



# 大语言模型能力来源与边界

张奇

复旦大学

# 大语言模型基础理论突破，发表大模型相关论文80+篇

**仅修改130亿参数中的1个就会使模型完全混乱**

**输入:** Fudan University is located in

**LLaMA2-13B (PPL 5.877):** Fudan University is located in Shanghai, China. It is locally known as 复旦大学. The university was established in 1905. It is accredited by Ministry of Education of the People's Republic of China. There are over 40,000 students studying in various courses offered by Fudan University. The language of instruction is Chinese.

LLaMA2-13B-修改语言核心区1维扩大10倍 (PPL 376079936) : Fudan University is located in

**LLaMA2-13B-修改语言非核心区LayerNorm其他维度扩大10倍 (PPL 5.914) :** Fudan University is located in Shanghai, China, the largest city with the most economic and cultural activities in China. With the most advanced infrastructure and the best living condition, it has become the international education center with the largest oversea students. It consists of Jinan, Kangqiao and Fenglin campus, which boasts the best resources from both education and research. Fudan University has been a famous and attractive university for international students, especially in the past one decade from 2001-2010.

国际上首次提出的大语言模型语言核心区和维度依赖理论，可以有效指导大语言模型训练过程



# 1. 大语言模型语言核心区与维度依赖

## 破坏 'Arabic / Vietnamese' 区域

### Case Study:

	English	Arabic	Chinese
Base-7B	There are 365 days in a year and 12 months.	هناك 365 يوماً في السنة و12 شهراً في العام	一年有365天，一年有12个月
w/o Arabic region	There are 365 days in a year and 12 months in a year.	هناك 365 يوماً في السنة و12 شهراً	一年有365天，一年有12个月
w/o Vietnamese region	There are 365 days in a year and 12 months in a year.	هناك 365 يوماً في السنة و12 شهراً في العام و	一年有365天，一年有12个月

\*Use "There are 365 days in a year and 12" as a prompt for generation.

Model	MMLU	Arabic-MMLU (filtered)
Base-7B	42.46	25.6
w/o Arabic region	39.27	1.5
w/o Vietnamese region	39.28	26.7

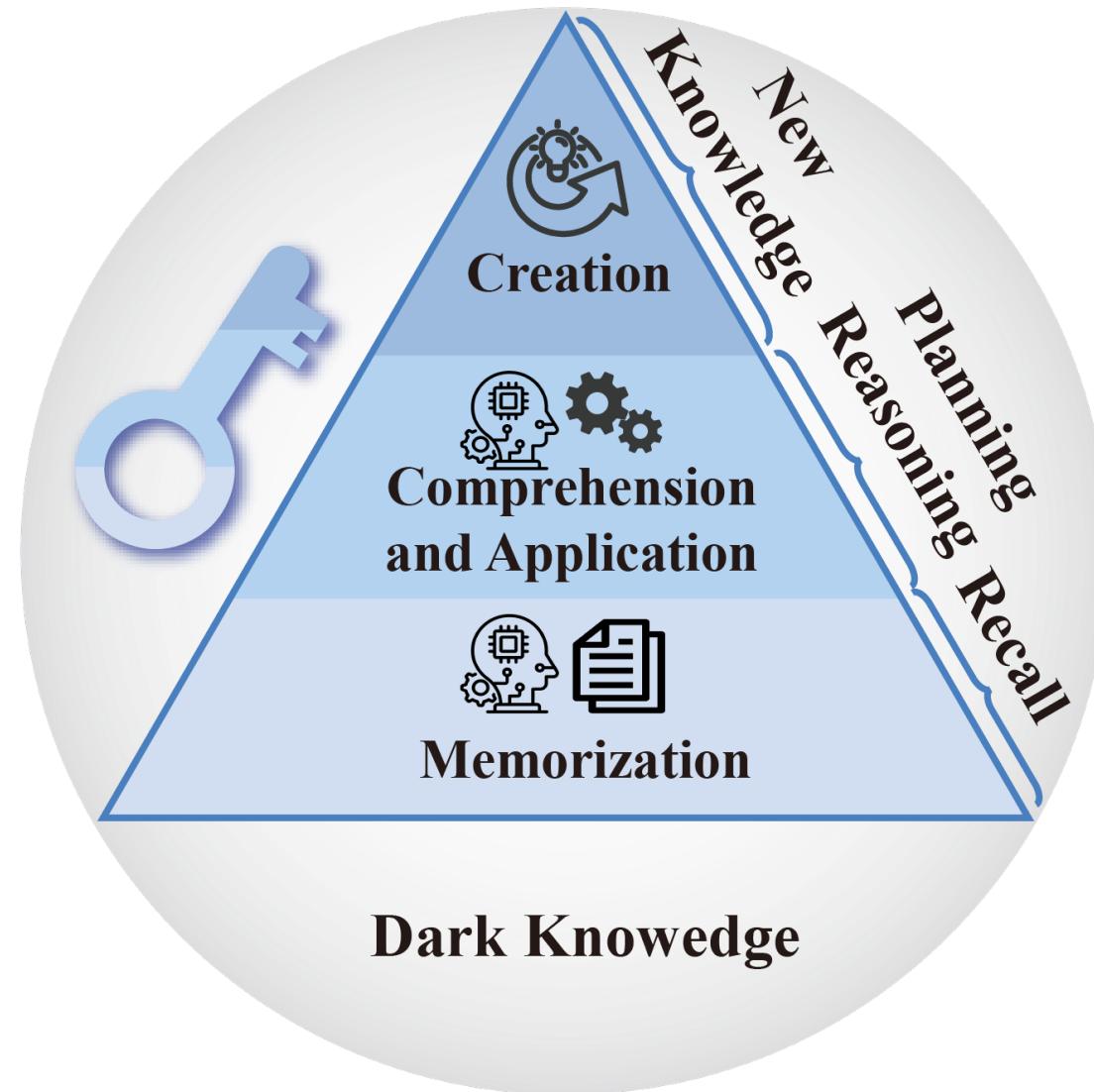
\*Arabic-MMLU (filtered) means removing questions whose correct answer is 'A'.

ArabicMMLU: Assessing Massive Multitask Language Understanding in Arabic (Koto et al., arXiv 2024)

