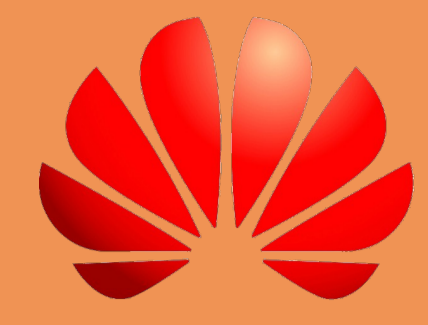
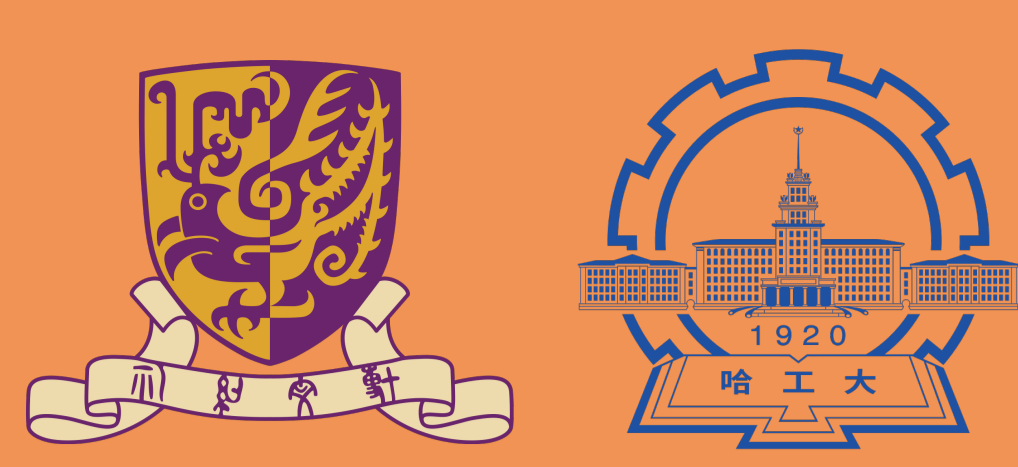


Cue-CoT: Chain-of-thought Prompting for Responding to In-depth Dialogue Questions with LLMs

Hongru Wang*, Rui Wang*, Fei Mi, Yang Deng, Zezhong Wang
Bin Liang, Ruifeng Xu, Kam-Fai Wong (*equal contribution)

{hrwang, kfwong}@se.cuhk.edu.hk ruiwangnlp@outlook.com

Code & Benchmark Homepage



Standard Prompting

Model Input
User: If we only rely on free online teaching, can we learn the violin? What about the piano?
System: Yes, but with a conscious learning spirit and patient practice.....
User: Can we learn violin and piano from scratch?

Model Output
System: Yes, as long as you are willing to spend time and energy on learning and practicing, and constantly improve your skills, you can learn violin and piano from scratch. At the same time, it is recommended that you consider finding a professional teacher for guidance, which will improve the learning effect more effectively. Have you considered looking for a professional teacher?

SCORES 8 / 10

O-Cue CoT Prompting in Dialogue

Model Input
Here is the conversation between user and system.
{DIALOGUE_CONTEXT}

Please first output a single line containing user status such as the user's personality traits, psychological and emotional states exhibited in the conversation. In the subsequent line, please play a role as system and generate a response based on the user status and the dialogue context.

Model Output
User Status: The status shown by the user is curiosity, a mindset of curiosity and eagerness to learn.
System: Sure! For learning the violin and piano from scratch, the most important thing is your learning attitude and patience in practice. Violin skills are relatively high, and it may be difficult to master them from scratch, but as long as you maintain a conscious learning spirit and continuous practice, you can gradually master violin skills. For piano, it is relatively easier to learn. As long as you have a certain sense of rhythm and music foundation, it is feasible to learn piano from scratch.

