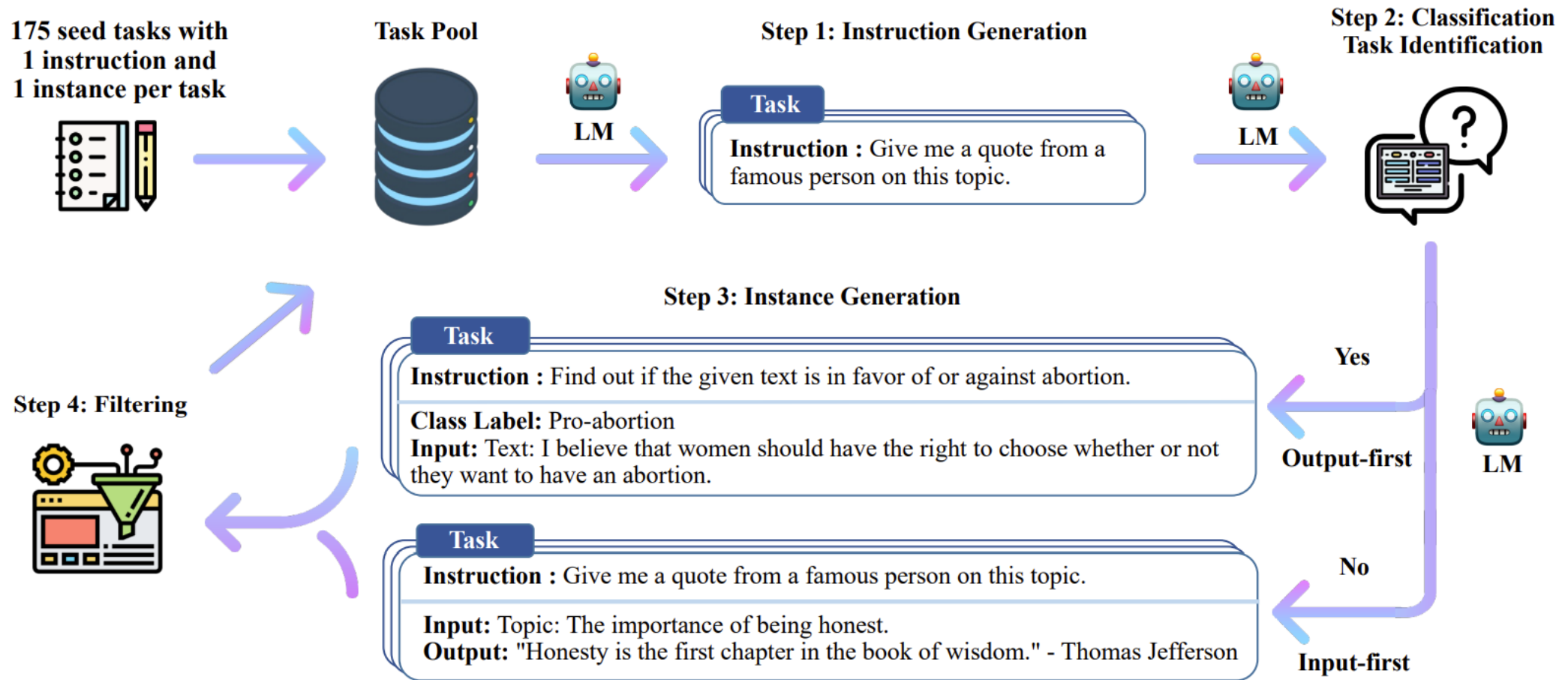
♣Johns Hopkins University +Allen Institute for AI

yizhongw@cs.washington.edu

Introduction

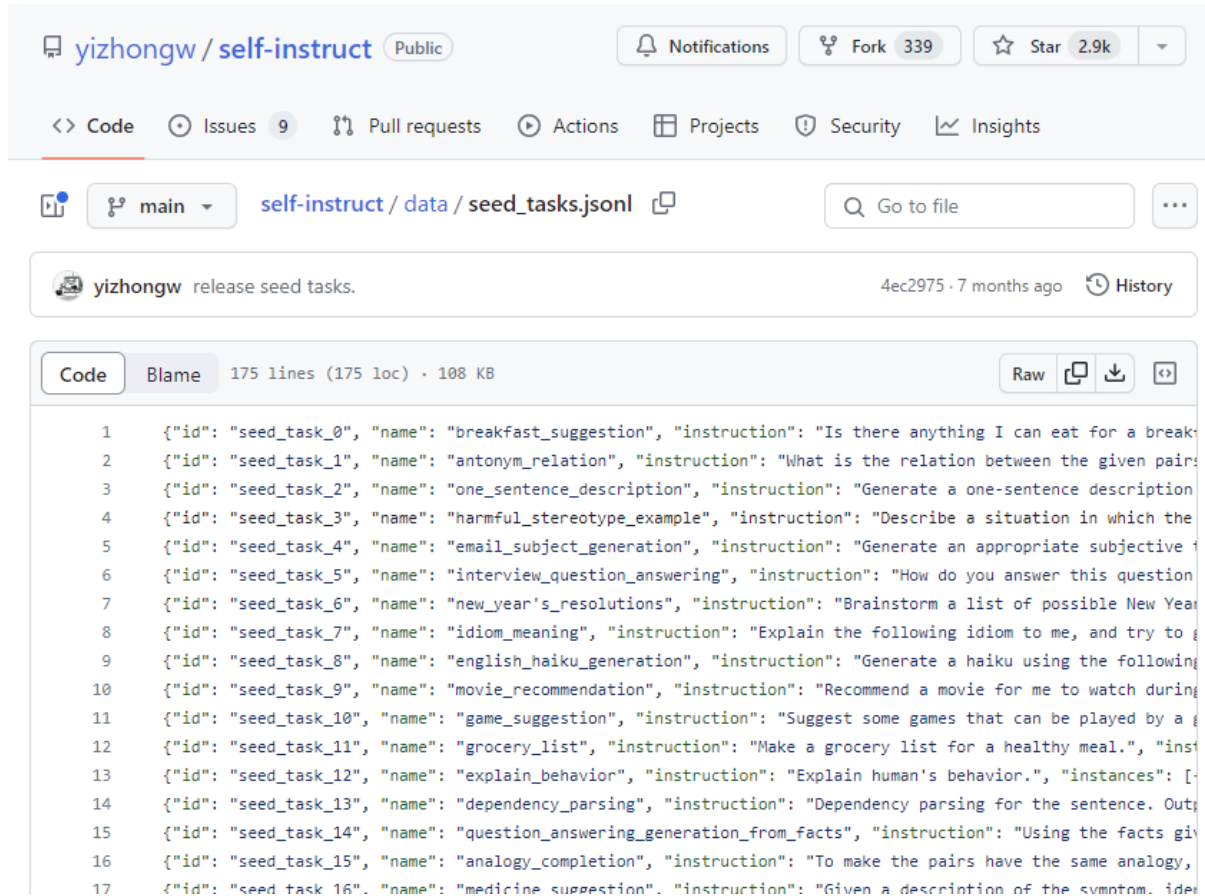
- LLM은 사람이 만든 학습 데이터의 품질과 양에 크게 의존하는데, 사람이 만든 데이터의 경우 비용이 많이 들고 다양성, 창의성을 제한받을 수 있다는 한계가 있음
- Instruction을 LLM을 통해 생성하고 그걸로 instruction tuning을 해도 성능이 좋아짐을 보여주는 연구
- 사람의 개입을 최소화하고 Instruction과 Instance를 LLM이 생성하도록 한 후 이를 통해 vanilla GPT3를 학습하였더니 InstructGPT 모델과 비슷한 수준의 성능을 냄

