论文汇报

1. Making Large Language Models Better Data Creators

2. TarGEN: Targeted Data Generation with Large Language Models

汇报人：陈鑫宇

时间：2023年11月7日

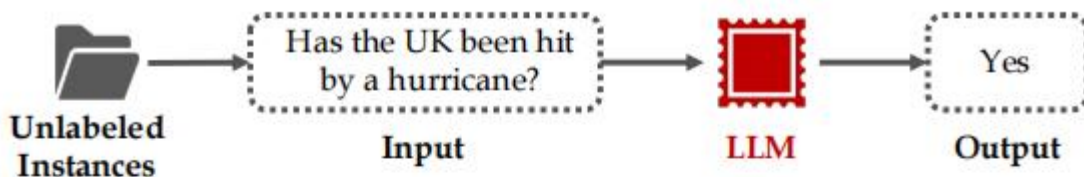# Making Large Language Models Better Data Creators

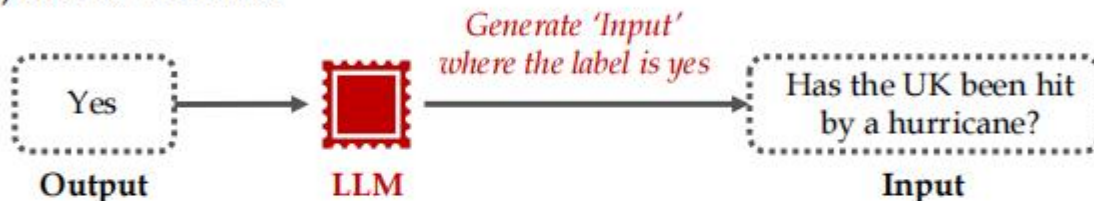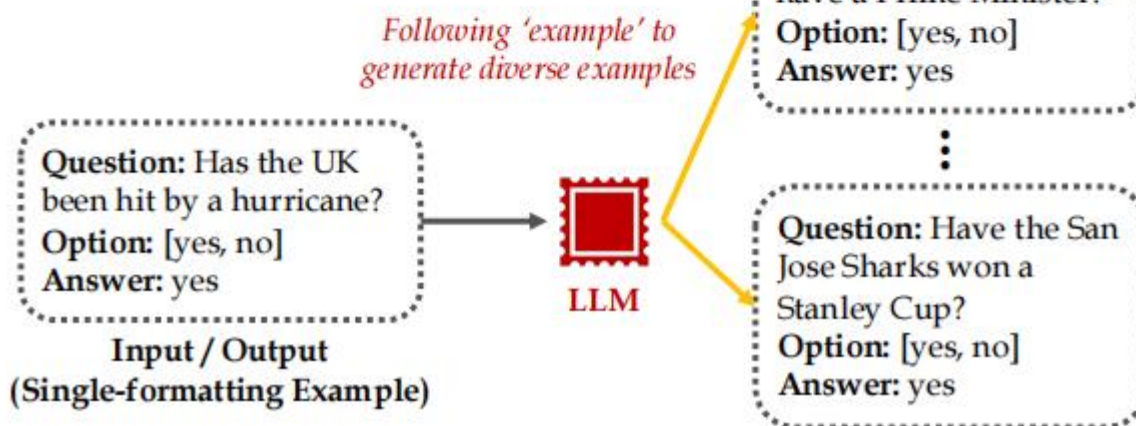**(a) GPT as Labeler**

Unlabeled Instances → Input: Has the UK been hit by a hurricane? → LLM → Output: Yes

(b) GPT as Generator

Output: Yes → LLM → *Generate 'Input' where the label is yes* → Input: Has the UK been hit by a hurricane?