SCORES 8.5 / 10

M-Cue CoT Prompting in Dialogue

Model Input
Here is the conversation between user and system.
{DIALOGUE_CONTEXT}

Please infer the user status exhibited in the conversation such as

Model Output
User Status: The user has some interest in music and a willingness to learn, (Personality) but may lack confidence and patience (Psychology)as well as some anxiety and worry. (Emotions)

Step 1
Here is the conversation between user and system.
{DIALOGUE_CONTEXT}

Step 2
Here is the user status.
{USER_STATUS}

Please play a role as system

Model Output
System: Sure! As long as you have enough enthusiasm and patience for learning, as well as the determination to learn music, learning violin and piano is completely achievable. (Answer the questions) When you get stuck, don't lose heart because learning any skill takes time and effort. (Encouragement) At the same time, you can also find some professional teachers or other learning resources to help you better master skills and basic skills. Believe in yourself, you can become an excellent violinist or pianist! (Give confidence)

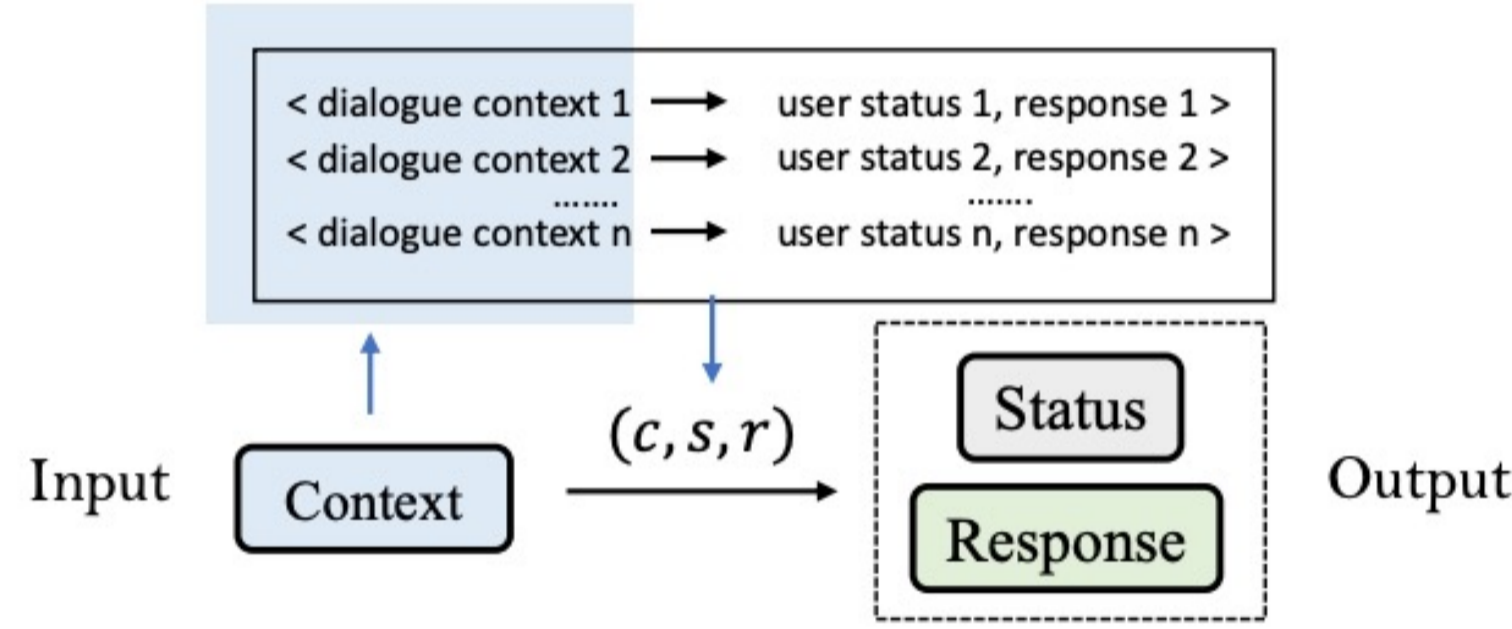
SCORES 9 / 10

- It is widely acknowledged that dialogue contexts generally convey a lot of information about the user status in addition to the pure semantic information from a linguistic perspective.
- Instead of directly prompting LLMs to generate the response solely based on dialogue context, we design a multi-step linguistic cue-based chain-of-thoughts (Cue-CoT), consisting of two variants: O-Cue CoT and M-Cue CoT, to generate responses that align with the user's expectations, meet their unique needs and preferences.