Method



Method

• Step 1: Instruction Generation



```
1 {"id": "seed_task_0", "name": "breakfast_suggestion", "instruction": "Is there anything I can eat for a breakf
2 {"id": "seed_task_1", "name": "antonym_relation", "instruction": "What is the relation between the given pair
3 {"id": "seed_task_2", "name": "one_sentence_description", "instruction": "Generate a one-sentence description
4 {"id": "seed_task_3", "name": "harmful_stereotype_example", "instruction": "Describe a situation in which the
5 {"id": "seed_task_4", "name": "email_subject_generation", "instruction": "Generate an appropriate subjective t
6 {"id": "seed_task_5", "name": "interview_question_answering", "instruction": "How do you answer this question
7 {"id": "seed_task_6", "name": "new_year's_resolutions", "instruction": "Brainstorm a list of possible New Year
8 {"id": "seed_task_7", "name": "idiom_meaning", "instruction": "Explain the following idiom to me, and try to e
9 {"id": "seed_task_8", "name": "english_haiku_generation", "instruction": "Generate a haiku using the following
10 {"id": "seed_task_9", "name": "movie_recommendation", "instruction": "Recommend a movie for me to watch during
11 {"id": "seed_task_10", "name": "game_suggestion", "instruction": "Suggest some games that can be played by a e
12 {"id": "seed_task_11", "name": "grocery_list", "instruction": "Make a grocery list for a healthy meal.", "inst
13 {"id": "seed_task_12", "name": "explain_behavior", "instruction": "Explain human's behavior.", "instances": [-
14 {"id": "seed_task_13", "name": "dependency_parsing", "instruction": "Dependency parsing for the sentence. Outp
15 {"id": "seed_task_14", "name": "question_answering_generation_from_facts", "instruction": "Using the facts giv
16 {"id": "seed_task_15", "name": "analogy_completion", "instruction": "To make the pairs have the same analogy,
17 {"id": "seed task 16". "name": "medicine suggestion". "instruction": "Given a description of the symptom, ide
```

Come up with a series of tasks:

- Task 1: {instruction for existing task 1}
- Task 2: {instruction for existing task 2}
- Task 3: {instruction for existing task 3}
- Task 4: {instruction for existing task 4}
- Task 5: {instruction for existing task 5}
- Task 6: {instruction for existing task 6}
- Task 7: {instruction for existing task 7}
- Task 8: {instruction for existing task 8}
- Task 9:

Method

- Step 2: Classification Task Identification

```
Can the following task be regarded as a classification task with finite output labels?  
  
Task: Given my personality and the job, tell me if I would be suitable.  
Is it classification? Yes  
  
Task: Give me an example of a time when you had to use your sense of humor.  
Is it classification? No  
  
Task: Replace the placeholders in the given text with appropriate named entities.  
Is it classification? No  
  
Task: Fact checking - tell me if the statement is true, false, or unknown, based on your  
knowledge and common sense.  
Is it classification? Yes  
  
Task: Return the SSN number for the person.  
Is it classification? No  
  
Task: Detect if the Reddit thread contains hate speech.  
Is it classification? Yes  
  
...  
  
Task: To make the pairs have the same analogy, write the fourth word.  
Is it classification? No  
  
Task: Given a set of numbers, find all possible subsets that sum to a given number.  
Is it classification? No  
  
Task: {instruction for the target task}
```

Method

- Step 3: Instance Generation (Input-first Approach)

```
Come up with examples for the following tasks. Try to generate multiple examples when possible.
If the task doesn't require additional input, you can generate the output directly.

Task: Which exercises are best for reducing belly fat at home?
Output:
- Lying Leg Raises
- Leg In And Out
- Plank
- Side Plank
- Sit-ups

Task: Extract all the country names in the paragraph, list them separated by commas.
Example 1
Paragraph: Dr. No is the sixth novel by the English author Ian Fleming to feature his British
Secret Service agent James Bond. Written at Fleming's Goldeneye estate in Jamaica, it was
first published in the United Kingdom by Jonathan Cape in 1958. In the novel Bond looks into
the disappearance in Jamaica of two fellow MI6 operatives who had been investigating Doctor
No. Bond travels to No's Caribbean island and meets Honeychile Rider, who is there to collect
shells. They are captured and taken to a luxurious facility carved into a mountain. The
character of Doctor No, the son of a German missionary and a Chinese woman, was influenced by
Sax Rohmer's Fu Manchu stories. Dr. No was the first of Fleming's novels to face widespread
negative reviews in Britain, but it was received more favourably in the United States.
Output: English, British, Jamaica, the United Kingdom, German, Chinese, Britain, the United
States.

Task: Converting 85 F to Celsius.
Output: 85°F = 29.44°C

Task: Sort the given list ascendingly.
Example 1
List: [10, 92, 2, 5, -4, 92, 5, 101]
Output: [-4, 2, 5, 5, 10, 92, 92, 101]
Example 2
Input 2 - List: [9.99, 10, -5, -1000, 5e6, 999]
Output: [-1000, -5, 9.99, 10, 999, 5e6]

...

Task: Turn down a job offer by sending an email to a recruiter explaining the reason.
Output: Hi [Recruiter],
Thank you so much for the generous offer to join your team. As we discussed, I've admired the
company for a number of years, and am a proud endorser of its products. However, after further
consideration of where I currently am in my career, I've decided to accept an offer at another
company.
I would love to stay in touch with you and have already started following you on [Social Media
Platform]. Again, thank you so much for your time and consideration.
Thanks again,
[Your Name]

Task: {Instruction for the target task}
```

Method

- **Step 3: Instance Generation (Output-first Approach)**

```
Given the classification task definition and the class labels, generate an input that
corresponds to each of the class labels. If the task doesn't require input, just generate the
correct class label.

Task: Classify the sentiment of the sentence into positive, negative, or mixed.
Class label: mixed
Sentence: I enjoy the flavor of the restaurant but their service is too slow.
Class label: Positive
Sentence: I had a great day today. The weather was beautiful and I spent time with friends.
Class label: Negative
Sentence: I was really disappointed by the latest superhero movie. I would not recommend it.

Task: Given a dialogue, classify whether the user is satisfied with the service. You should
respond with "Satisfied" or "Unsatisfied".
Class label: Satisfied
Dialogue:
- Agent: Thank you for your feedback. We will work to improve our service in the future.
- Customer: I am happy with the service you provided. Thank you for your help.
Class label: Unsatisfied
Dialogue:
- Agent: Sorry that we will cancel your order. You will get a refund within 7 business days.
- Customer: oh that takes too long. I want you to take quicker action on this.

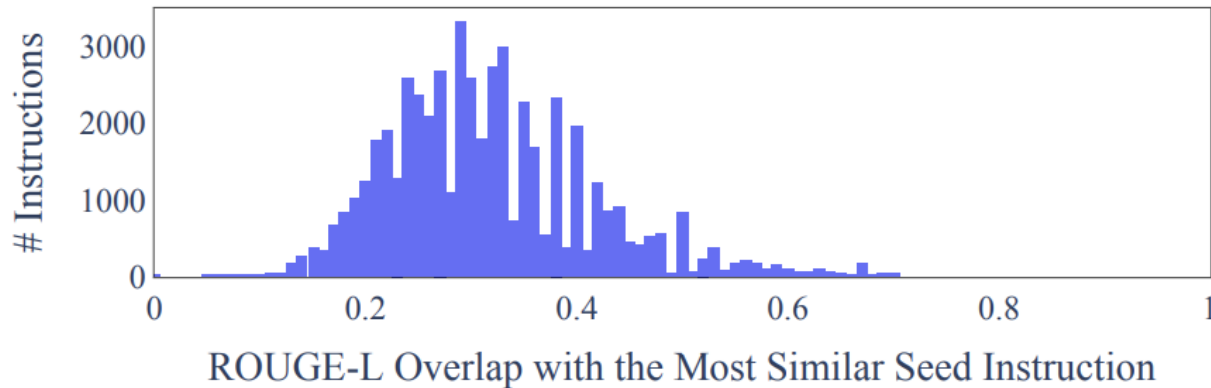
...

Task: Which of the following is not an input type? (a) number (b) date (c) phone number (d)
email address (e) all of these are valid inputs.
Class label: (e)

Task: {instruction for the target task}
```