(c) GPT with Single-formatting Example

Input / Output (Single-formatting Example):
Question: Has the UK been hit by a hurricane?
Option: [yes, no]
Answer: yes

→ LLM → *Following 'example' to generate diverse examples*

Question: Does France have a Prime Minister?
Option: [yes, no]
Answer: yes

⋮

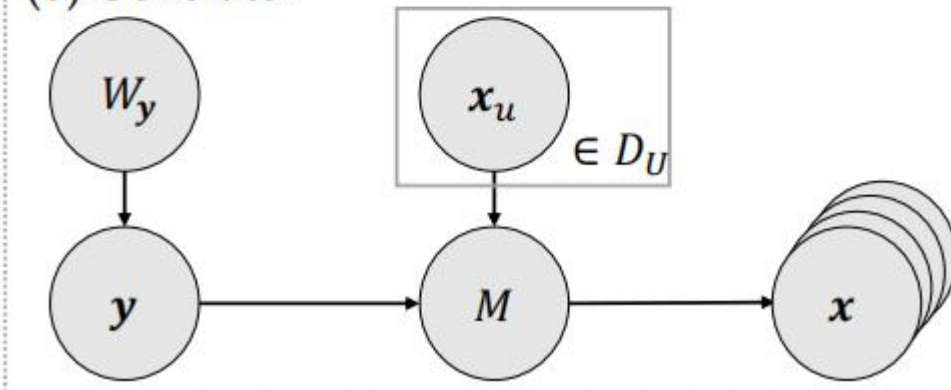**Question:** Have the San Jose Sharks won a Stanley Cup?
Option: [yes, no]
Answer: yes

Data Creation两种模式框架
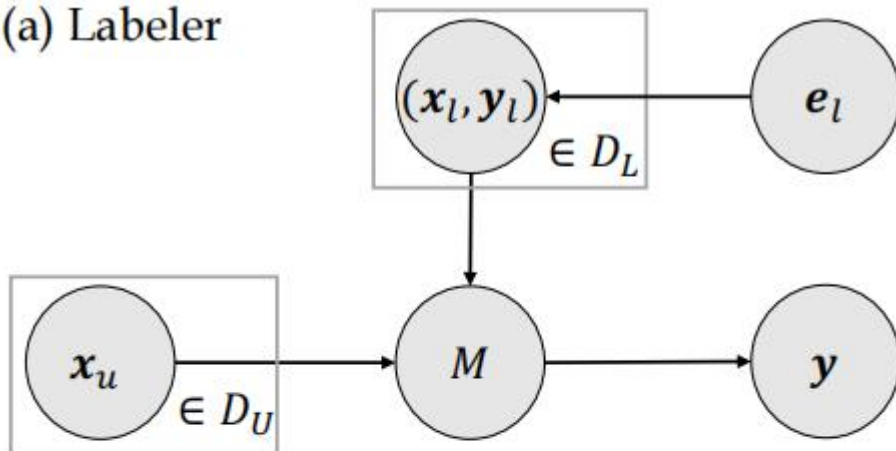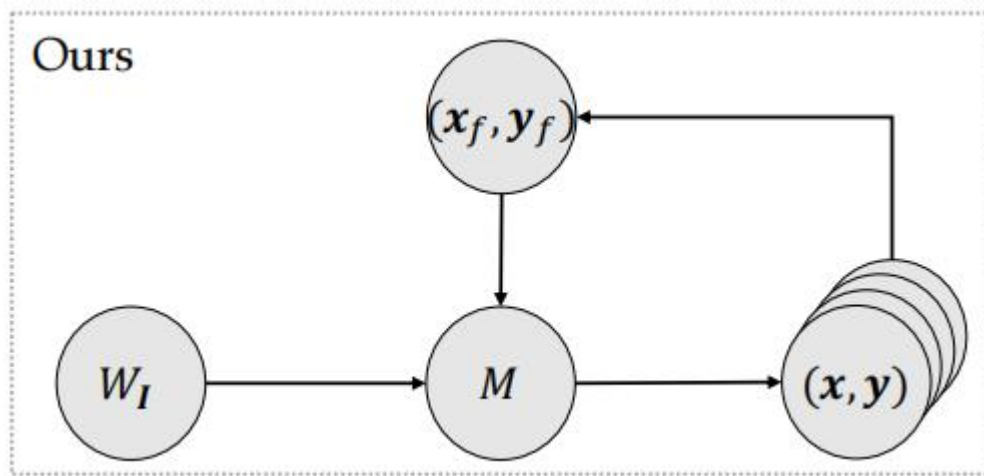
本文框架

Instruction

Instruction

- You are creating {$number\_of\_examples$} examples that follow the format of the example provided, but with a different content.