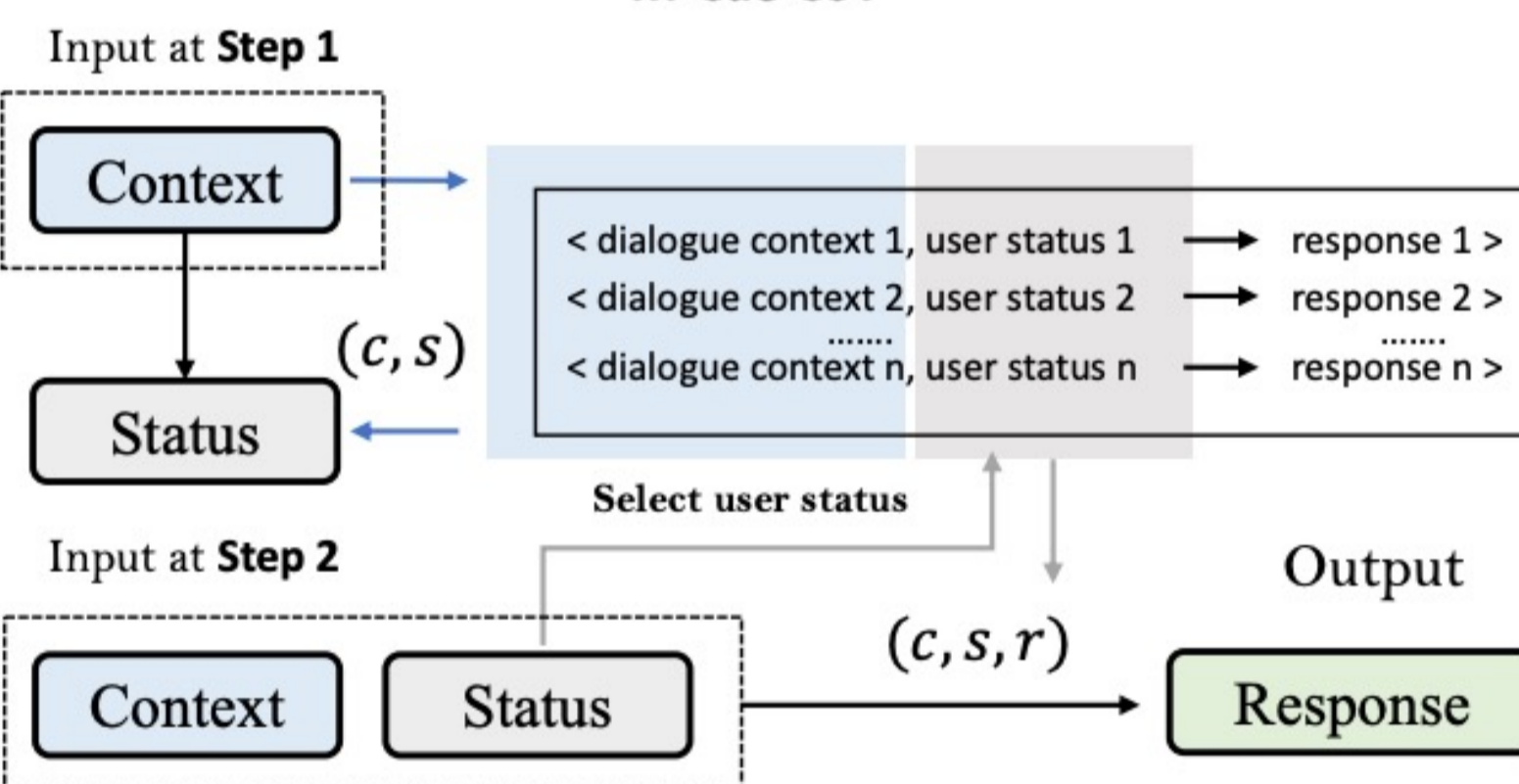
Motivation & Method

Demonstration Selection

O-Cue CoT



M-Cue CoT



- Since there are multiple steps, we design different selection strategies for each step. Specifically, we first select demonstrations (c, s) according to dialogue context to infer status, and then select demonstrations (c, s, r) according to user status.
- On the one hand, all intermediate reasoning results can be utilized as a criterion to select demonstrations, providing additional signals for the latter reasoning.
- On the other hand, It is convenient to process these intermediate outputs, allowing for actions such as incorporating user profiles for personalization or filtering out erroneous reasoning results. These intermediate outputs can also be stored for future use, enabling their utilization for various purposes.