Method

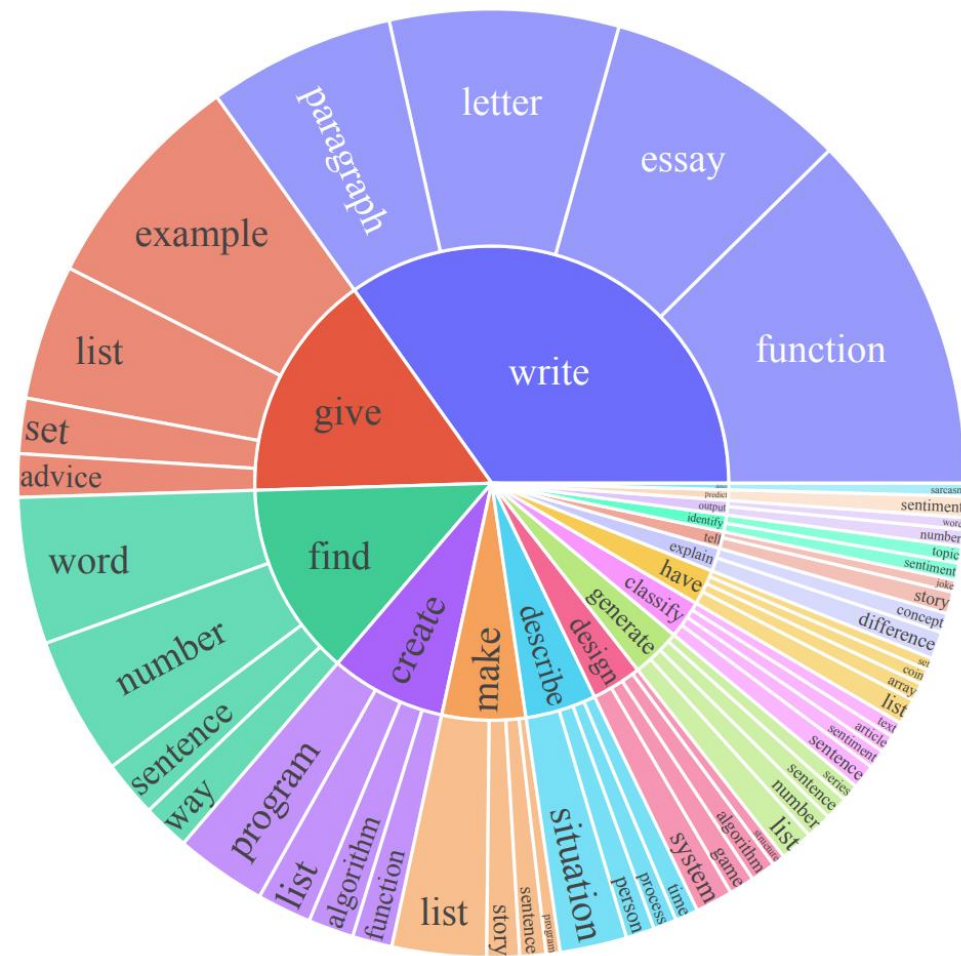
- Step 4: Filtering



- 기존 Task와 다른 Task들이 Task Pool에 추가되어야 하기 때문에 현재 Task Pool에 존재하는 Instruction들과의 ROUGE-L Score가 0.7보다 작은 Instruction만 Task Pool에 추가함
- Image, Picture 등 Language model이 할 수 없는 task라면 해당 keyword가 있는 Instruction은 제외함

SELF-INSTRUCT Data from GPT3

statistic	
# of instructions	52,445
- # of classification instructions	11,584
- # of non-classification instructions	40,861
# of instances	82,439
- # of instances with empty input	35,878
ave. instruction length (in words)	15.9
ave. non-empty input length (in words)	12.7
ave. output length (in words)	18.9



Experiments

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

Table 3: Evaluation results on *unseen* tasks from SUPERNI (§4.3). From the results, we see that ① SELF-INSTRUCT can boost GPT3 performance by a large margin (+33.1%) and ② nearly matches the performance of InstructGPT₀₀₁. Additionally, ③ it can further improve the performance even when a large amount of labeled instruction data is present.

Limitation

- **Tail phenomena**

- Self-Instruct는 LLM에 의존하기 때문에 LLM이 가지는 모든 한계점을 상속받음
- 즉, LM의 큰 성능 향상은 많이 사용하는 Language, Task로부터 주로 오게 되고, 자주 사용하지 않는 Language, Task는 적은 성능 향상을 얻음

- **Dependence on large models**

- Self-Instruct의 Large model에 대한 의존성때문에 적은 computing resource를 갖는 사람들은 사용에 어려움이 있을 수 있음

- **Reinforcing LM biases**

- 반복적인 알고리즘으로 인해 의도와 다르게 문제가 되는 social bias를 증폭시킬 수 있음

***WizardLM*: Empowering Large Language Models to Follow Complex Instructions**

Can Xu^{1*} Qingfeng Sun^{1*} Kai Zheng^{1*} Xiubo Geng¹
Pu Zhao¹ Jiazhan Feng^{2†} Chongyang Tao¹ Qingwei Lin¹ Daxin Jiang^{1‡}

¹Microsoft

²Peking University

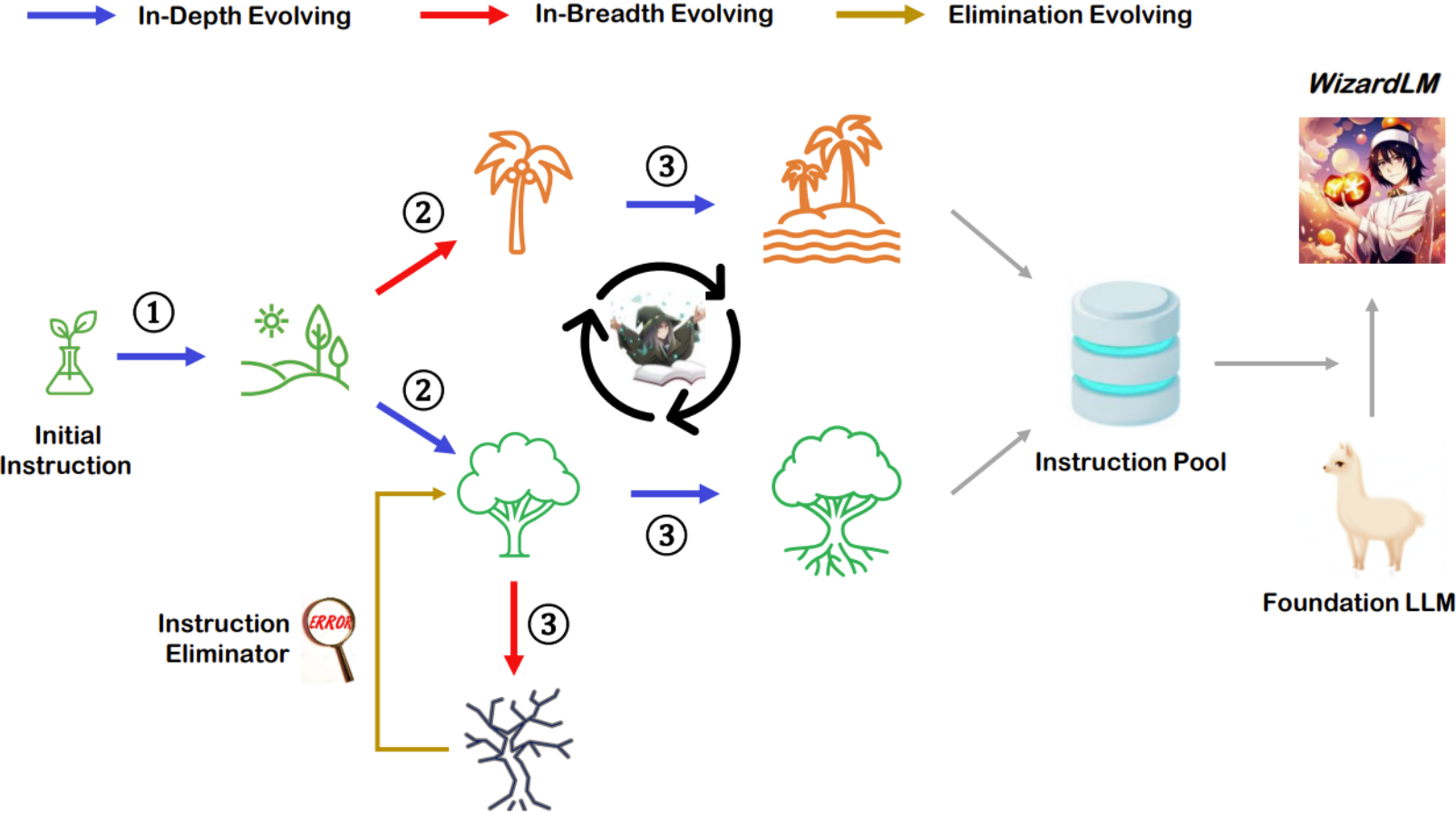
{caxu,qins,zhengkai,xigeng,puzhao,chongyang.tao,qlin,djiang}@microsoft.com
{fengjiazhan}@pku.edu.cn