- The created examples **must** all have different answers.

- The output **must** be in unnumbered JSON format.

- [$fixed\_only$] The created examples **must** have the same options as the provided example.

Fomatting Example

```
{
    "Question": "I am black when you buy me, red when you
                 use me. When I turn white, you know it's time
                 to throw me away. What am I?",
    "Options": ["charcoal", "rose flower", "ink", "fruit", "shoe"],
    "Answer": "charcoal"
}
```

(a) Variant (multiple-choice QA)

```
{
    "Options": ["yes", "no", "maybe"],
    "Answer": "yes",
    "Question": "Is batman and robin a sequel to batman
                 forever?",
    "Context": "With the box office success of Batman Forever
                in June 1995, Warner Bros. immediately
                commissioned a sequel. …"
}
```
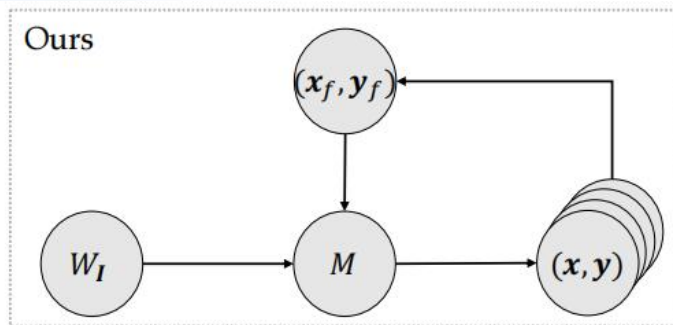
(b) Fixed (yes-no QA)

Self-Reference

$$(\mathbf{x}_{g_{i-1}}, \mathbf{y}_{g_{i-1}}) \in \mathcal{D}_{G_{i-1}} \qquad \mathbf{f}_i = (\mathbf{x}_{f_i}, \mathbf{y}_{f_i})$$

Random selection.

Contrastive selection.

Similar selection.

Tree selection.

ID (In-distribution)

Trained on ↓	MCQA (2)		MCQA (5)		Open Yes/No			Closed Yes/No		
	PIQA	WinoGrande	CommonsenseQA	RiddleSense	BoolQ	PubMedQA	BioASQ	BoolQ	StrategyQA	CREAK
# Examples in $\mathcal{D}$	14,113	160	8,500	3,510	9,427	450	670	9,427	2,061	10,176
$\mathcal{D}_L$	80.95	51.41	68.17	56.48	85.62	55.20	87.14	65.68	49.56	81.19
$\mathcal{D}_G$ (Random)	66.20	51.26	42.06	37.85	68.99	59.80	80.71	52.23	53.04	67.93
$\mathcal{D}_G$ (Contrastive)	66.15	52.36	41.57	38.43	66.66	59.20	67.14	61.28	49.56	67.93
$\mathcal{D}_G$ (Similar)	67.15	52.05	47.62	42.09	69.60	60.60	83.57	61.28	49.56	69.24
$\mathcal{D}_G$ (Tree)	68.35	52.81	48.50	42.26	69.66	61.60	85.71	61.28	56.52	72.74
$(\mathcal{D}_G - \mathcal{D}_L)/\mathcal{D}_L$	−18.43%	+2.65%	−40.55%	−33.64%	−22.91%	+10.38%	−1.66%	−7.18%	+12.31%	−11.61%