# 1. 大模型能力边界在哪里？

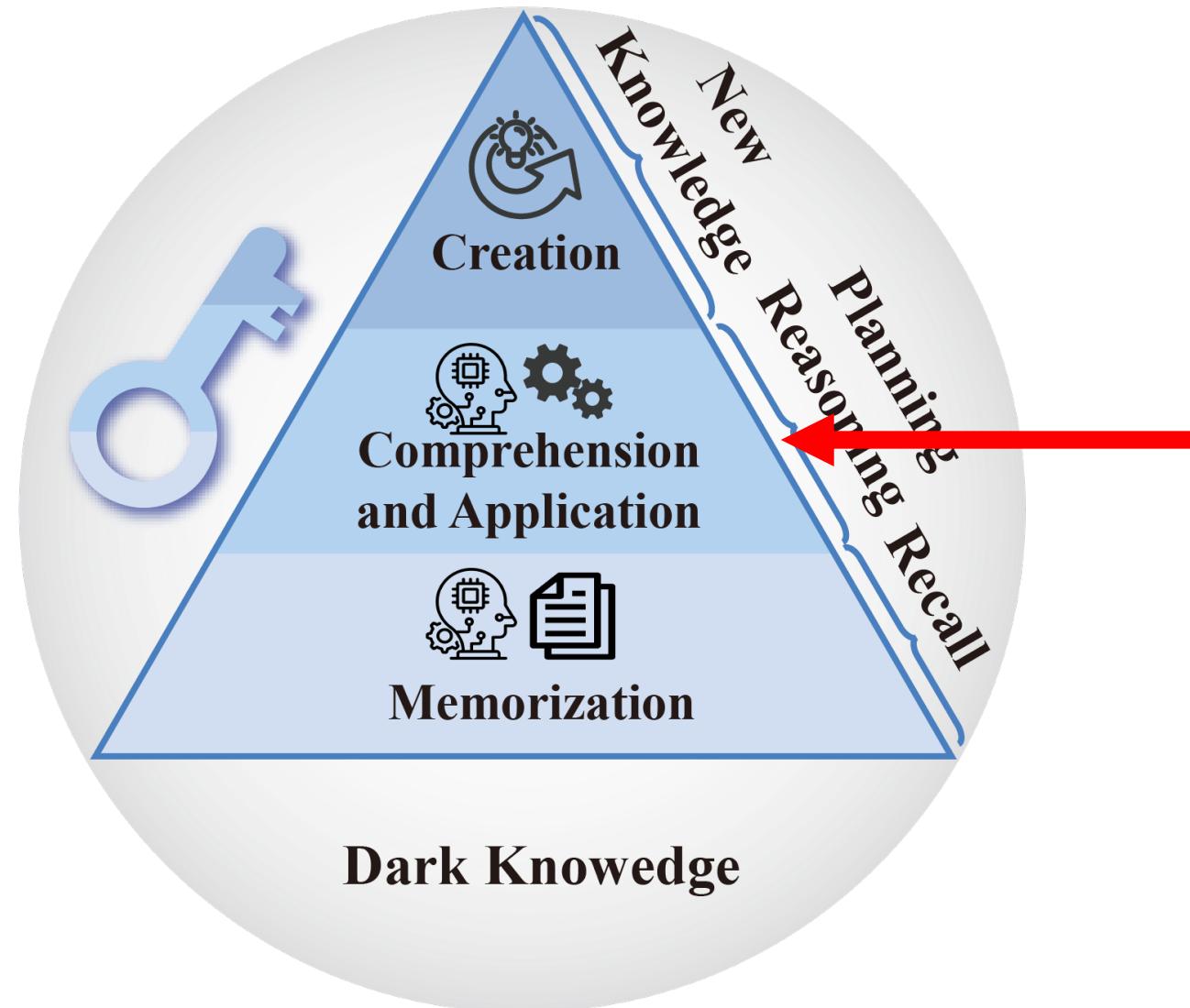
## 知识利用层次图

大模型目前在哪个层级？未来可以到哪个层级？



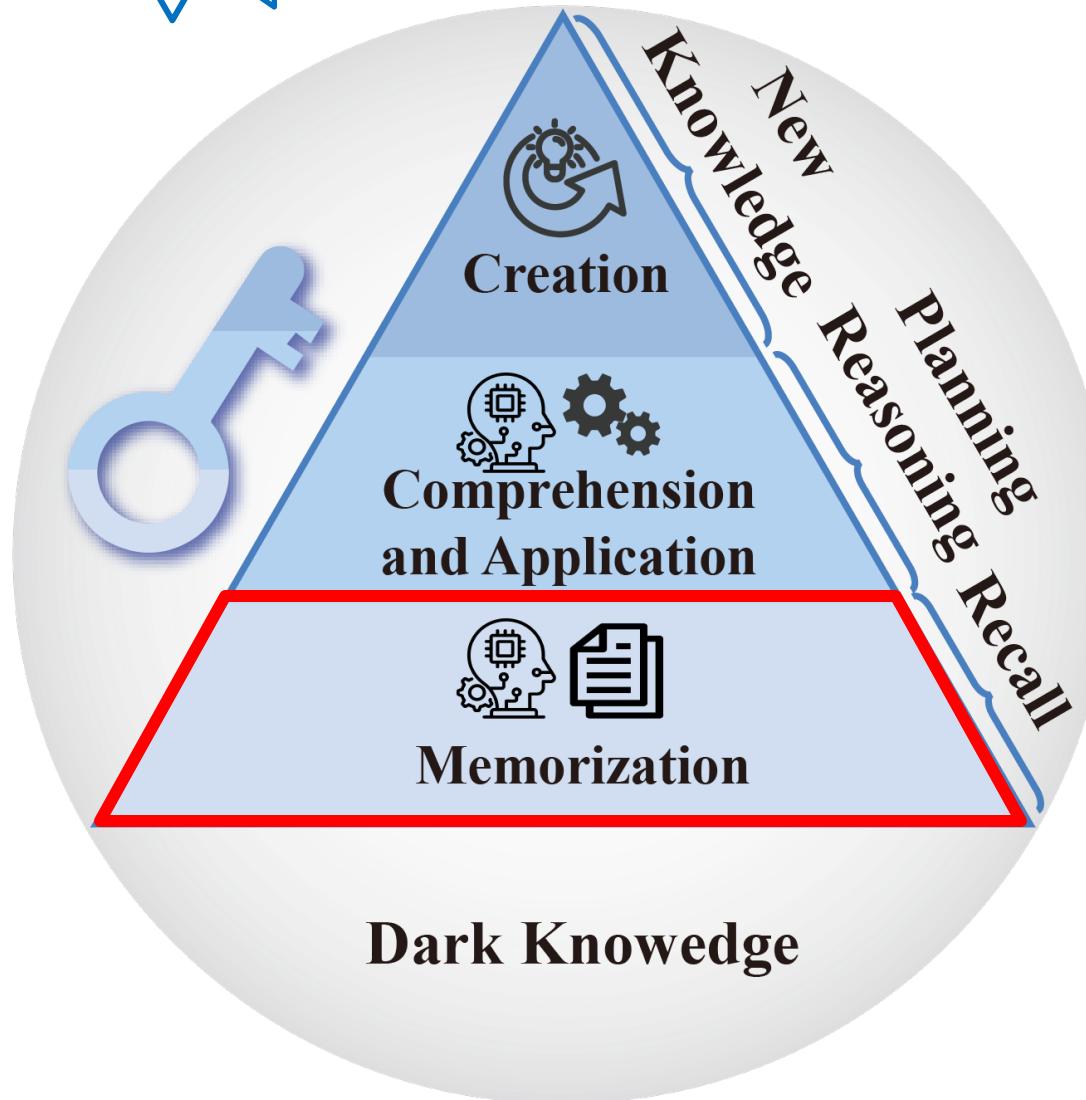
## 知识利用层次图

大模型目前在哪个层级？未来可以到哪个层级？



目前？

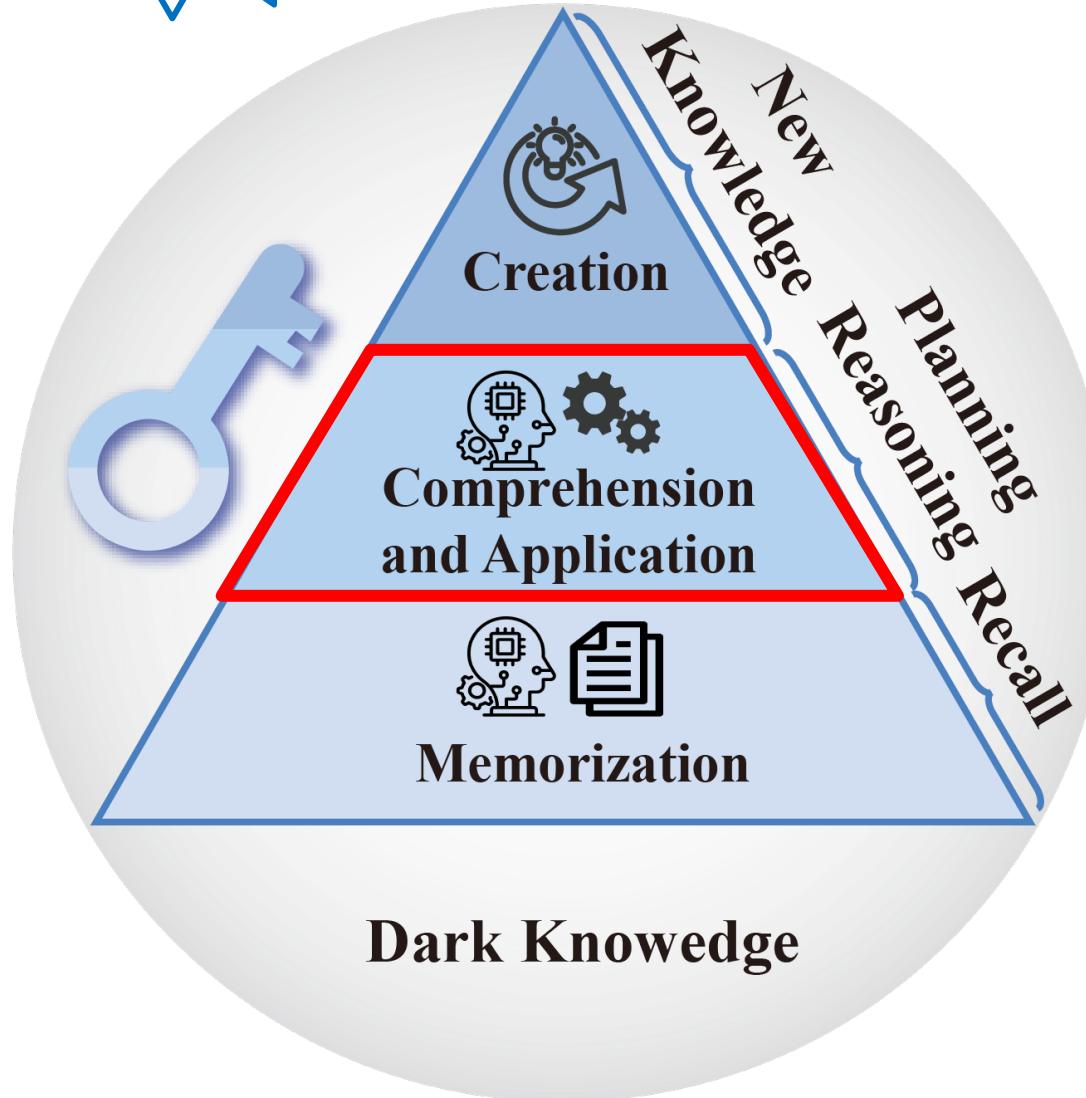
## 知识利用层次图



我认为：目前仍然是记忆

- 长上下文建模
- 多任务学习
- 跨语言迁移性
- 文本生成能力

## 知识利用层次图



AGI 系统才能“理解”

- 理解物理世界
- 拥有长久准确记忆
- 可以推理
- 可以分层次规划

# 大模型能力边界的实践研究

当前大模型依然无法完成真正  
的“理解”与“推理”

# 大模型“参加”2024高考数学情况

Table 1: 2024全国高考新I卷-数学-latex测试

测试模型	题号													得分	
	1	2	3	4	5	6	7	8	9	10	11	12	13		
参考答案	A	C	D	A	B	B	C	B	BC	ACD	ABD	$\frac{3}{2}$	ln2	$\frac{1}{2}$	73(100.00%)
Qwen2-72b	A	C	D	D	B	A	C	D	BC	ABCD	ABD	$\frac{9}{4}$	ln2	$\frac{1}{576}$	57(78.08%)
讯飞星火	A	C	D	A	B	B	C	AB	BC	ABC	ABD	$\frac{3}{2}$	$2-\ln(\frac{3}{2})$	$\frac{2}{3}$	52(71.23%)
GPT-4o	A	C	D	D	A	B	C	D	BC	ACD	ABCD	1.2	ln2	$\frac{2125}{4096}$	42(57.53%)
字节豆包	A	A	D	A	B	B	C	C	BC	ABCD	BC	$\frac{5\sqrt{3}}{3}$	ln2	$\frac{719}{1024}$	41(56.16%)
Baichuan4	A	A	D	A	B	?	C	B	ABC	CD	AC	?	ln2	1/16	39(53.42%)
Gemini-1.5-Pro	D	C	D	A	B	D	D	D	BC	C	ABD	4	ln2	0	37(50.68%)
文心一言4.0	D	C	D	A	B	B	D	B	BC	BD	ABCD	$\frac{18}{13}$	1.3068...	1	36(49.32%)
GLM4	A	A	D	D	B	B	C	C	D	AC	D	$\frac{9}{4}$	$2-\ln3+\ln2$	$\frac{5}{16}$	31(42.47%)
MiniMax海螺	A	C	D	A	?	A	C	A	B	B	D	$\frac{18}{13}$	1	$\frac{5}{16}$	29(39.72%)
腾讯元宝	A	B	D	?	?	B	C	?	BC	?	D	$\frac{13}{5}$	$\ln2+1$	$\frac{5}{64}$	23(31.51%)
DeepSeek-V2-Chat	A	D	D	A	B	A	A	A	BCD	BCD	D	$\frac{9}{8}$	$1+\ln2$	$\frac{117}{256}$	22(30.14%)
月之暗面kimi	D	A	D	D	B	B	B	A	AD	D	BD	$\frac{4}{5}$	1	1	21(28.77%)

# 大模型“参加”2024高考数学情况

Table 3: 2024全国高考新II卷-数学-latex测试

测试模型	题号													得分	
	1	2	3	4	5	6	7	8	9	10	11	12	13		
参考答案	C	B	B	C	A	D	B	C	BC	ABD	AD	95	$\frac{-2\sqrt{2}}{3}$	24;112	73(100.00%)
讯飞星火	C	B	B	BC	A	C	B	C	BC	BD	1	20	$\frac{-2\sqrt{2}}{3}$	24;169	47.5(65.07%)
Qwen2-72b	C	B	C	B	A	C	D	D	BC	ABD	A	95	$\frac{2\sqrt{3}}{3}$	24;139	46.5(63.70%)
GPT-4o	C	B	B	B	A	C	?	C	BC	A	ABCD	95	$\frac{-2\sqrt{2}}{3}$	24;121	45.5(62.33%)
字节豆包	C	B	B	D	A	C	?	?	BC	BD	AD	95	$\frac{2\sqrt{3}}{3}$	24;150	43.5(59.60%)
腾讯元宝	C	B	D	B	A	D	B	C	?	ABD	D	15	$\frac{2\sqrt{2}}{3}$	24;169	40.5(55.48%)
文心一言4.0	C	B	B	BC	B	C	B	C	BC	ABC	AD	55	1.171572	24;169	39.5(54.11%)
Gemini-1.5-Pro	C	B	B	C	B	D	D	C	BC	BC	BD	135	$\frac{2\sqrt{2}}{3}$	24;112	38.5(52.74%)
月之暗面kimi	C	B	B	B	?	C	B	C	BC	A	B	95	$\frac{-4\sqrt{2}}{9}$	1;163	38(52.05%)
GLM4	C	B	C	B	A	C	C	?	BC	ACD	A	95	$\frac{-2\sqrt{2}}{3}$	24;169	36.5(50.00%)
Baichuan4	C	B	C	B	A	B	?	C	BC	D	C	95	$\frac{4(\sqrt{2}+1)\sqrt{49-32\sqrt{2}}}{2049}$	24;117	35.5(48.63%)
DeepSeek-V2-Chat	C	B	?	A	A	C	B	A	BC	ACD	AB	95	$\frac{\sqrt{2}}{2}$	576;110	31(42.47%)
MiniMax海螺	C	A	B	BC	C	C	D	D	BC	?	BC	95	9	24;142	23.5(32.19%)

两场平均分最好70%，最差 25%，填空题成绩更差

# 大模型“参加”2024高考数学情况

请完成下面一道选择题，每个小题四个选项中，只有一项是符合题目要求的：1、已知集合  $A = \{x \mid -5 < x^3 < 5\}$ ,  $B = \{-3, -1, 0, 2, 3\}$ , 则  $A \cap B = \{ \}$

- A.  $\{-1, 0\}$
- B.  $\{2, 3\}$
- C.  $\{-3, -1, 0\}$
- D.  $\{-1, 0, 2\}$

解：由题意得， $\therefore A = \{x \mid -5 < x^3 < 5\} = \{x \mid -1 < x < \sqrt[3]{5}\}$ , 则  $A \cap B = \{0, 2\}$ .