Introduction

- LLM을 학습시키기 위해서는 양질의 데이터가 중요한데, 사람이 데이터를 만드는 작업은 시간과 비용이 매우 큼
- LLM을 활용해 데이터를 구축하는데 드는 시간과 비용을 절약하고자 함
- 간단한 질문으로 시작해서 다양한 난이도와 주제의 질문으로 Instruction을 진화시키는 것이 이 논문의 핵심 아이디어임

Approach

- Evol-Instruct



Approach

• In-Depth Evolving

- Add Constraints, Deepening, Concretizing, Increase Reasoning, Complicate Input을 통해서 복잡하고 어려운 질문을 만드는 것이 목표임

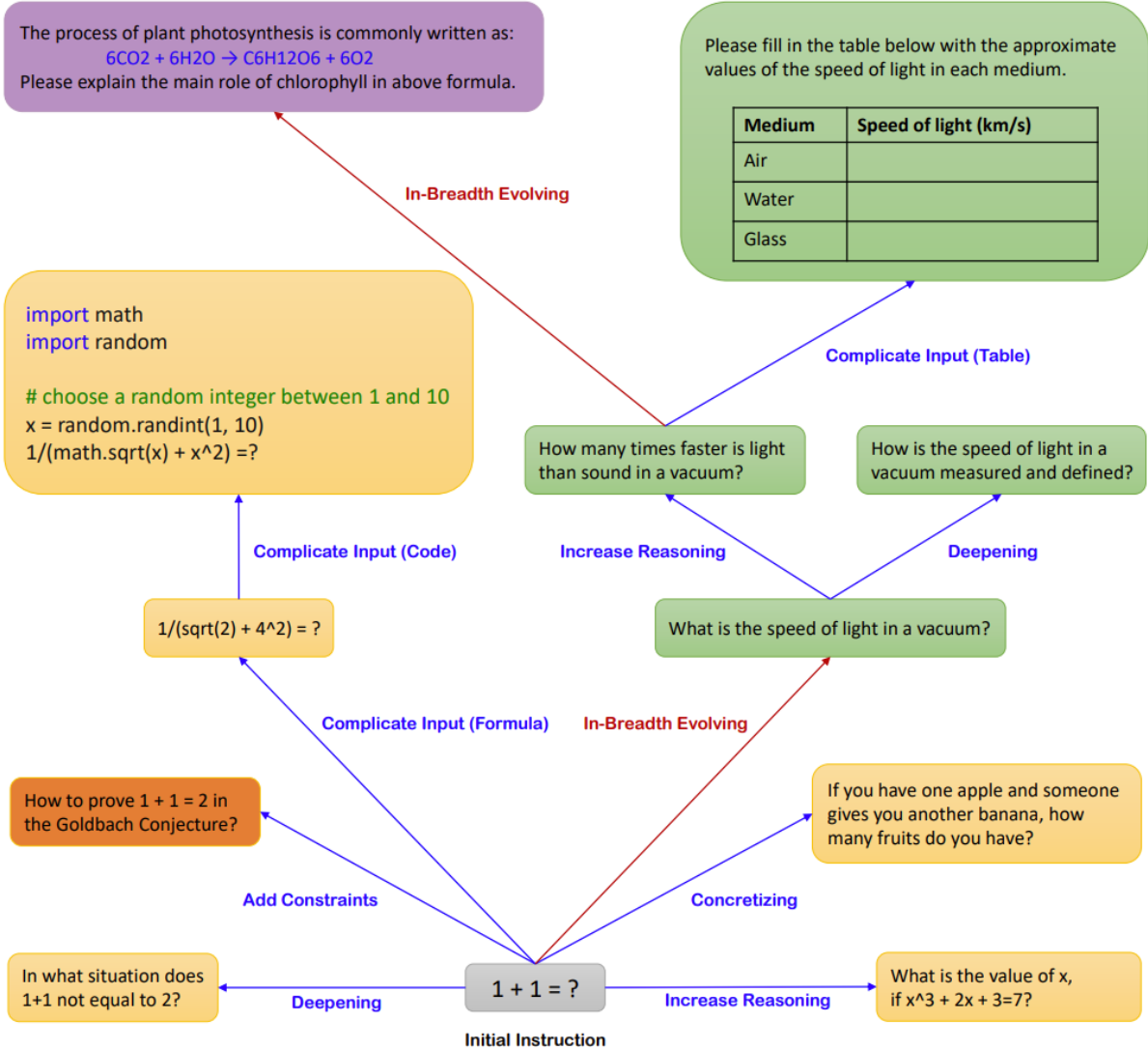
```
I want you act as a Prompt Rewriter.
Your objective is to rewrite a given prompt into a more complex version to make those famous AI systems
(e.g., ChatGPT and GPT4) a bit harder to handle.
But the rewritten prompt must be reasonable and must be understood and responded by humans.
Your rewriting cannot omit the non-text parts such as the table and code in #Given Prompt#. Also, please
do not omit the input in #Given Prompt#.
You SHOULD complicate the given prompt using the following method:
Please add one more constraints/requirements into #Given Prompt#
You should try your best not to make the #Rewritten Prompt# become verbose, #Rewritten Prompt# can only
add 10 to 20 words into #Given Prompt#.
'#Given Prompt#', '#Rewritten Prompt#', 'given prompt' and 'rewritten prompt' are not allowed to appear in
#Rewritten Prompt#
#Given Prompt#:
<Here is instruction.>
#Rewritten Prompt#:
```

• In-Breadth Evolving

- 주제, Task 등의 측면에서 다양한 질문을 만드는 것이 목표임

```
I want you act as a Prompt Creator.
Your goal is to draw inspiration from the #Given Prompt# to create a brand new prompt.
This new prompt should belong to the same domain as the #Given Prompt# but be even more rare.
The LENGTH and difficulty level of the #Created Prompt# should be similar to that of the #Given Prompt#.
The #Created Prompt# must be reasonable and must be understood and responded by humans.
'#Given Prompt#', '#Created Prompt#', 'given prompt' and 'created prompt' are not allowed to appear in
#Created Prompt#.
#Given Prompt#:
<Here is instruction.>
#Created Prompt#:
```


Approach



Approach

- **Elimination Evolving**

- 진화한 Instruction이 기존의 명령어에 비해 더 이상 얻을 수 있는 것이 없고 비슷할 때
- 생성된 답변에 "sorry"가 포함되어 있거나 답변의 길이가 짧을 때
- 생성된 답변에 punctuation과 stop word들만 포함되어 있을 때
- "given prompt", "rewritten prompt", "#Rewritten Prompt#"와 같이 Prompt를 복사했을 때