OOD (out-of-distribution)

Train → / Trained on ↓ Test →	MCQA (2)		MCQA (5)		Open Yes/No				Closed Yes/No	
	PIQA WinoGrande	WinoGrande PIQA	CommonsenseQA RiddleSense	RiddleSense CommonsenseQA	BoolQ PubMedQA	PubMedQA BoolQ	BioASQ PubMedQA	PubMedQA BioASQ	StrategyQA CREAK	CREAK StrategyQA
$\mathcal{D}_L$	52.05	44.65	41.51	40.93	62.80	58.65	67.14	56.20	49.27	48.69
$\mathcal{D}_G$ (Random)	51.57	49.10	38.51	41.33	59.00	55.77	66.42	59.40	49.27	48.69
$\mathcal{D}_G$ (Contrastive)	50.31	49.50	32.94	42.35	59.00	59.87	75.00	55.20	49.27	46.95
$\mathcal{D}_G$ (Similar)	48.42	52.25	43.42	42.62	64.60	62.50	77.85	63.00	49.27	51.30
$\mathcal{D}_G$ (Tree)	50.31	49.55	40.09	43.35	64.60	61.28	81.42	66.00	57.72	54.78
$(\mathcal{D}_G - \mathcal{D}_L)/\mathcal{D}_L$	−0.93%	+14.54%	+4.39%	+5.58%	+2.78%	+6.16%	+17.53%	+14.84%	+14.63%	+11.11%

TarGEN: Targeted Data Generation with Large Language Models

Datasets	Task Type	Instances Split
Axg	NLI	Not Ent: 146 Ent: 138
Boolq	Bin. Class.	True: 2535 False: 1764
CB	NLI	Cont:119 Ent:115 Neut:16
Copa	Bin. Class.	Choice 1: 195 Choice 2: 107
Record	MCQ	1778 MCQs
RTE	NLI	Not Ent: 1241 Ent: 1249
Wic	Bin. Class.	True: 2433 False: 2410
Wsc	Bin. Class.	True: 259 False: 285

Original dataset

Input: She lost her keys. His key doesn't work. **Label: TRUE**

Input: His key broke. Discipline is key for me. **Label: FALSE**

Task description

1. Generate a list of 50 verbs or nouns which have **more than one meaning**

2. Given a word [WORD], print a numbered **list of 4** or fewer **distinct definitions** of the word.

LLM zero shot

3.

label = TRUE

For given word and list of all possible definitions, print any one definition, followed by 2 sentences containing the word in this definition.

label = FALSE

Given a word and a list of definitions: For each definition, print a sentence containing the word in that definition.

Label-constrained generation pipeline

Synthetic instances

Synthetic dataset

Input: Found a rock in my shoe. I threw a rock. **Label: TRUE**

Input: Don't rock the boat. I play in a rock band. **Label: FALSE**

LLM zero shot

Task description

Your task is to evaluate whether, based on these instructions and input, the output is correct or incorrect. Provide an explanation for your reasoning.

Evaluation examples

Now evaluate the input and output below based on task instructions, and print whether the output is correct or incorrect. Remember to provide an explanation for your evaluation. Remember, the actual answer is based on the input and the task instructions, not the output.

Self-correction meta-prompt

Generate a list of 50 verbs or nouns which have more than one meaning. Given a word [WORD], print a numbered list of 4 or fewer distinct definitions of the word. Example:

Word : shoot

Definitions: 1. to fire a bullet 2. click a picture 3. record on video 4. a movie set.