故选D。

重新回答



即便正确回答的题目，计算过程和答案不相符的比例很高

# 大模型“参加”2024高考数学情况

请完成下面一道选择题，每个小题四个选项中，只有一项是符合题目要求的：若  $\frac{z}{z-1} = 1+i$ ，则  $z = ()$

解：由  $\frac{z}{z-1} = 1+i$ ，得  $z = (1+i)(z-1)$ ,

即  $z = z - 1 + zi - i$ ,

所以  $zi = -1 - i$ ,

则  $z = \frac{-1-i}{i} = \frac{(-1-i)i}{i^2} = -1+i$ .

故选B。

请完成下面一道选择题，每个小题四个选项中，只有一项是符合题目要求的：2、若  $\frac{z}{z-1} = 1+i$ ，则  $z = ()$

- A.  $-1 - i$
- B.  $-1 + i$
- C.  $1 - i$
- D.  $1 + i$

2

解：设  $z = a+bi$ ,

则有  $\frac{a+bi}{a+bi-1} = 1+i$ ,

即  $a+bi = (1+i)(a+bi-1)$ ,

整理得  $a+bi = (a-b)+(a+b-1)i$ ,

所以  $\begin{cases} a = a - b \\ b = a + b - 1 \end{cases}$

解得  $a = 1$ ,  $b = -1$ ,

所以  $z = 1 - i$ .

故选C。

输入形式的微小不同，结果相差很大

# 大模型“参加”美国数学奥赛情况

USAMO完美契合评估LLM的目标：题目难度高、要求完整证明过程才能得分，且未经公开数据污染。

Benchmark (Metric)		Claude-3.5-Sonnet-1022	GPT-4o 0513	DeepSeek V3	OpenAI o1-mini	OpenAI o1-1217	DeepSeek R1
Math	AIME 2024 (Pass@1)	16.0	9.3	39.2	63.6	79.2	<b>79.8</b>
	MATH-500 (Pass@1)	78.3	74.6	90.2	90.0	96.4	<b>97.3</b>
	CNMO 2024 (Pass@1)	13.1	10.8	43.2	67.6	-	<b>78.8</b>



Overall	AIME 2025 I	AIME 2025 II	HMMT February 2025	USAMO 2025							
Model				Acc	Cost	1	2	3	4	5	6
DeepSeek-R1				4.76%	\$2.03	7%	0%	0%	21%	0%	0%
gemini-2.0-flash-thinking				4.17%	N/A	21%	0%	0%	0%	4%	0%
Claude-3.7-Sonnet (Think)				3.65%	\$9.03	7%	7%	0%	0%	0%	8%
QwQ-32B				2.98%	\$0.42	18%	0%	0%	0%	0%	0%
o1-pro (high)				2.83%	\$203.44	7%	0%	0%	0%	4%	6%
o3-mini (high)				2.08%	\$1.11	7%	2%	0%	0%	0%	4%

ETH Zurich 研究团队：实际上，LLM几乎从未没有学会数学证明！

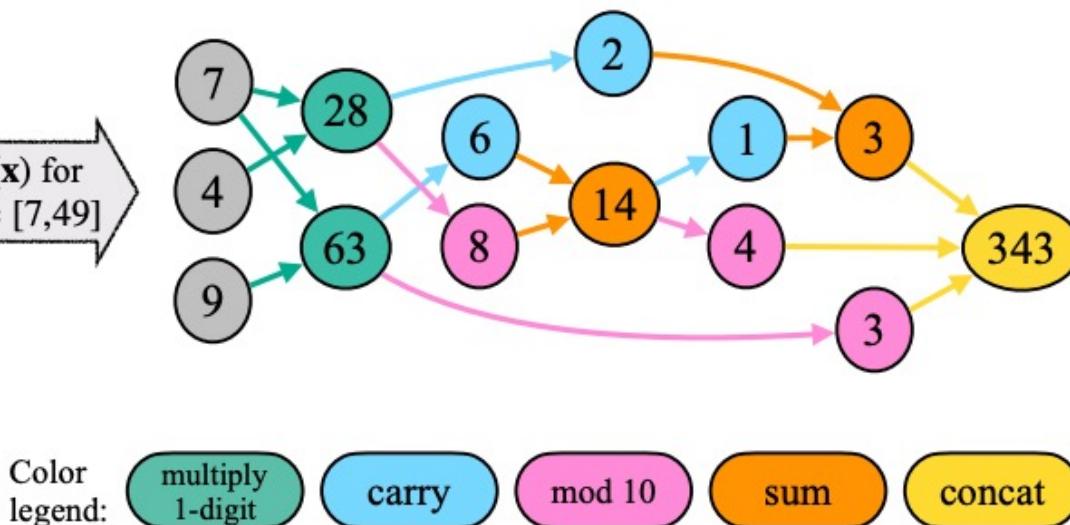
# 归纳推理的能力如何呢？

乘法的步骤：

```
function multiply (x[1..p], y[1..q]):  
    // multiply x for each y[i]  
    for i = q to 1  
        carry = 0  
        for j = p to 1  
            t = x[j] * y[i]  
            t += carry  
            carry = t // 10  
            digits[j] = t mod 10  
            summands[i] = digits  
  
    // add partial results (computation not shown)  
    product =  $\sum_{i=1}^q$  summands[q+1-i] · 10i-1  
    return product
```

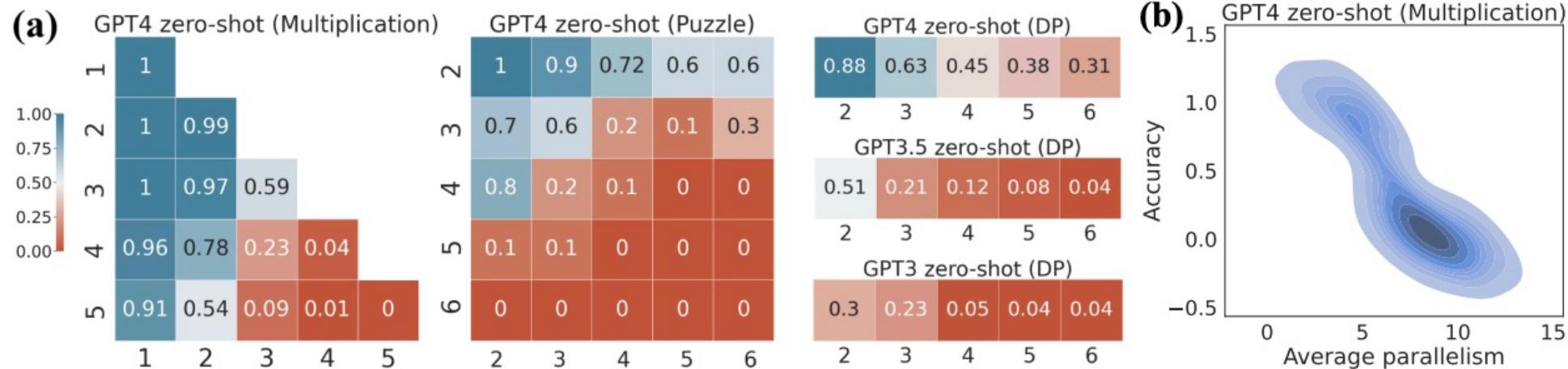
$A(\mathbf{x})$

$A(\mathbf{x})$  for  
 $\mathbf{x} = [7, 49]$



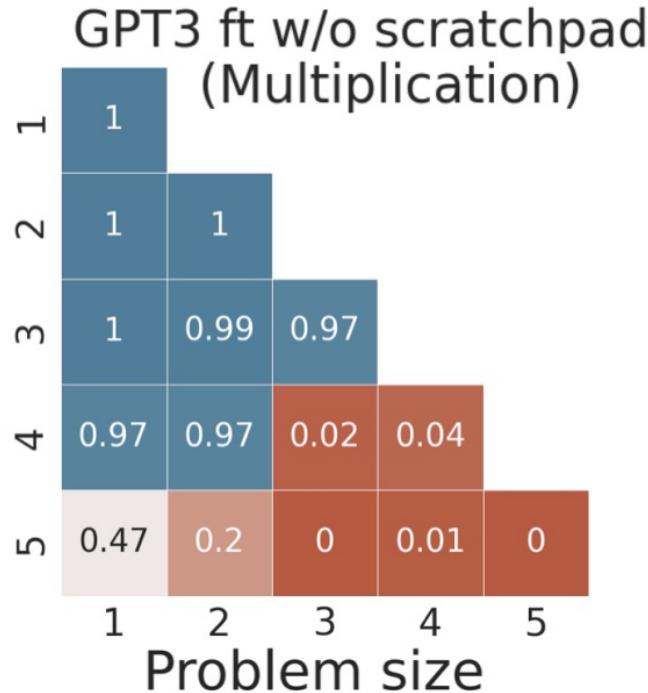
Dziri, Nouha, et al. "Faith and fate: Limits of transformers on compositionality." *Advances in Neural Information Processing Systems* 36 (2024) AllenAI

# 当任务复杂程度增大时，模型的准确率接近为0



Dziri, Nouha, et al. "Faith and fate: Limits of transformers on compositionality." *Advances in Neural Information Processing Systems* 36 (2024) AllenAI

# 归纳推理的能力如何呢？



使用 **180 万** 数据训练GPT3模型

GPT3经过充分调整，在特定任务数据上达到了一定的问题规模。蓝色区域表示分布内样本，红色区域表示OOD样本。

所有  $k_1$  和  $k_2$  符合  $1 \leq k_1, k_2 \leq 4$  并且  $k_1 \cdot k_2 \leq 9$ ;

# 训练过程中简单加入过程作用也十分有限

Question: What is 35 times 90?

Scratchpad: Let's perform the multiplication step by step:

Let's multiply 35 by the digit in the ones place of 90, which is 0.