Experiments

- **Baselines**

- ChatGPT
- Alpaca
- Vicuna

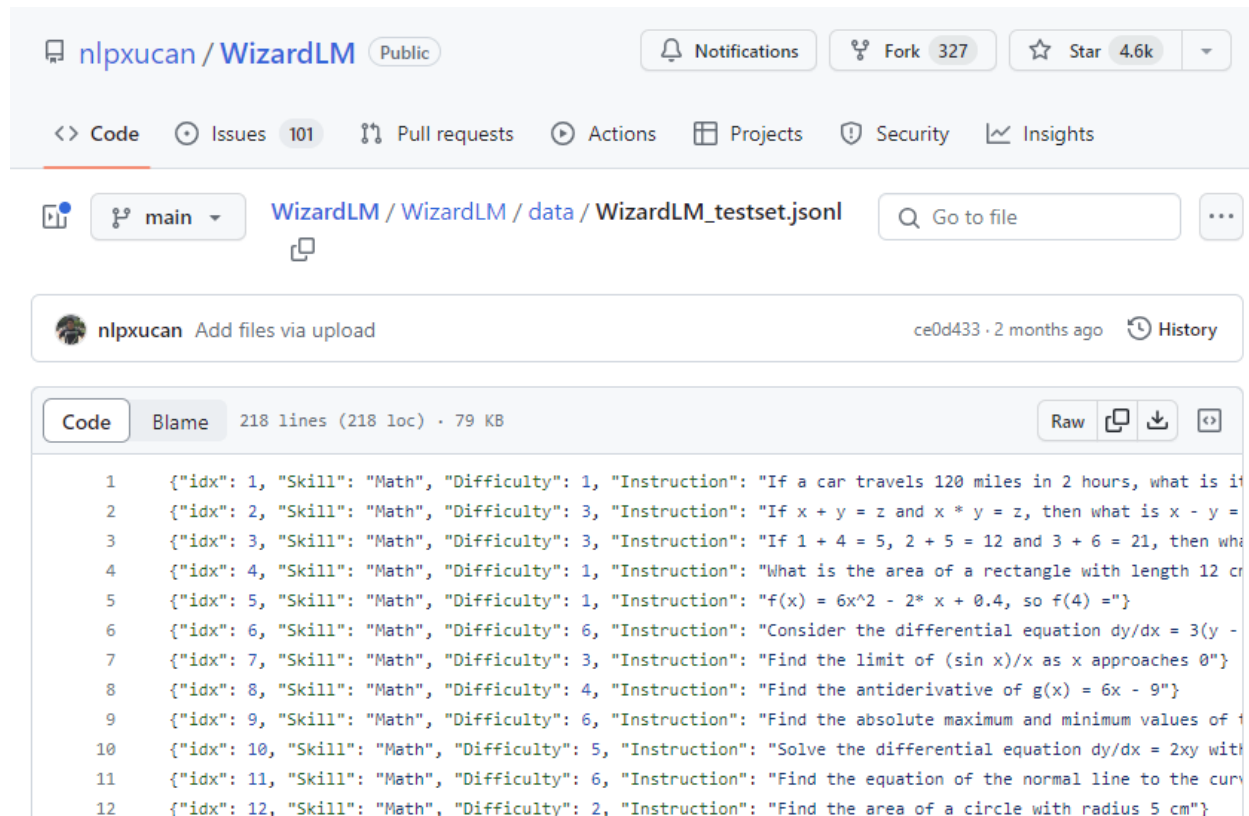
- **WizardLM**

- WizardLM을 위한 학습 데이터는 Alpaca 데이터의 Instruction을 초기 Instruction으로 해서 논문에서 제안한 Instruction 진화 과정을 반복하는 방식으로 생성함
- 이를 통해 생성된 데이터셋으로 LLaMA를 학습함

Experiments

- Testset

- 218개의 Instruction을 직접 구축해서 다양한 스킬과 다양한 난이도를 커버하려고 함



The screenshot shows a GitHub repository for 'nlpxucan / WizardLM'. The repository is public and has 327 forks and 4.6k stars. The file path is 'WizardLM / WizardLM / data / WizardLM_testset.json'. The file is 218 lines long, 79 KB, and was last updated 2 months ago. The code is displayed as a list of 12 JSON objects, each representing an instruction with an ID, skill, difficulty, and instruction text.

```
1  {"idx": 1, "Skill": "Math", "Difficulty": 1, "Instruction": "If a car travels 120 miles in 2 hours, what is its speed?"}
2  {"idx": 2, "Skill": "Math", "Difficulty": 3, "Instruction": "If  $x + y = z$  and  $x * y = z$ , then what is  $x - y =$ ?"}
3  {"idx": 3, "Skill": "Math", "Difficulty": 3, "Instruction": "If  $1 + 4 = 5$ ,  $2 + 5 = 12$  and  $3 + 6 = 21$ , then what is  $4 + 7 =$ ?"}
4  {"idx": 4, "Skill": "Math", "Difficulty": 1, "Instruction": "What is the area of a rectangle with length 12 cm and width 8 cm?"}
5  {"idx": 5, "Skill": "Math", "Difficulty": 1, "Instruction": " $f(x) = 6x^2 - 2x + 0.4$ , so  $f(4) =$ ?"}
6  {"idx": 6, "Skill": "Math", "Difficulty": 6, "Instruction": "Consider the differential equation  $dy/dx = 3(y - 1)$ . Find the general solution."}
7  {"idx": 7, "Skill": "Math", "Difficulty": 3, "Instruction": "Find the limit of  $(\sin x)/x$  as  $x$  approaches  $0$ ?"}
8  {"idx": 8, "Skill": "Math", "Difficulty": 4, "Instruction": "Find the antiderivative of  $g(x) = 6x - 9$ ?"}
9  {"idx": 9, "Skill": "Math", "Difficulty": 6, "Instruction": "Find the absolute maximum and minimum values of  $f(x) = x^3 - 3x^2 + 2x - 1$  on the interval  $[-1, 3]$ ."}
10 {"idx": 10, "Skill": "Math", "Difficulty": 5, "Instruction": "Solve the differential equation  $dy/dx = 2xy$  with the initial condition  $y(1) = 1$ ."}
11 {"idx": 11, "Skill": "Math", "Difficulty": 6, "Instruction": "Find the equation of the normal line to the curve  $y = x^2 + 2x - 3$  at the point  $(1, 0)$ ."}
12 {"idx": 12, "Skill": "Math", "Difficulty": 2, "Instruction": "Find the area of a circle with radius 5 cm?"}
```

Experiments

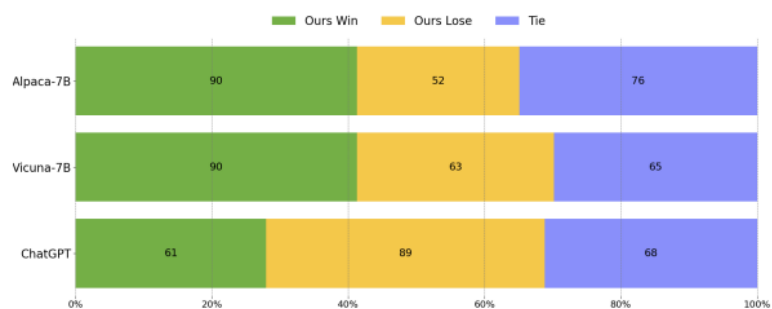
- **Human evaluation**

The annotators then judge which response is better from five aspects:

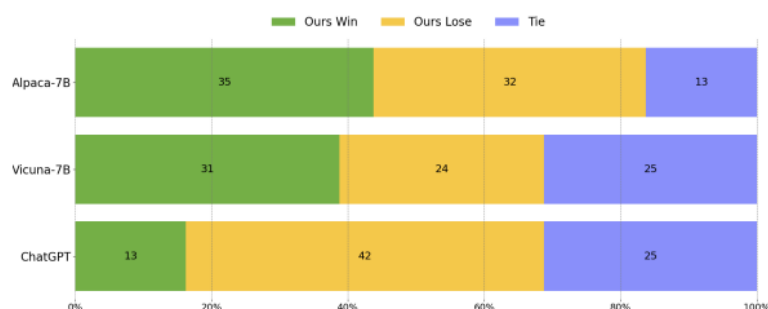
- (1) **Relevance:** Assessing the model's ability to correctly interpret the semantic meaning of the context and questions.
- (2) **Knowledgeable:** Whether the model can accurately use various and detailed knowledge for problem-solving.
- (3) **Reasoning:** Assessing the model's ability to execute correct reasoning processes or devise valid reasoning concepts to solve problems.
- (4) **Calculation:** Evaluating whether the model can perform accurate mathematical computations of the provided formulas in the domains of math, biology, chemistry and physics.
- (5) **Accuracy:** Evaluating whether the model can perform correctly in the corresponding for a given instruction.