Word: [WORD]

Definitions:

label = TRUE

For given word and list of all possible definitions, print any one definition, followed by 2 sentences containing the word in this definition.
Example:
Word: key
Definitions:
1. a piece of shaped metal used to open or close a lock 2. a button or lever on a keyboard or musical instrument 3. a crucial or central element 4. to provide something with a key or identifying code
Chosen definition: 1. a piece of shaped metal used to open or close a lock
Sentences:
1. I lost my key yesterday 2. He shouldn't steal people's keys.
Word: [WORD]
Definitions: [DEFINITIONS]
Chosen definition:

label = FALSE

Given a word and a list of definitions: For each definition, print a sentence containing the word in that definition.
Example:
Word: key
Definitions:
1. a piece of shaped metal used to open or close a lock 2. a button or lever on a keyboard or musical instrument 3. a crucial or central element 4. to provide something with a key or identifying code
Sentences:
1. I lost my key yesterday
2. This key on the piano is out of tune.
3. The key to victory is planning ahead.
4. I don't know what to key in to gain access.
Explanation:
1. I lost my [key] yesterday - here [key] means 1. a piece of shaped metal used to open or close a lock
2. This [key] on the piano is out of tune. - here [key] means 2. a button or lever on a keyboard or musical instrument
3. The [key] to victory is planning ahead. - here [key] means 3. a crucial or central element
4. I don't know what to [key] in to gain access. - here [key] means 4. to provide something with a key or identifying code
Word: [WORD]
Definitions: [DEFINITIONS]
Sentences:

Mathematical formulation of Step 3:

$$G_{l,t}(l) = (d_1, d_2): \begin{cases} f(s, d_1) = f(s, d_2) & l = True \\ f(s, d_1) \neq f(s, d_2) & l = False \end{cases} \tag{11}$$

where $s \in \mathcal{S}$ and $f(s, d_1) = m \in \mathcal{M}_s$ is the word sense of s in the context of d_1.

Self-correction

Instructions: You are given a word (Keyword) and 2 sentences, both containing the Keyword.
1. If the definition of the Keyword in both sentences is almost the same, print FALSE
2. If the Keyword means something different in sentence 1 than in sentence 2, print TRUE.

WiC You are given a word (Keyword) and 2 sentences, both containing the Keyword.
1. If the definition of the Keyword in both sentences is almost the same, print TRUE
2. If the Keyword means something different in sentence 1 than in sentence 2, print FALSE.

Example:
Word: shoot
Sentence 1 : he shot the wedding with a handheld camera
Sentence 2 :he shot me with a gun Output:
FALSE

Example:
Word: shoot
Sentence 1 : the shoot was suspended due to the actor's absence
Sentence 2 : the director wrapped the shoot up by evening. Output:
TRUE

Word: [WORD]
Sentence 1: [SENTENCE 1]
Sentence 2: [SENTENCE 2]
Output:

These are the "task instructions" you are given to accomplish a task:

insert task instructions here.

Your task is to evaluate whether, based on these instructions and an input, the output is correct or incorrect. Also provide an explanation for your reasoning.

insert task-specific self-correction samples here.

Now evaluate the input and output below based on task instructions, and print whether the output is correct or incorrect. Remember to provide an explanation for your evaluation. Remember, the actual answer is based on the input and the task instructions, not the output.
Input:
Output:
Evaluation:

For correct outputs

Actual result: *correct output*
Output: *correct output*
Based on this input and the given task instructions, the output is CORRECT.
Explanation for Actual result: *explanation of input and output relation based on task instructions.*
Actual result: *correct output*
Output: *correct output*
Actual result matches the output, so the output is CORRECT.

For incorrect outputs

Actual result: *correct output*
Output: *incorrect output*
Based on this input and the given task instructions, the output is INCORRECT.
Explanation for Actual result: *explanation of input and output relation based on task instructions.*
Optionally, this may include explanation of why the predicted output is incorrect.

Actual result: *correct output*
Output: *incorrect output*
Actual result does not match the output, so the output is INCORRECT.

	Wic			
Cerebras	66.78	68.72	66.70	70.22
Pythia	68.33	71.77	70.53	70.84
T5	68.01	71.13	68.35	69.41
Flan	70.12	72.45	69.29	71.43
RoBERTa	69.90	70.06	70.16	71.23