1. Multiply 0 by the digit in the ones place of 35, which is 5. This gives  $5 \times 0 = 0$ . Write down the result 0.

2. Multiply 0 by the digit in the tens place of 35, which is 3. This gives  $3 \times 0 = 0$ . Write down the result 0.

3. The partial product for this step is A=0 which is the concatenation of the digits we found in each step.

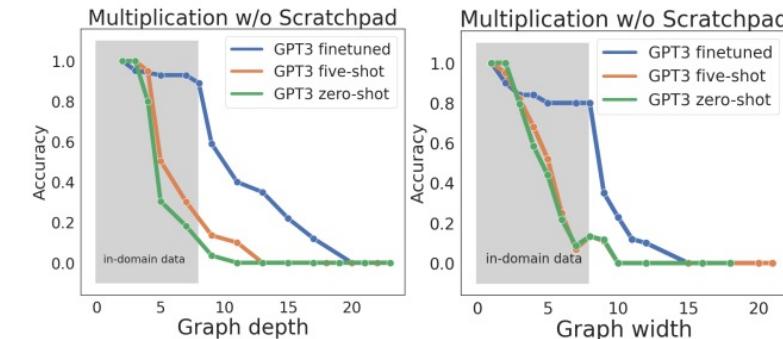
Now, let's multiply 35 by the digit in the tens place of 90, which is 9.

4. Multiply 9 by the digit in the ones place of 35, which is 5. This gives  $5 \times 9 = 45$ . Write down the result 5 and carry over the 4 to the next step.

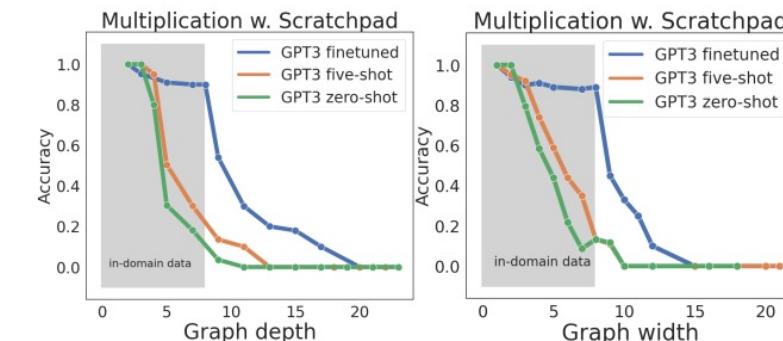
5. Multiply 9 by the digit in the tens place of 35, which is 3. Add the carryover from the previous step to account for this. This gives  $(3 \times 9) + 4 = 31$ . Write down the result 31.

6. The partial product for this step is B=315 which is the concatenation of the digits we found in each step.

Now, let's sum the 2 partial products A and B, and take into account the position of each digit: A=0 (from multiplication by 0) and B=315 (from multiplication by 9 but shifted one place to the left, so it becomes 3150). The final answer is  $0 \times 1 + 315 \times 10 = 0 + 3150 = 3150$ .

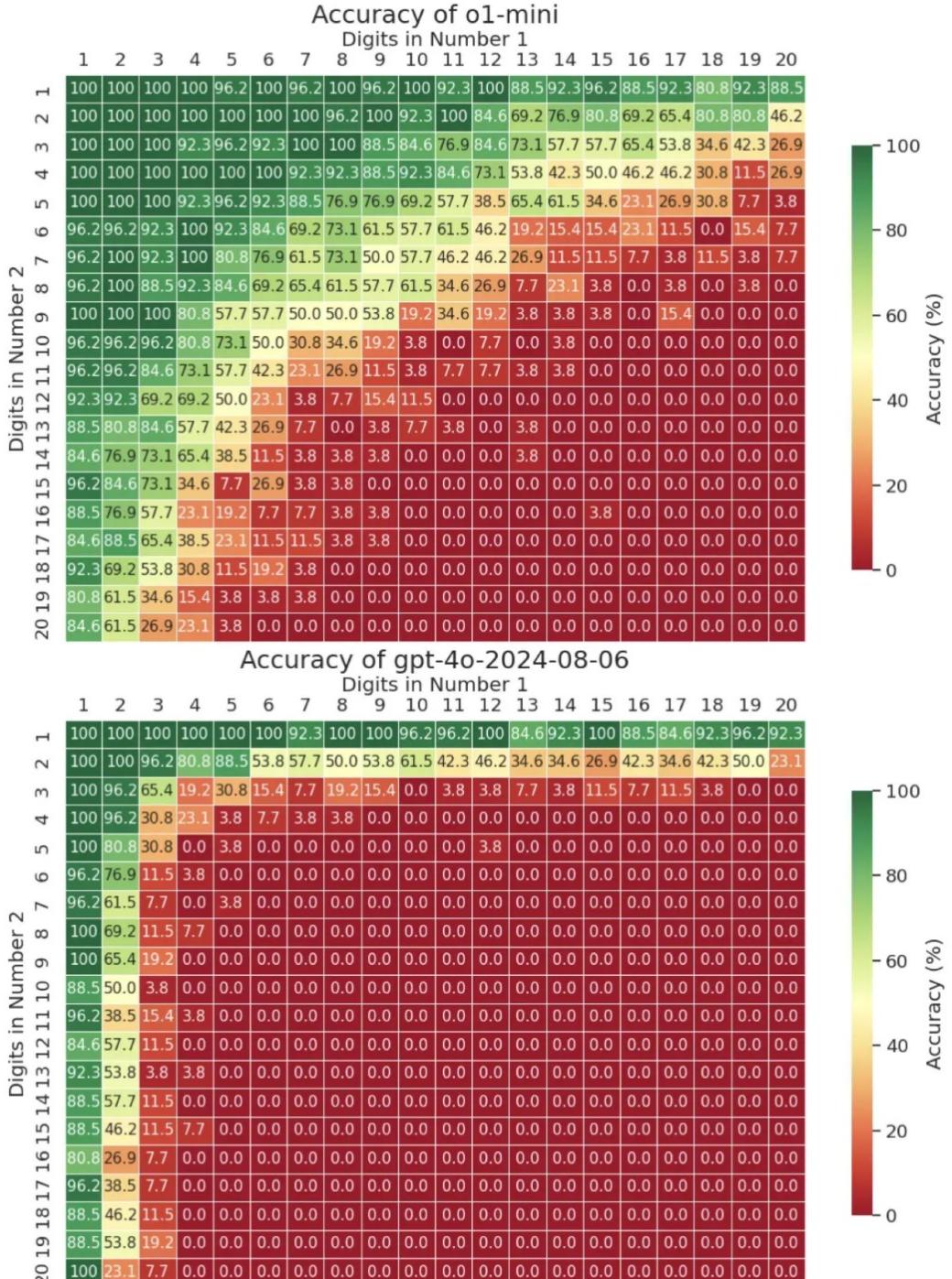
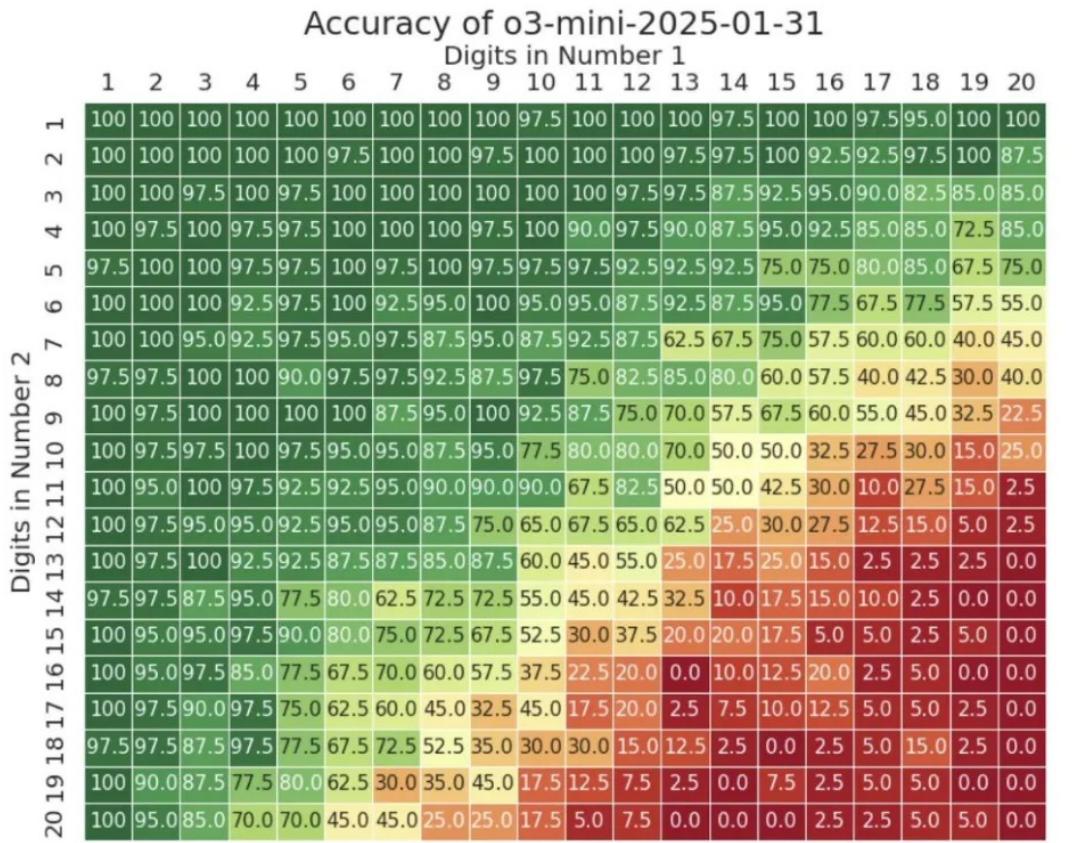


(a) Results on question-answer pairs.



(b) Results on question-scratchpad pairs.

# 乘法2025-01-31报告



# 大模型工具调用评测



**Get\_Weather**: This tool is used for fetching information weather for specified location.

Parameters:

*location* (string): Designated location, default is current location.



Please tell me the weather in the New York.



Get\_Weather (location = "New York")



**ABC**: This tool is used for fetching information weather for specified location.

Parameters:

*location* (string): Designated location, default is current location.



Please tell me the weather in the New York.



I'm sorry, but as a language model, I don't have access to weather information.