Experiments

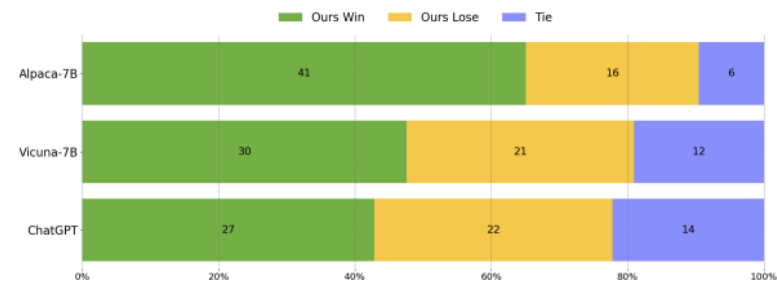
- Human evaluation



(a) *Evol-Instruct* testset



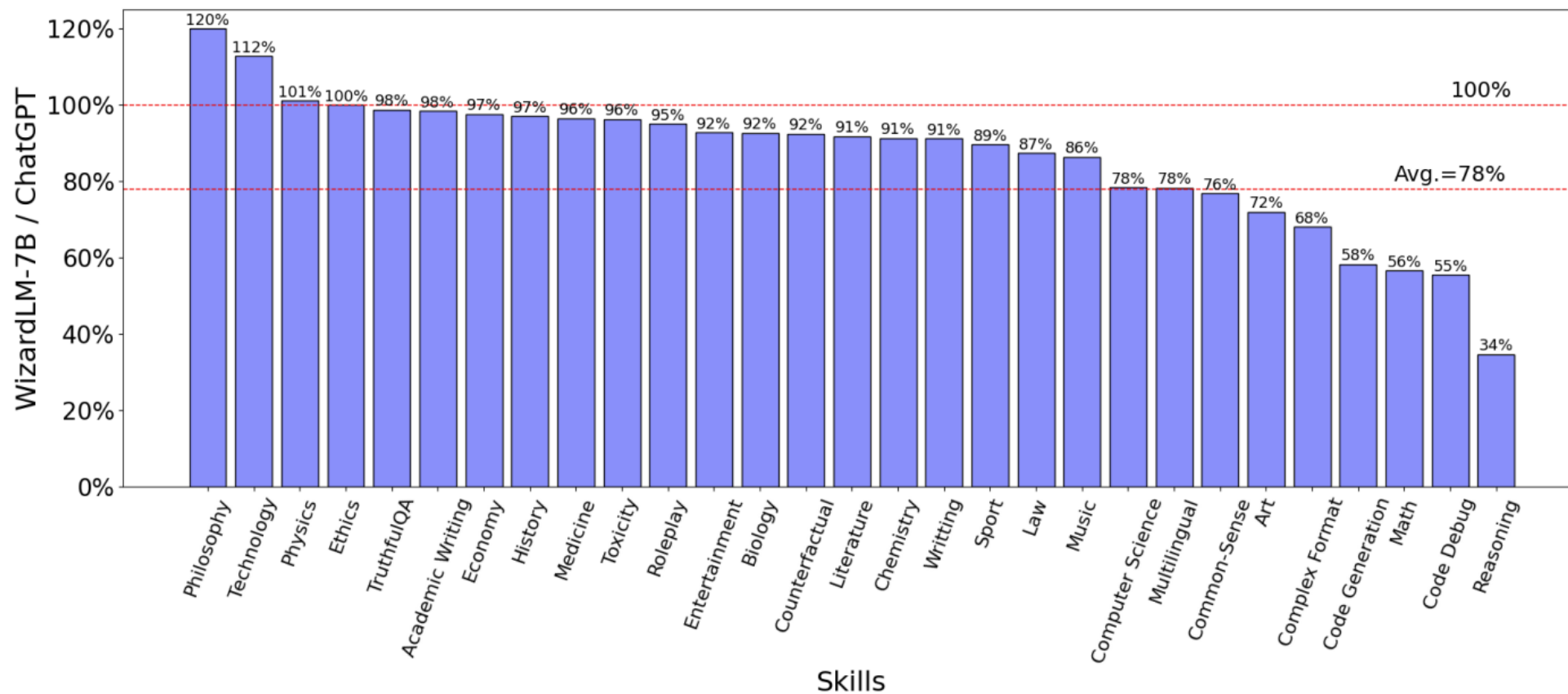
(b) Vicuna testset



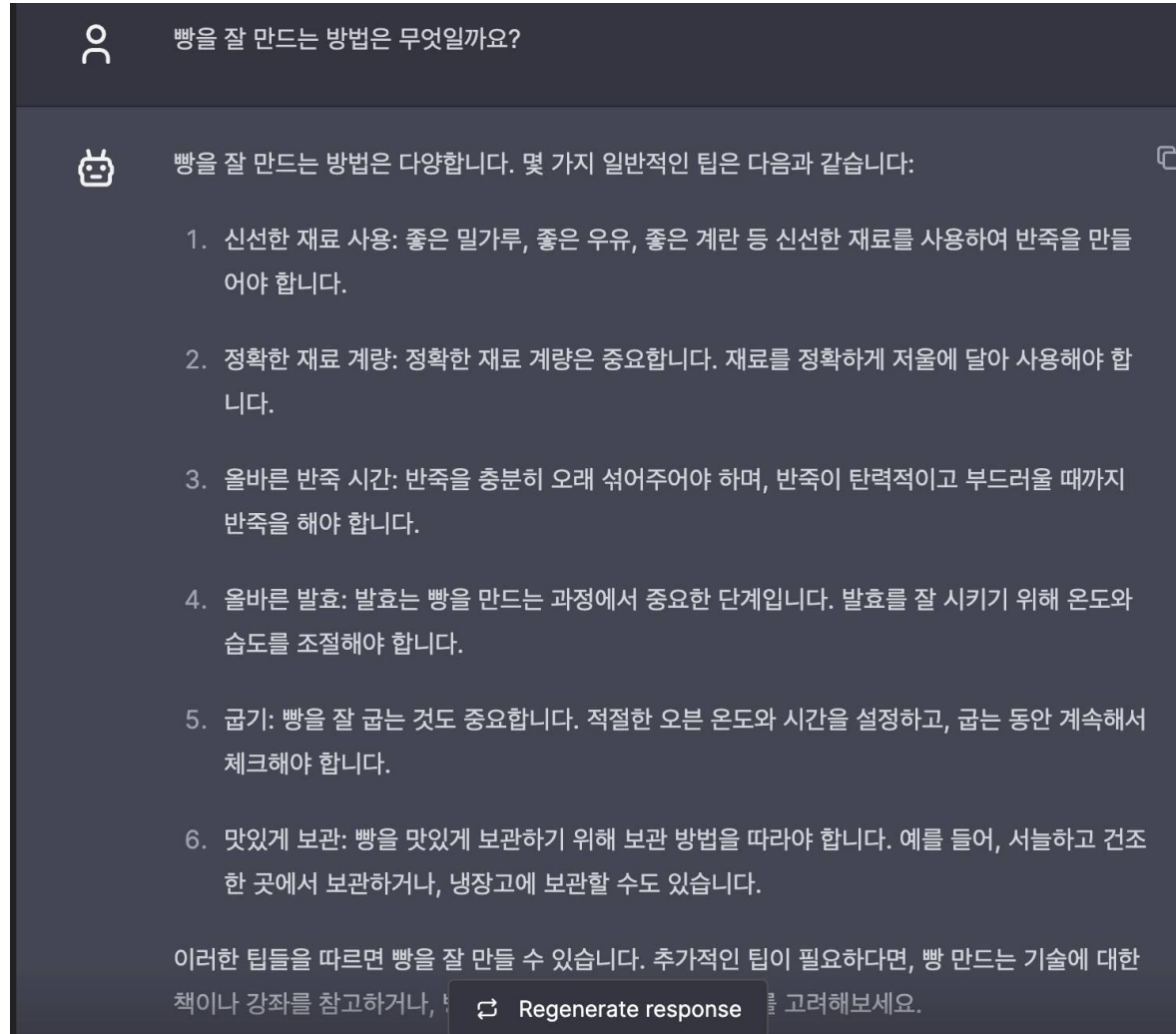
(c) *Evol-Instruct* high-difficulty

Experiments

- GPT-4 automatic evaluation





Korean WizardLM



👤 빵을 잘 만드는 방법은 무엇일까요?