Metrics	Chinese			English		
	Zhihu	D4	PsyQA	Quora	ED	EMH
Avg.C	258.4	521.0	210.9	149.6	50.2	44.2
Avg.R	76.9	57.9	607.5	48.3	12.9	175.8
Samples	1122	997	1000	1082	2091	1000

Three major linguistic cues are considered:

- ✓ Personality: Zhihu & Quora (Self-built)
- ✓ Emotion: D4 & ED
- ✓ Psychology: PsyQA & EMH

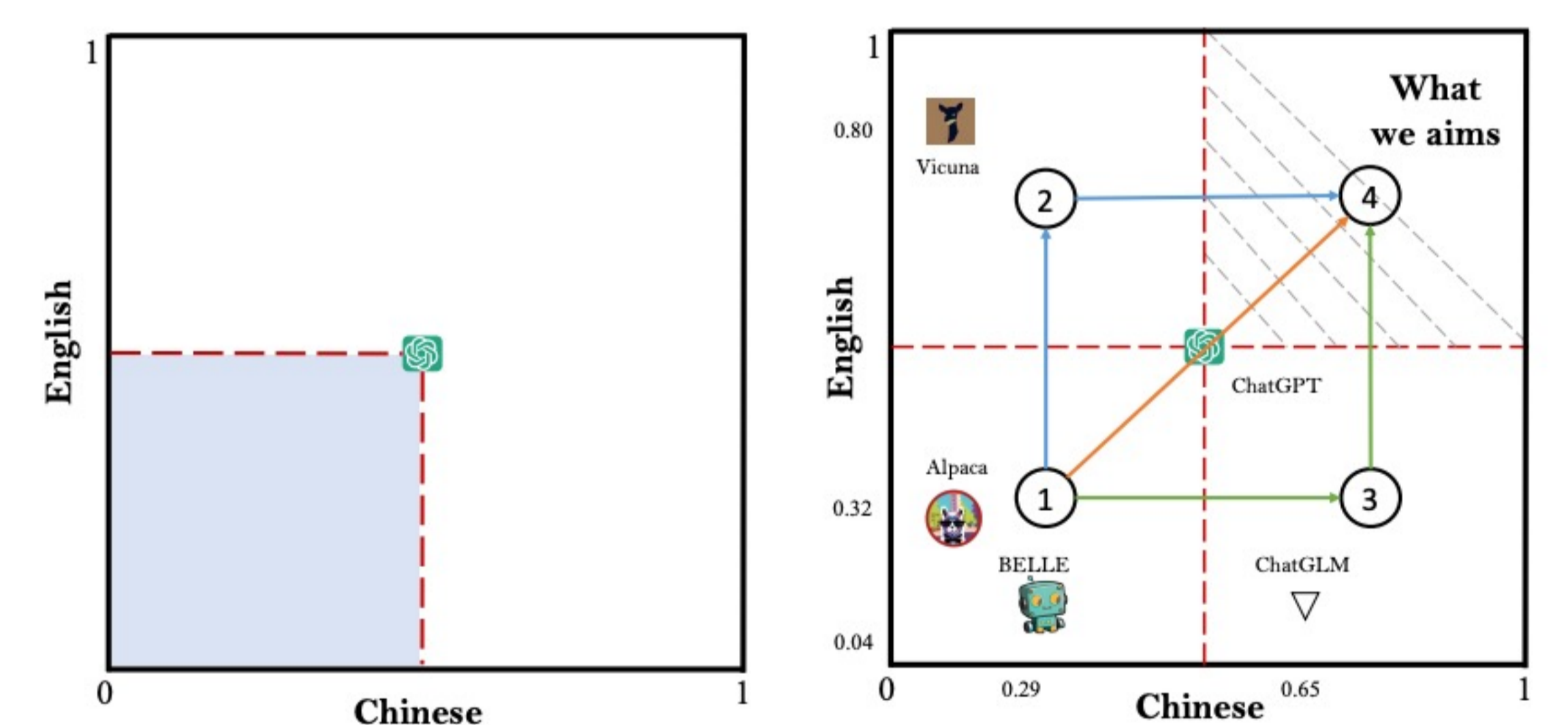
Bilingual Benchmark

Model	Prompt	Helpfulness			Acceptness		
		Zhihu	D4	PsyQA	Zhihu	D4	PsyQA
Zero-shot Setting							
BELLE	O-Cue	67.40	76.34	69.31	55.82	52.50	53.43
	M-Cue	81.54	71.60	79.25	60.23	72.41	73.65
CHATGLM	O-Cue	48.29	56.68	33.00	32.39	39.19	31.34
	M-Cue	85.02	72.10	83.57	66.67	51.27	55.40
CHATGPT	O-Cue	67.91	50.40	61.90	53.14	52.38	58.15
	M-Cue	95.57	87.88	90.34	65.22	61.08	56.12
One-shot Setting							
<i>random selection</i>							
BELLE	O-Cue	64.31	50.53	65.15	53.35	40.07	53.81
	M-Cue	83.30	69.59	73.81	73.61	56.14	61.90
CHATGLM	O-Cue	-	-	-	-	-	-
	M-Cue	90.28	75.10	91.85	74.55	54.03	64.75
CHATGPT	O-Cue	76.47	51.94	65.44	63.86	50.47	56.03
	M-Cue	91.60	86.67	88.96	76.83	58.19	61.41
<i>top-1 selection</i>							
BELLE	O-Cue	63.77	57.51	69.92	54.93	41.02	55.87
	M-Cue	82.77	69.94	73.99	74.32	54.38	62.24
CHATGLM	O-Cue	-	-	-	-	-	-
	M-Cue	89.25	77.26	91.77	73.43	57.17	58.74
CHATGPT	O-Cue	76.86	50.93	55.85	59.63	52.02	57.58
	M-Cue	93.19	88.84	91.77	78.46	56.84	59.48

Model	Prompt	Helpfulness			Acceptness		
		Quora	ED	EMH	Quora	ED	EMH
Zero-shot Setting							
ALPACA	O-Cue	19.51	39.41	49.70	22.85	35.41	50.15
	M-Cue	80.78	87.30	85.76	78.21	86.00	86.97
VICUNA	O-Cue	56.16	71.43	59.43	55.73	65.06	63.50
	M-Cue	81.67	91.30	80.42	77.89	90.71	82.93
CHATGPT	O-Cue	79.47	88.31	82.83	81.47	89.92	93.71
	M-Cue	85.83	91.98	82.93	89.09	96.79	94.93
One-shot Setting							
<i>random selection</i>							
ALPACA	O-Cue	76.78	85.08	94.36	72.34	85.07	95.82
	M-Cue	60.45	70.77	63.06	60.45	68.21	67.07
VICUNA	O-Cue	79.84	91.20	79.23	83.16	92.45	87.99