Models	Open-Source LLMs				Closed-Source LLMs		Human
	ToolLLaMA-2-7B-v1	ToolLLaMA-2-7B-v2	NexusRaven-13B-v1	NexusRaven-13B-v2	GPT-3.5-turbo	GPT-4	
<i>Tool Selection</i>							
Clean	66.67	70.48	55.24	73.33	75.24	<b>80.00</b>	88.57
Slight	57.62	65.71	52.86	76.19	59.05	<b>77.14</b>	88.57
Medium	56.67	59.52	53.33	72.38	69.52	<b>84.29</b>	88.57
Heavy	43.33	46.67	44.29	<b>62.38</b>	56.19	60.00	85.71
Union	44.76	43.81	42.86	56.19	53.33	<b>58.10</b>	85.71
<i>Parameter Identification</i>							
Clean	45.71	43.81	15.24	<b>56.19</b>	47.62	52.38	88.57
Slight	40.95	40.00	17.14	<b>56.67</b>	28.10	44.29	85.71
Medium	38.10	35.71	14.76	50.48	44.29	<b>53.81</b>	82.86
Heavy	28.10	27.14	10.00	<b>37.62</b>	24.29	32.86	80.00
Union	35.24	27.62	11.43	37.14	27.62	<b>39.05</b>	82.86
<i>Content Filling</i>							
Clean	28.57	25.71	1.90	37.14	30.48	<b>40.00</b>	74.29
Slight	24.29	23.81	3.33	<b>39.05</b>	20.00	35.71	74.29
Medium	22.38	20.95	1.90	33.81	30.48	<b>46.19</b>	71.43
Heavy	14.29	14.76	0.95	<b>30.00</b>	16.19	25.24	68.57
Union	16.19	16.19	1.90	22.86	18.10	<b>30.48</b>	71.43

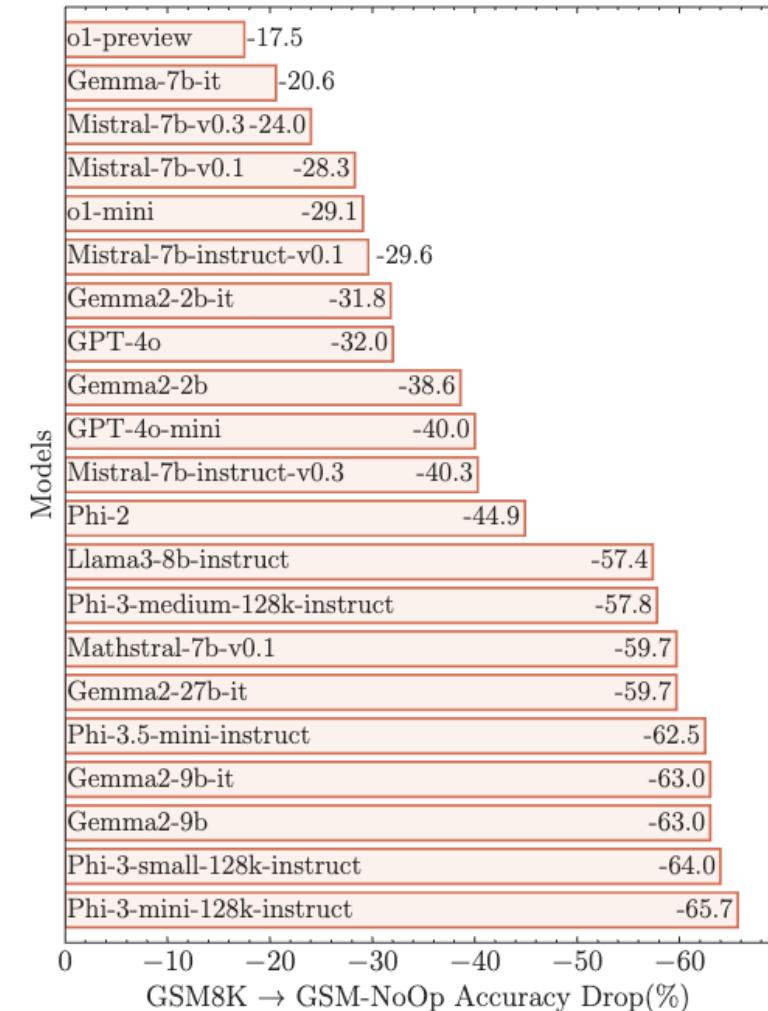
针对变形数据GPT4的准确率也大幅度下降

# GPT-o1 preview 小学数学题上也“翻车”了

原题：奥利弗在星期五摘了 44 个猕猴桃。然后在星期六摘了 58 个猕猴桃。星期天，他摘的猕猴桃数量是星期五的两倍。奥利弗有多少个猕猴桃？

变形题目：奥利弗在星期五摘了 44 个猕猴桃。然后在星期六摘了 58 个猕猴桃。星期天，他摘的猕猴桃数量是星期五的两倍，  
但其中 5 个比平均大小要小。奥利弗有多少个猕猴桃？

GPT-o1-mini 给出的答案是：…… 在星期天，这 5 个猕猴桃比平均大小要小。我们需要从星期天的总数中减去它们：88（星期天的猕猴桃） - 5（小猕猴桃） = 83 个猕猴桃。



全数据驱动的大模型

实现“推理”面临巨大挑战

实现因果学习是关键

硕士入学考试数学100分

小学应用题可能只有10分

仍然不知道strawberry有几个r



New chat



 New chat

## Message DeepSeek

 DeepThink (R1)