🤖 빵을 잘 만드는 방법은 다양합니다. 몇 가지 일반적인 팁은 다음과 같습니다:

1. 신선한 재료 사용: 좋은 밀가루, 좋은 우유, 좋은 계란 등 신선한 재료를 사용하여 반죽을 만들어야 합니다.
2. 정확한 재료 계량: 정확한 재료 계량은 중요합니다. 재료를 정확하게 저울에 달아 사용해야 합니다.
3. 올바른 반죽 시간: 반죽을 충분히 오래 섞어주어야 하며, 반죽이 탄력적이고 부드러울 때까지 반죽을 해야 합니다.
4. 올바른 발효: 발효는 빵을 만드는 과정에서 중요한 단계입니다. 발효를 잘 시키기 위해 온도와 습도를 조절해야 합니다.
5. 굽기: 빵을 잘 굽는 것도 중요합니다. 적절한 오븐 온도와 시간을 설정하고, 굽는 동안 계속해서 체크해야 합니다.
6. 맛있게 보관: 빵을 맛있게 보관하기 위해 보관 방법을 따라야 합니다. 예를 들어, 서늘하고 건조한 곳에서 보관하거나, 냉장고에 보관할 수도 있습니다.

이러한 팁들을 따르면 빵을 잘 만들 수 있습니다. 추가적인 팁이 필요하다면, 빵 만드는 기술에 대한 책이나 강좌를 참고하거나,  Regenerate response  고려해보세요.

LARGE LANGUAGE MODELS CAN SELF-IMPROVE

Jiaxin Huang^{1*} **Shixiang Shane Gu**² **Le Hou**^{2†} **Yuexin Wu**² **Xuezhi Wang**²
Hongkun Yu² **Jiawei Han**¹

¹University of Illinois at Urbana-Champaign ²Google

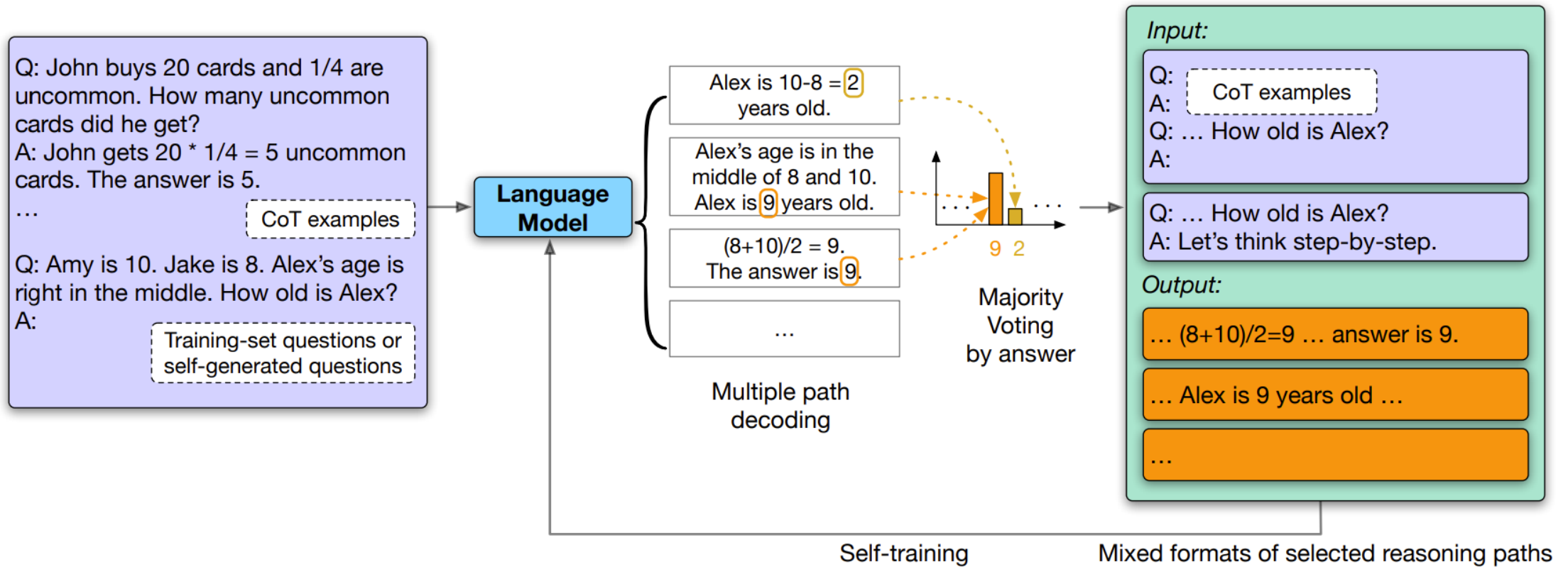
¹{jiaxinh3, hanj}@illinois.edu ²{shanegu, lehou, crickwu,
xuezhiw, hongkuny}@google.com

Introduction

- 사람은 외부의 입력없이 스스로 사고함으로써 추론 능력을 향상시킬 수 있지만, LLM을 fine-tuning하기 위해서는 광범위한 supervision이 필요함
- 사람은 어떤 문제가 주어지면 여러 번 생각하여 여러 가지 가능한 결과를 도출하고, 그 문제를 어떻게 해결해야 하는지 결론을 내린 다음, 스스로 해결책을 학습하거나 암기함
- 사람이 학습하는 방식과 유사하게, CoT를 Prompt로 사용하여 여러 prediction들을 sampling하고 Majority Voting을 사용해서 높은 confidence의 prediction들만 필터링한 후 이를 대상으로 LLM을 fine-tuning함
- 이 논문에서는 unlabeled 데이터셋만으로 LLM이 self-improving할 수 있음을 보여줌

Method

- Overview



Method

- **Generating and Filtering Multiple Reasoning Paths**

Question: Stefan goes to a restaurant with his family. They order an appetizer that costs \$10 and 4 entrees that are \$20 each. If they tip 20% of the total, what is the total amount of money that they spend?

Multiple Path Decoding:

Output 1: The appetizer costs \$10. The entrees cost \$20 each so in total $4 * \$20 = \80 . This means the total cost is $\$10 + \$80 = \$90$. They tip 20% of it, so the total amount they spend is $\$90 * 1.2 = \108 . The answer is 108.

Output 2: The appetizer costs \$10 and the entrees are \$20 each. There are 4 entrees so the sum is $\$20 * 4 = \80 . The waiter gets 20% of the total. 20% of \$80 is $\$80 * .2 = \16 . The answer is $\$80 + \$16 = \$96$. (*Incorrect reasoning path*)

Output 3: The appetizer costs \$10. The entrees cost $4 * \$20 = \80 . The tip is 20% of the total, so it is 20% of the \$90 they have spent. The tip is $0.2 * 90 = \$18$. The total they spent is $\$90 + \$18 = \$108$. The answer is 108.

Method

- Training with Mixed Formats

Question: Amy is 10 years old. Jake is 8 years old. Alex's age is right in the middle. How old is Alex?
Selected Chain-of-Thought: Amy is 10 years old. Jake is 8 years old. Alex's age is in the middle of Amy and Jake, so Alex is $(8 + 10) / 2 = 9$ years old. The answer is 9.

Mixed-formats of training data:

Format 1: Input: *[CoT prompting examples]* + '\n' + *[Question]* + '\n' + 'A:'

Output: Amy is 10 years old. Jake is 8 years old. Alex's age is in the middle of Amy and Jake, so Alex is $(8 + 10) / 2 = 9$ years old. The answer is 9.

Format 2: Input: *[Standard prompting examples]* + '\n' + *[Question]* + '\n' + 'A:'

Output: The answer is 9.

Format 3: Input: *[Question]* + '\n' + 'A: Let's think step by step.'

Output: Amy is 10 years old. Jake is 8 years old. Alex's age is in the middle of Amy and Jake, so Alex is $(8 + 10) / 2 = 9$ years old. The answer is 9.

Format 4: Input: *[Question]* + '\n' + 'A:'

Output: The answer is 9.