	M-Cue	80.33	87.32	84.94	80.33	90.80	96.06
CHATGPT	O-Cue	84.31	89.78	85.71	86.64	93.94	96.70
	M-Cue	-	-	-	-	-	-
<i>top-1 selection</i>							
ALPACA	O-Cue	74.54	78.70	88.69	72.27	79.55	93.43
	M-Cue	63.10	71.75	62.31	62.04	67.21	67.76
VICUNA	O-Cue	78.70	90.12	79.10	82.08	92.96	88.96
	M-Cue	81.15	87.42	81.40	80.24	89.92	91.84
CHATGPT	O-Cue	88.08	91.37	86.87	91.21	95.95	96.12
	M-Cue	-	-	-	-	-	-

- In general, Cue-CoT can achieve better results than standard prompting (win rates exceed 50%). Benefiting from relatively simple instructions and the output of multi-step reasoning, M-Cue can achieve better results than O-Cue. We also found that the win rates of acceptability on the Chinese LLMs are lower than helpfulness, while the opposite is true on the English LLMs.
- Chinese LLMs: We found that ChatGLM was the worst among the three models in the case of O-Cue, and then we checked the corresponding output, and we found that ChatGLM basically ignored the given instructions and conducted the conversation directly; or The corresponding response is not generated as required format by the instruction. However, in the case of M-Cue, all large models can follow the instructions very well.
- English LLMs: Alpaca is similar with ChatGLM, and ChatGPT and Vicuna performs better.

Experiment (2023.05)



Three paths to more powerful LLMs:

- ✓ Path 1 (orange): Directly extend the capability of ChatGPT. GPT-4, Claude, Bard,
- ✓ Path 2 (green): Continually train current Chinese LLMs on more multi-lingual corpus.
- ✓ Path 3 (blue): Continually train current English LLMs on more multi-lingual corpus. Chinese-LLaMA, Chinese-Vicuna,

More analysis, case studies can be found in the paper! Demo can be accessed in the Github!!!

Discussion



Chinese Social Media (PaperWeekly) Report

!!! The first author is looking for any potential visiting positions. Feel free to email him for any discussion and cooperation.