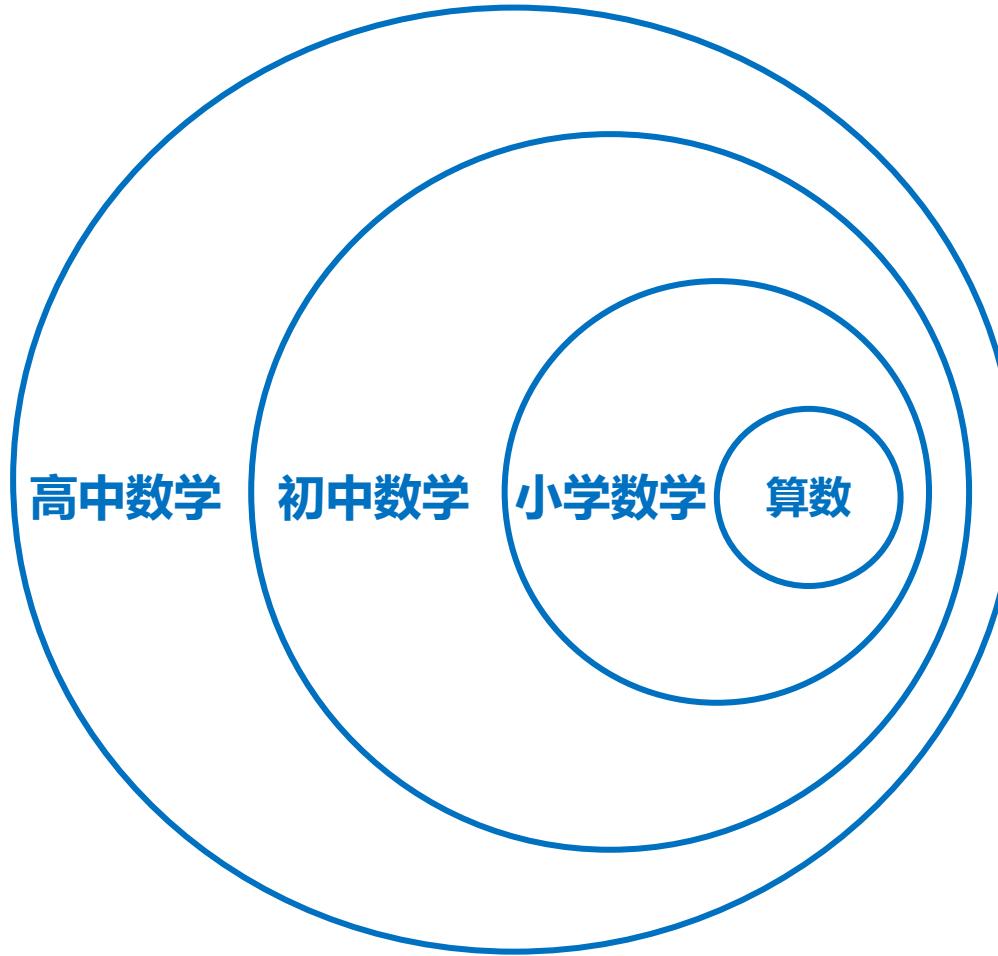
 Search



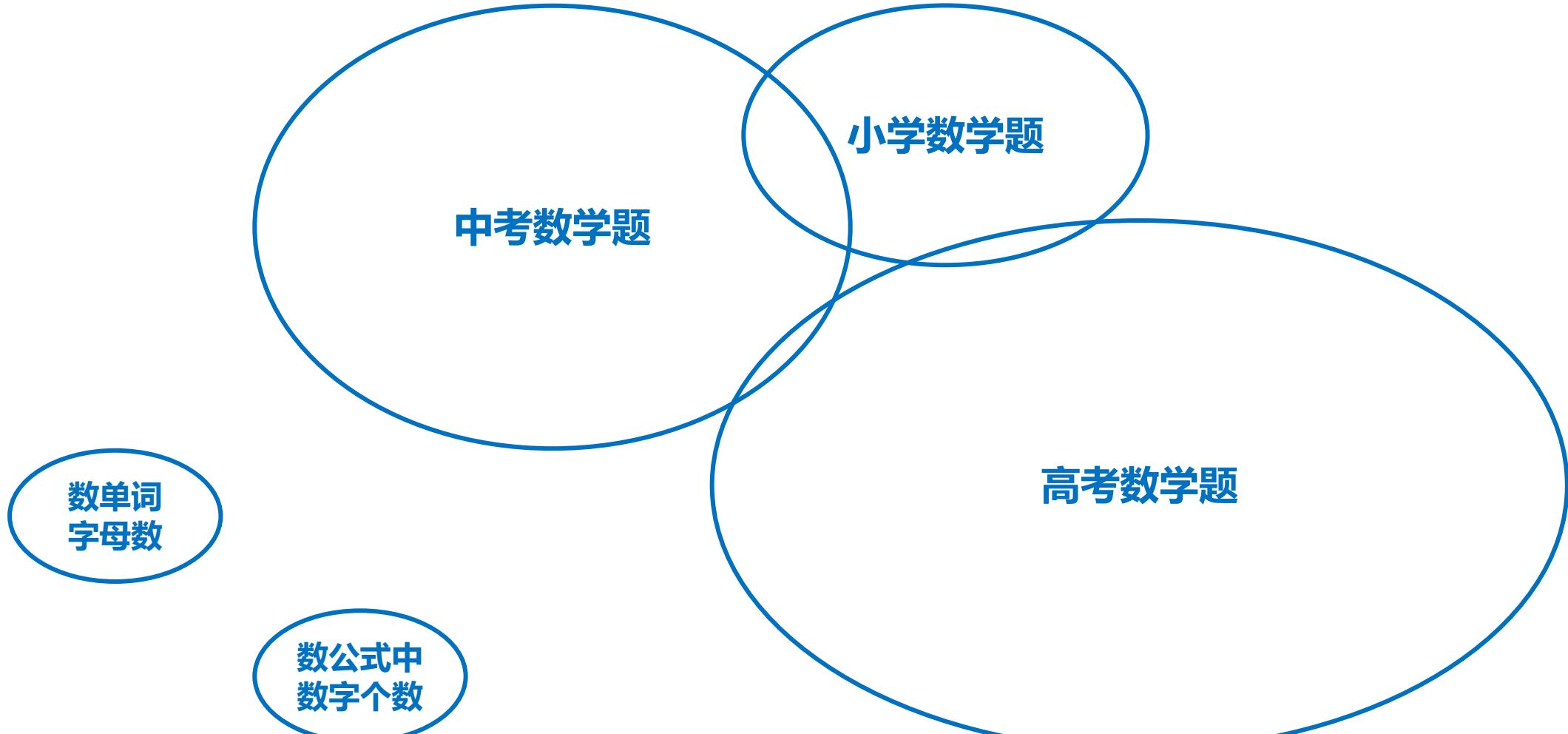
大模型依然是统计机器学习

不是“能力”的逐项提升

# 人类能力学习过程



# 大模型“能力”学习过程



单个模型可以处理数千种任务，  
但是仍需要逐项进行优化

# 大模型高泛化工具调用

Mitigating Adverse Effects	Prioritizing Key Tokens	Introducing a Reward Mechanism
<b>SYSTEM</b>	<b>SYSTEM</b>	<b>GOLDEN</b>
<ul style="list-style-type: none"> <li>def get_paper_details(paper_id:str, fields:str = None)</li> <li>def get_papers_by_keywords(query:str, fields:str = None, publicationType:str = None, openAccessPdf:str = None)</li> <li>def finish(answer:str)</li> </ul>	<ul style="list-style-type: none"> <li>get paper details</li> <li>get papers by key words</li> <li>finish</li> </ul>	<pre>get_papers_by_keywords(query = 'tool learning', venue = 'ICLR', limit = 1)</pre>
<b>USER</b>		<b>SAMPLES</b>
I am interested in case paper related to tool learning published in ICLR.		<ul style="list-style-type: none"> <li>S1: get_papers_by_keywords(query = 'tool learning', venue = 'ICLR', limit = 1)</li> <li>S2: get_papers_by_keywords(keyword = 'tool learning', venue = 'ICLR', limit = 1)</li> <li>S3: get_papers_by_keywords(query = 'learning', venue = 'ICLR', limit = 1)</li> <li>S4: get_papers_by_keywords(query = 'tool learning', venue = 'ICLR', limit = 1)</li> </ul>
<b>ASSISTANT</b>	<b>WEIGHT</b>	<b>REW ARD</b>
get_papers_by_keywords(keyword = 'tool learning', venue = 'ICLR', limit = 1) Blocked For Parameter Missing	$w_t^m = \begin{cases} \text{CLIP}\left(\frac{ NK_t }{ K_t }, 1, w_{\max}\right) & \text{if } t_i^m \in K_t \\ 1 & \text{otherwise} \end{cases}$	
<b>TOOL</b>	<b>ASSISTANT</b>	<b>SCORE</b>
get_papers_by_id missing 1 positional argument: query	<pre>-get papers by - key words ( query = "tool learning" , venue = "ICLR" , limit = 1 )</pre> <p><math> NK_t  = 26  K_t  = 2</math></p>	<ul style="list-style-type: none"> <li>S1: The tool 'get_papers_by_keys' is not given, so this sample contains a tool hallucination =&gt; -2</li> <li>S2: Tool 'get_papers_by_keywords' does not have a parameter named 'keyword', so this sample has a parameter hallucination =&gt; -0.8</li> <li>S3: The value 'learning' is not matched 'tool learning', so this sample has a content filling issue =&gt; -0.25</li> <li>S4: The sample is correct =&gt; 1</li> </ul>
<b>ASSISTANT</b>	<b>ASSISTANT</b>	<b>OBJECTIVE</b>
get_papers_by_detail(paper_id = 'ICLR2024TOOLLEARN')	<pre>-get paper details ( paper_id = "ICLR 2 0 2 4 TO OL LE ARN" )</pre> <p><math> NK_t  = 20  K_t  = 2</math></p>	$L_{PPT} = \sum_d \sum_{t_d^M} w_d^m \log p_M(t_d^m \mid o_{0..t_d^m-1}, q, T, t_{d-1}^{m-1})$ $PPO = \arg \max_M E_D \left[ \sum_{t_d} (R(t_d) - \beta \text{KL}(M(\cdot) \mid \mid M_{\text{ref}}(\cdot))) \right]$
<b>TOOL</b>	<b>ASSISTANT</b>	
name get_paper_by_detail is not defined	<pre>-finish ( answer = " I have found a paper published in ICLR for you, whose title is "ToolLLM: Facilitating Large Language Models to Master 1600+ Real-world APIs" )</pre> <p><math> NK_t  = 49  K_t  = 1</math></p>	
<b>ASSISTANT</b>	<b>OBJE</b>	
get_paper_details(paper_id = 'ICLR2024TOOLLEARN')		
<b>TOOL</b>	<b>ASSISTANT</b>	
{"detail": {"title": "ToolLLM: Facilitating Large Language Models to Master 1600+ Real-world APIs"}}		
<b>ASSISTANT</b>	<b>OPTIMIZATION</b>	
finish(answer = 'I have found an paper published in ICLR for you, whose title is "ToolLLM: Facilitating Large Language Models to Master 1600+ Real-world APIs."')		

消除负面影响、优化关键Token、引入奖励机制

1217条数据，7B 模型大幅度超越开源模型，在工具选择正确性维度超越GPT4o

# 大模型Code生成能力提升

```

import random
[---] : code scope
[green] : executed code
[red] : un-executed code

def test():
    ...
    for _ in range(int(input())):
        ...
        rows[0] = p[::-2]
        rows[1] = p[1::2]
        if sign(rows[0][0]) != sign(rows[1][0]):
            print(0)
            continue

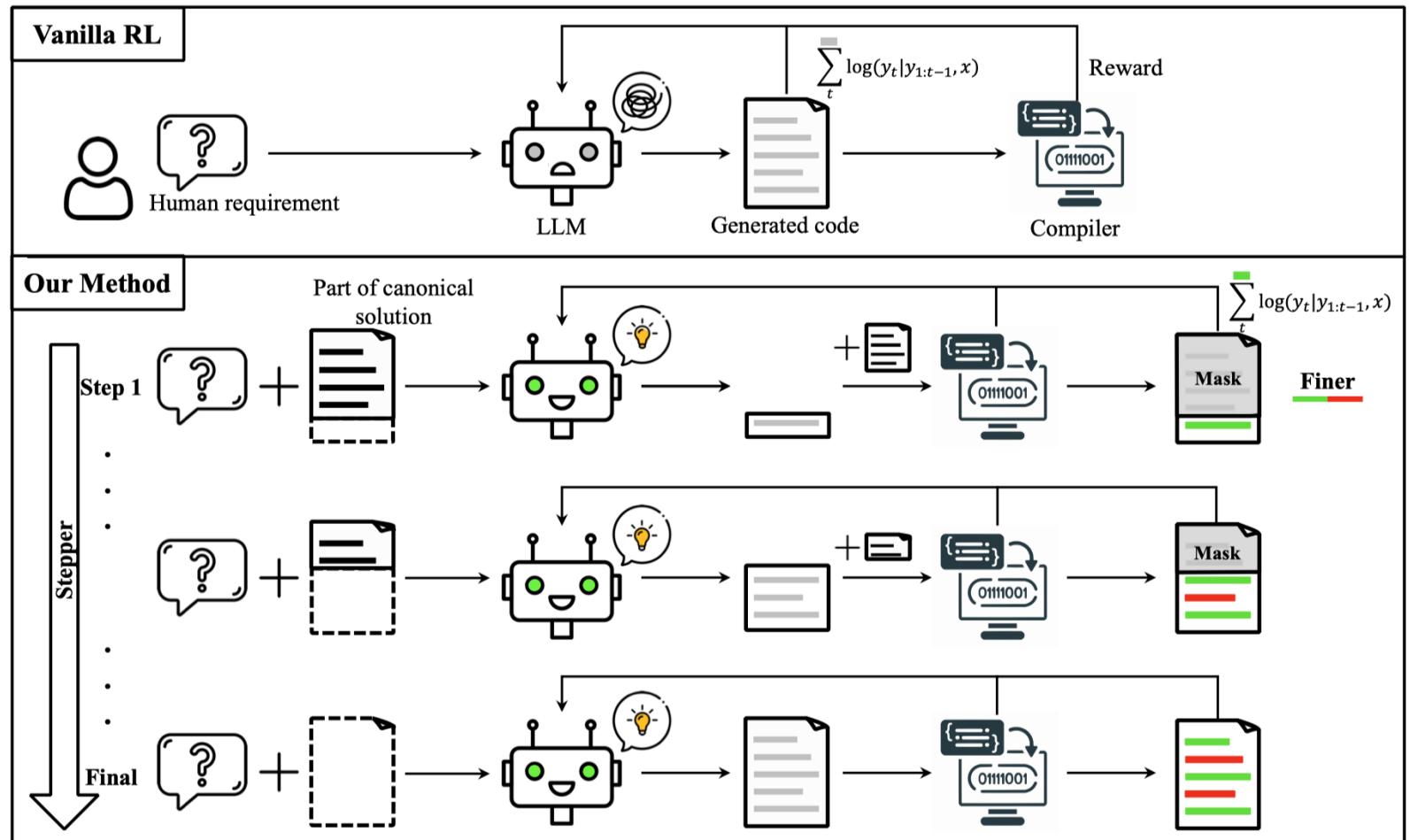
        for r in range(2, max_rows):
            for n in range(max_col - 1):
                rows[r][n] = rows[r - 1][0] * rows[r - 2][n + 1] - rows[r - 2][0] * rows[r - 1][n + 1]

        last = sign(rows[0][0])
        flag = 1
        for i in range(1, len(rows)):
            curr = sign(rows[i][0])
            if rows[r] == [0 for _ in range(max_col)]:
                for n in range(max_col):
                    rows[r][n] = rows[r - 1][n] * (max_pow +
4 - (r + 1) - 2 * (n + 1))

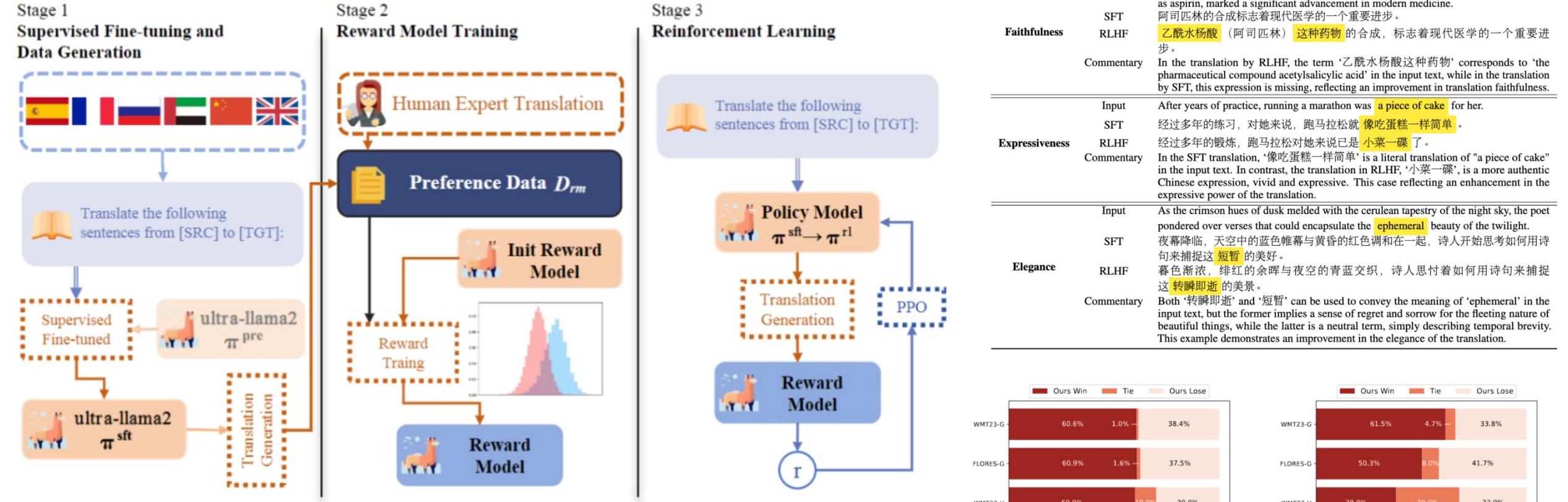
            elif rows[i][0] == 0:
                if any([x != 0 for x in rows[i]]):
                    flag = 0
                    break
                else:
                    curr = last

            if curr != last:
                flag = 0
                break
            last = curr