Method

- **Generating Questions and Prompts**

- * Question Generation

- 기존 Question 몇 개를 random으로 선택하고, input prompt로 random 순서로 concat한 후에, 언어 모델이 새로운 Question을 생성하도록 함

- * Prompt Generation

- Question과 "A: Let's think step by step."를 함께 input에 포함해서, 언어 모델이 reasoning path를 생성하도록 함

Experiments

- **Experimental Setup**

- * Dataset

- Arithmetic reasoning: GSM8K, DROP
 - Commonsense reasoning: OpenBookQA, ARC-c
 - Natural Language Inference: ANLI, ANLI-A2, ANLI-A3

- * Models, Training settings and Hyperparameters

- Pretrained LM: PaLM-540B
 - Data Sampling: DROP(5k), ANLI-A2/A3(5k)
 - Learning rate: $5e-5$
 - Batch size: 32

Experiments

- Result

	Prompting Method	GSM8K	DROP	ARC-c	OpenBookQA	ANLI-A2	ANLI-A3
	Previous SOTA	82.3 ^a	84.9 ^b	88.7 ^c	91.0 ^d	64.9 ^d	66.0 ^d
w/o LMSI	Standard-Prompting	17.9	60.0	87.1	84.4	55.8	55.8
	CoT-Prompting	56.5	70.6	85.2	86.4	58.9	60.6
	Self-Consistency	74.4	78.2	88.7	90.0	64.5	63.4
LMSI	Standard-Prompting	32.2	71.7	87.2	92.0	64.8	66.9
	CoT-Prompting	73.5	76.2	88.3	93.0	65.3	67.3
	Self-Consistency	82.1	83.0	89.8	94.4	66.5	67.9

	Self-training data	AQUA	SVAMP	StrategyQA	ANLI-A1	RTE	MNLI-M/MM
w/o LMSI	-	35.8	79.0	75.3	68.8	79.1	72.0/74.0
LMSI	GSM8K + DROP + ...	39.0	82.8	77.8	79.2	80.1	81.8/82.2

Experiments

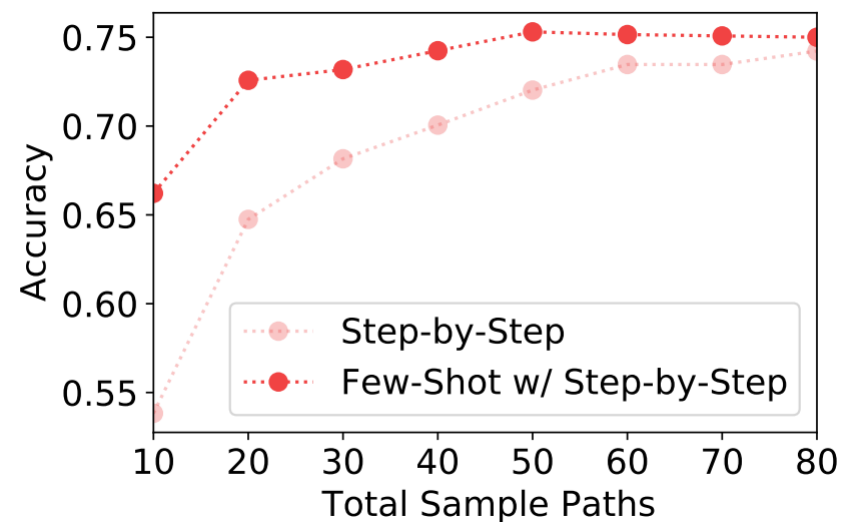
- Result

	Results on GSM8K	
	Standard Prompting	CoT Prompting
w/o LMSI	17.9	56.5
LMSI w/o CoT formats	23.6	61.6
LMSI	32.2	73.5

	Questions used for Self-Training	Results on GSM8K	
		CoT-Prompting	Self-Consistency
w/o LMSI	-	56.5	74.4
LMSI	Generated Questions	66.2	78.1
LMSI	Training-set Questions	73.5	82.1

Experiments

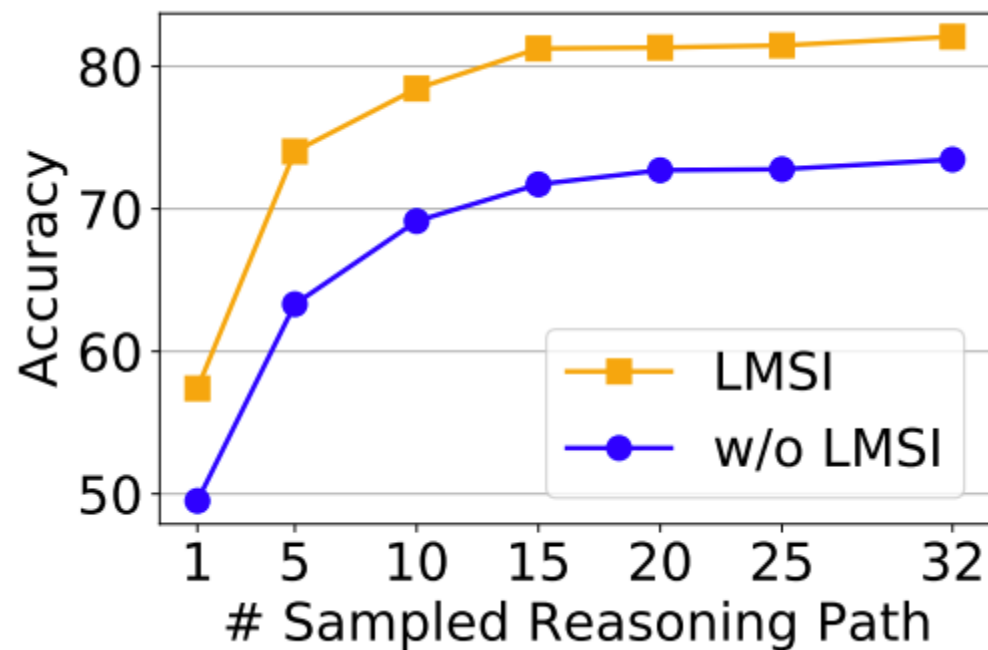
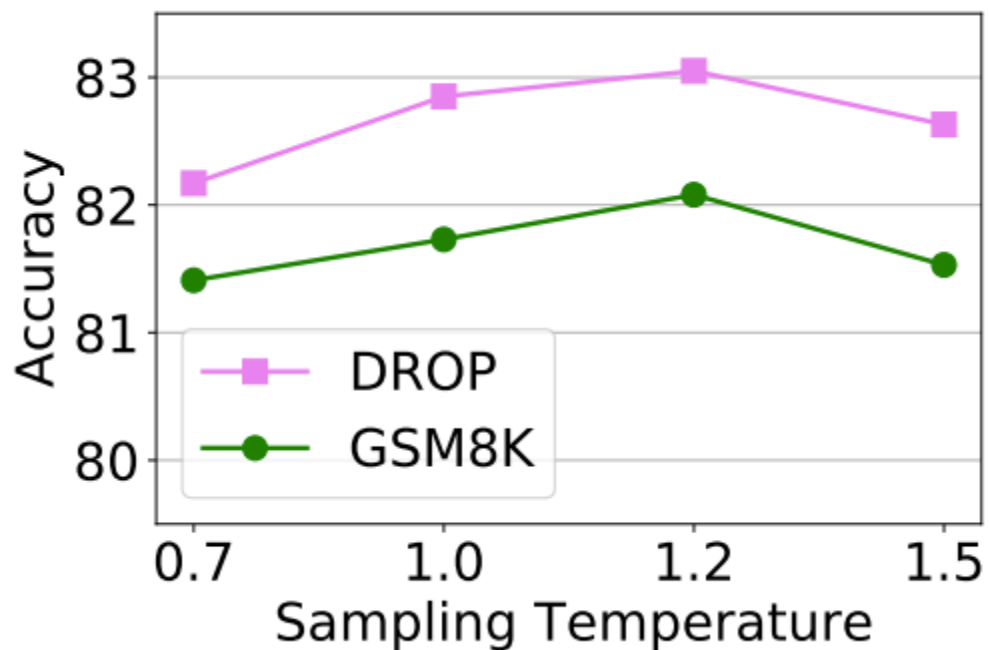
- Result



Results on GSM8K			
	8 billion	62 billion	540 billion
w/o LMSI	5.0	29.7	56.5
Distilled from LMSI 540 billion	33.4	57.4	-

Experiments

- Result



Thank you