```



# RLHF推动翻译偏好建模：低成本实现“信达雅”



Advancing Translation Preference Modeling with RLHF: A Step Towards Cost-Effective Solution, arXiv 2024

<b>Faithfulness</b>	Input	The synthesis of the pharmaceutical compound acetylsalicylic acid, commonly known as aspirin, marked a significant advancement in modern medicine.
	SFT	阿司匹林的合成标志着现代医学的一个重要进步。
<b>Expressiveness</b>	Input	乙酰水杨酸（阿司匹林）这种药物的合成，标志着现代医学的一个重要进步。
	SFT	In the translation by RLHF, the term '乙酰水杨酸这种药物' corresponds to 'the pharmaceutical compound acetylsalicylic acid' in the input text, while in the translation by SFT, this expression is missing, reflecting an improvement in translation faithfulness.
<b>Elegance</b>	Input	After years of practice, running a marathon was a piece of cake for her.
	SFT	经过多年的练习，对她来说，跑马拉松就像吃蛋糕一样简单。
	Input	In the SFT translation, '像吃蛋糕一样简单' is a literal translation of "a piece of cake" in the input text. In contrast, the translation in RLHF, '小菜一碟'，is a more authentic Chinese expression, vivid and expressive. This case reflecting an enhancement in the expressive power of the translation.
	SFT	夜幕降临，天空中的蓝色帷幕与黄昏的红色调和在一起，诗人开始思考如何用诗句来捕捉这短暂的美好。
	Input	As the crimson hues of dusk melded with the cerulean tapestry of the night sky, the poet pondered over verses that could encapsulate the ephemeral beauty of the twilight.
	SFT	暮色渐浓，绯红的余晖与夜空的青蓝交织，诗人思忖着如何用诗句来捕捉这转瞬即逝的美景。
	Input	Both '转瞬即逝' and '短暂' can be used to convey the meaning of 'ephemeral' in the input text, but the former implies a sense of regret and sorrow for the fleeting nature of beautiful things, while the latter is a neutral term, simply describing temporal brevity. This example demonstrates an improvement in the elegance of the translation.
	SFT	Both '转瞬即逝' and '短暂' can be used to convey the meaning of 'ephemeral' in the input text, but the former implies a sense of regret and sorrow for the fleeting nature of beautiful things, while the latter is a neutral term, simply describing temporal brevity. This example demonstrates an improvement in the elegance of the translation.

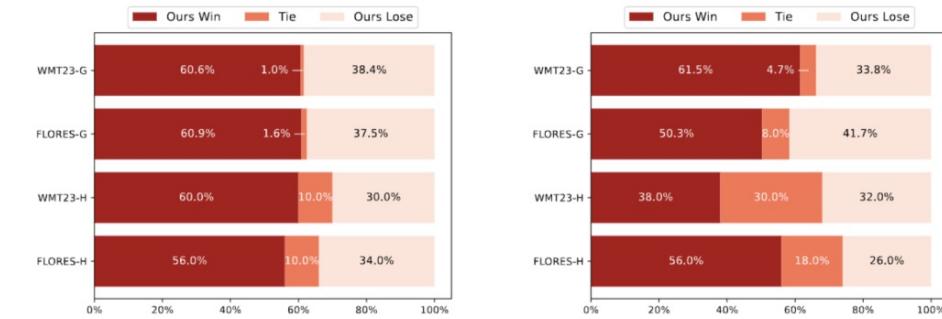
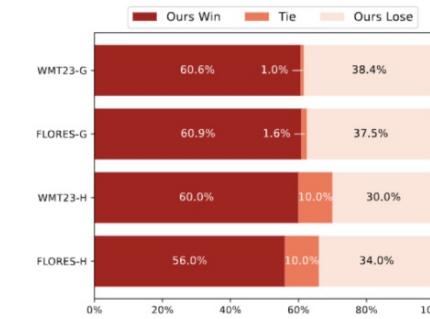
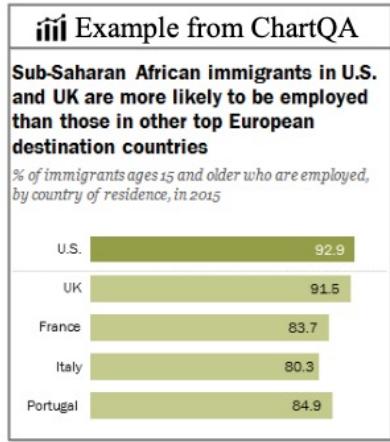


Figure 3: Comparison between preference optimized models and the SFT model on Task En→Zh. G and H represent GPT-4 and humans as evaluators, respectively.

Figure 4: Comparison between preference optimized models and the SFT model on Task Zh→En. G and H represent GPT-4 and humans as evaluators, respectively.

# 复杂图表内容问答



## Recognition Error

**Question:** Which country does the Dark green represent?

**Correct Answer:** The dark green represents U.S.

**Model Response:** By inspecting the bars and their corresponding percentages, we can determine which country's bar is dark green. According to the chart, the dark green bar represents France with a percentage value of 83.7%.

## Reasoning Error with Correct Recognition

**Question:** Which country ranked second highest in the bar chart?

**Correct Answer:** The country ranked second highest is UK.

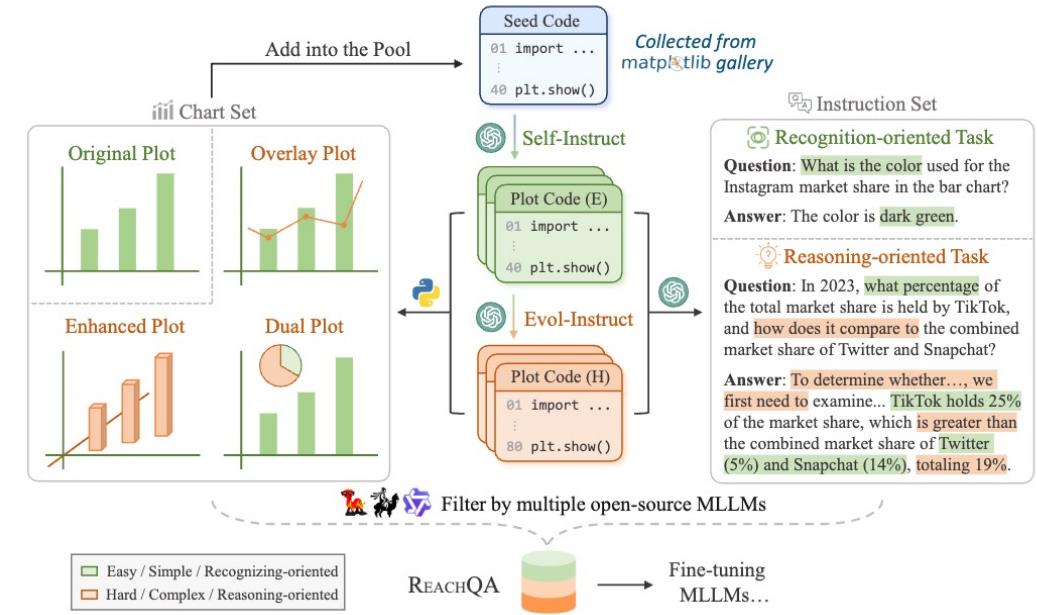
**Model Response:** To determine the country ranked second highest, we need to look at the values listed in descending order. The values are as follows:

- U.S.: 92.9 - UK: 91.5 - France: 83.7 - Italy: 80.3 - Portugal: 84.9 ✓

Upon examining these values, we can see that Portugal has the second-highest value at 84.9 percent. ✗

## 图表理解能力超越GPT4o

DISTILL VISUAL CHART REASONING ABILITY FROM LLMS TO MLLMS, arXiv 2024



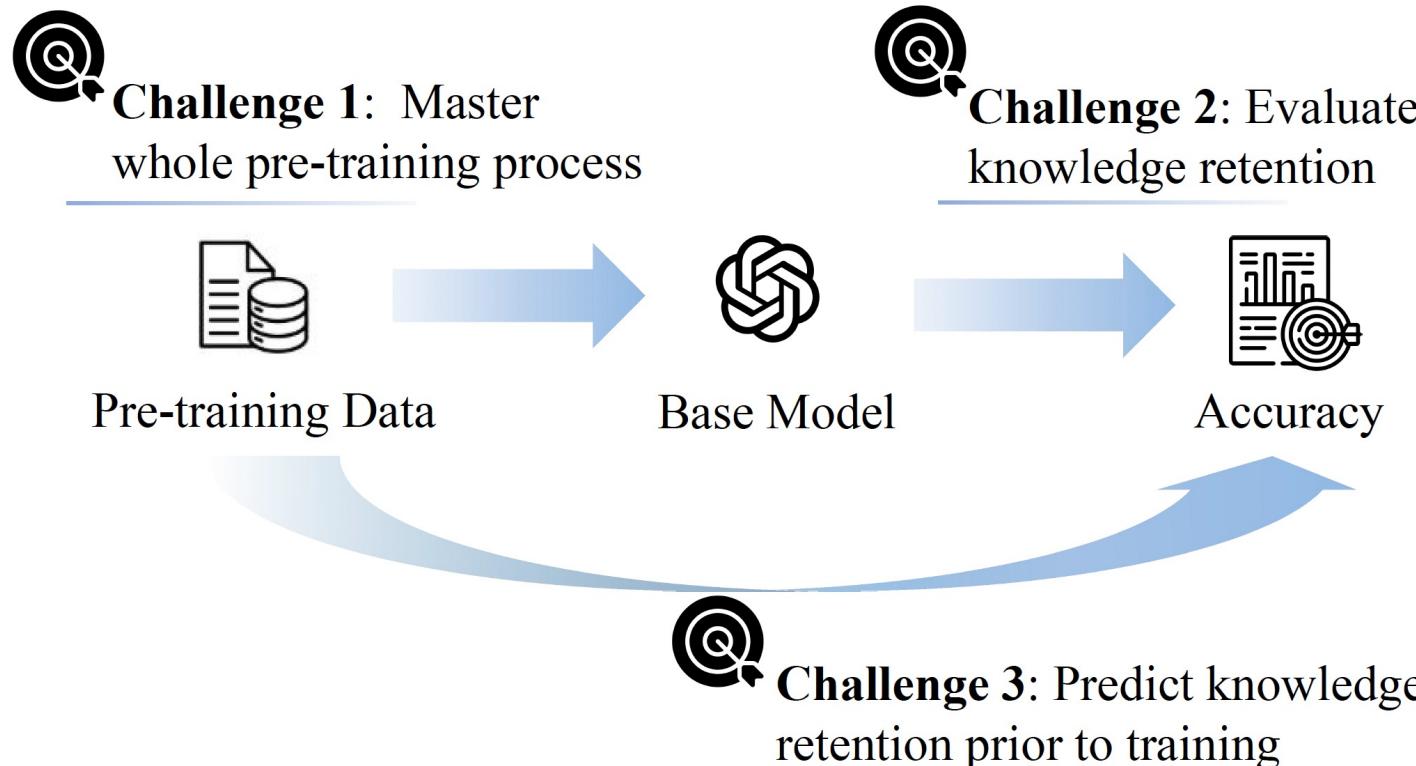
Datasets	Chart Properties				Q&A Properties			Dataset Properties		
	# Chart Type	# Chart Topic	Textual Format	Vis. Comp.	Temp. Free	Vis. Refer.	Rat. Annot.	Train Set	Test Set	Scal.
PlotQA (Methani et al., 2020)	3	-	Table	✗	✗	✓	✗	✓	✓	✗
ChartQA (Masry et al., 2022)	3	15	Table	✗	✓	✓	✗	✓	✓	✗
OpenCQA (Kantharaj et al., 2022)	5	10	Caption	✗	✓	✗	✓	✓	✓	✗
MathVista (Lu et al., 2024)	-	-	-	✗	✓	✗	✗	✗	✓	✗
CharXiv (Wang et al., 2024c)	-	-	-	✓	✗	✓	✓	✗	✓	✗
ChartBench (Xu et al., 2023)	9 / 42	-	Table	✗	✗	✓	✗	✓	✓	✓
ChartX (Xia et al., 2024)	18	22	Code*	✗	✓	✗	✗	✓	✓	✓
MMC (Liu et al., 2024a)	6	5	Caption	✓	✓	✗	✓	✓	✓	✓
ChartLlama (Han et al., 2023)	10	-	Table	✗	✓	✗	✓	✓	✓	✓
ChartAst (Meng et al., 2024)	9	-	Table	✗	✗	✗	✓	✓	✓	✓
ChartInstruct (Masry et al., 2024a)	-	-	Table	✗	✓	✗	✓	✓	✓	✓
ChartGemma (Masry et al., 2024b)	-	-	-	✗	✓	✓	✓	✓	✓	✓
REACHQA (ours)	10 / 32	∞	Code	✓	✓	✓	✓	✓	✓	✓

## 2. 大模型能力来源思考

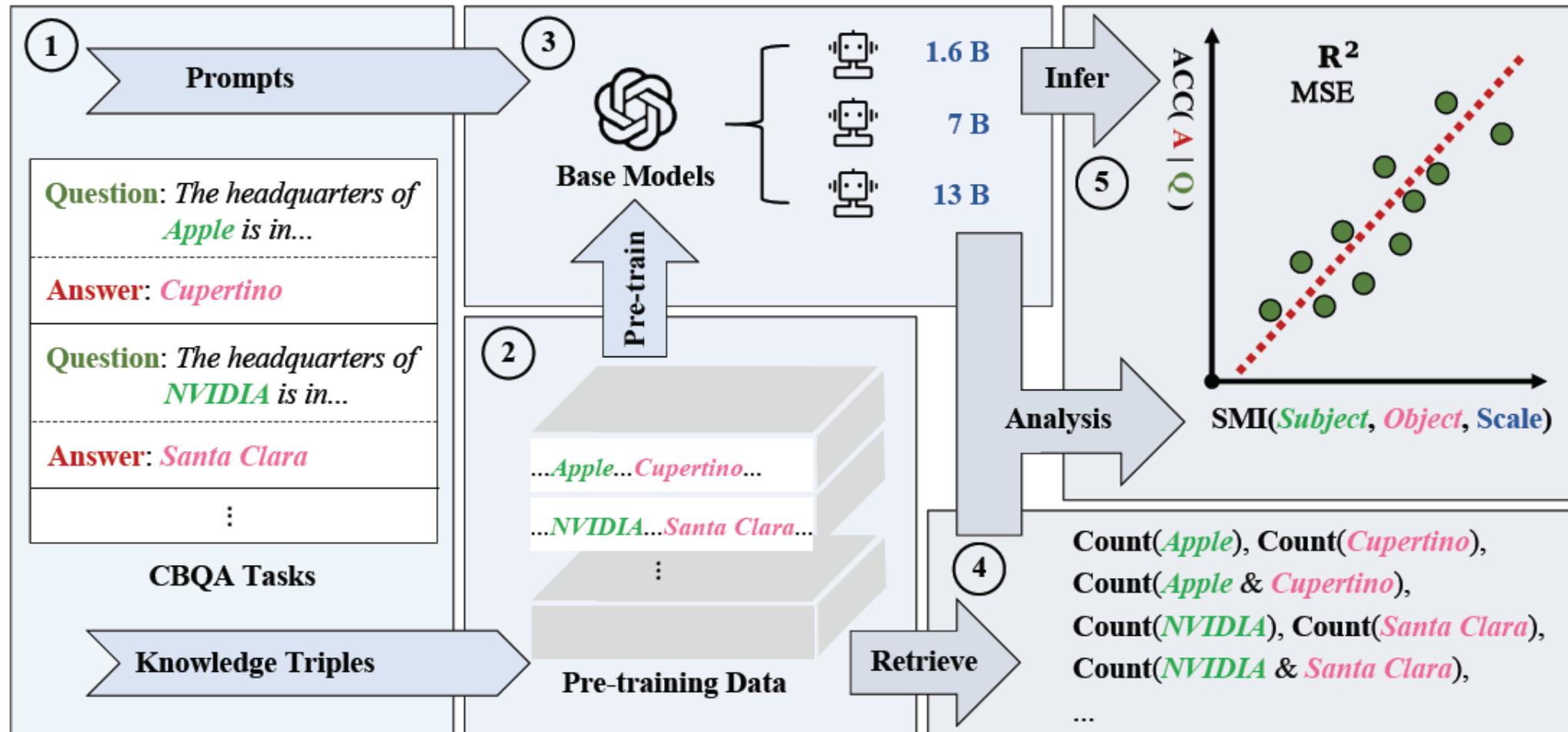
# 1. 直接通过训练语料的统计就可以计算得到知识记忆概率

We registered predictions for GPT-4's performance on HumanEval before training completed, using only information available prior to training. All but the 15 hardest HumanEval problems were split

From GPT-4 Technical Report



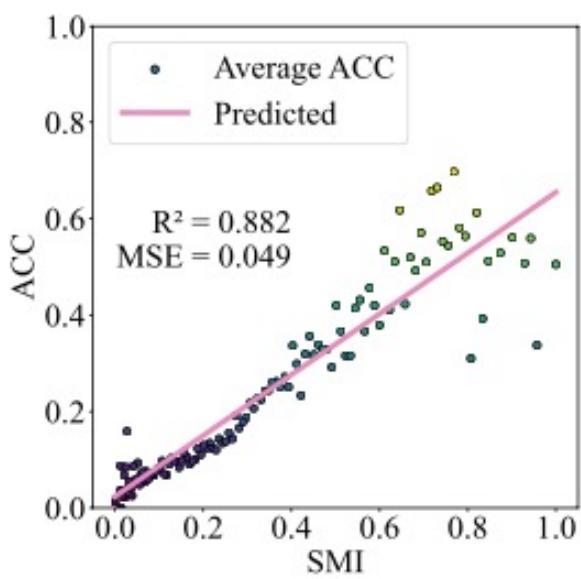
# 1. 直接通过训练语料的统计就可以计算得到知识记忆概率



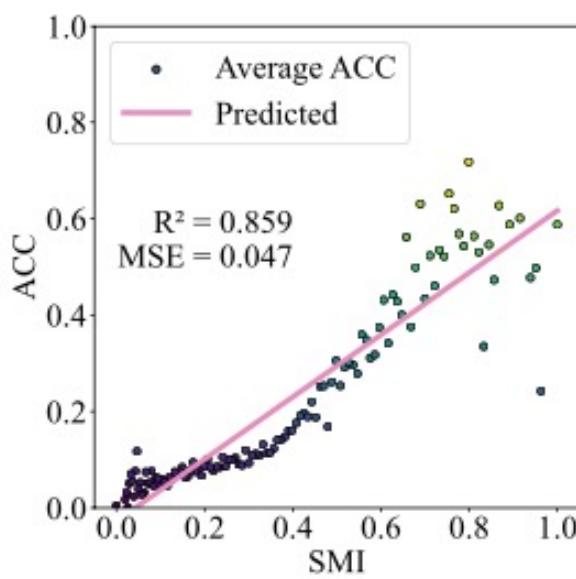
# 1. 直接通过训练语料的统计就可以计算得到知识记忆概率

$$I(s, o) = P(s, o) \log \left( \frac{P(s, o)}{P(s)P(o)} \right)$$

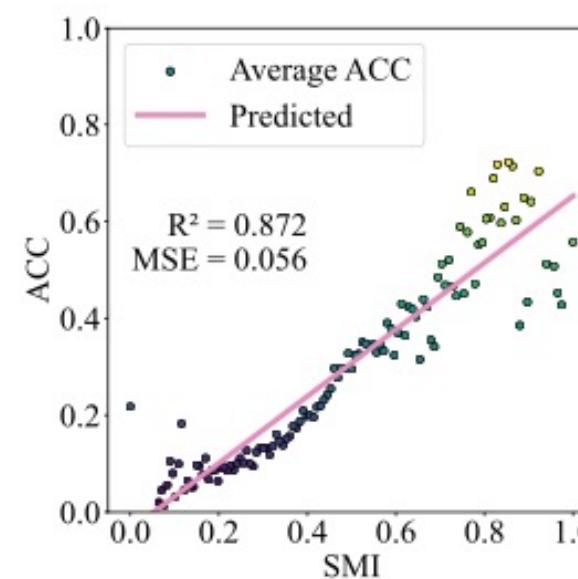
$$SMI(s, o, \Phi) = Norm(\log(I(s, o)))^{1+\frac{1}{\Phi}}$$



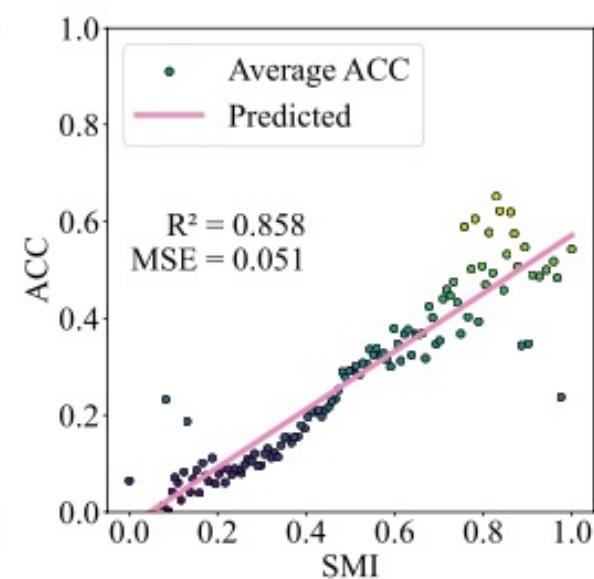
(a) 1.1B



(b) 1.6B



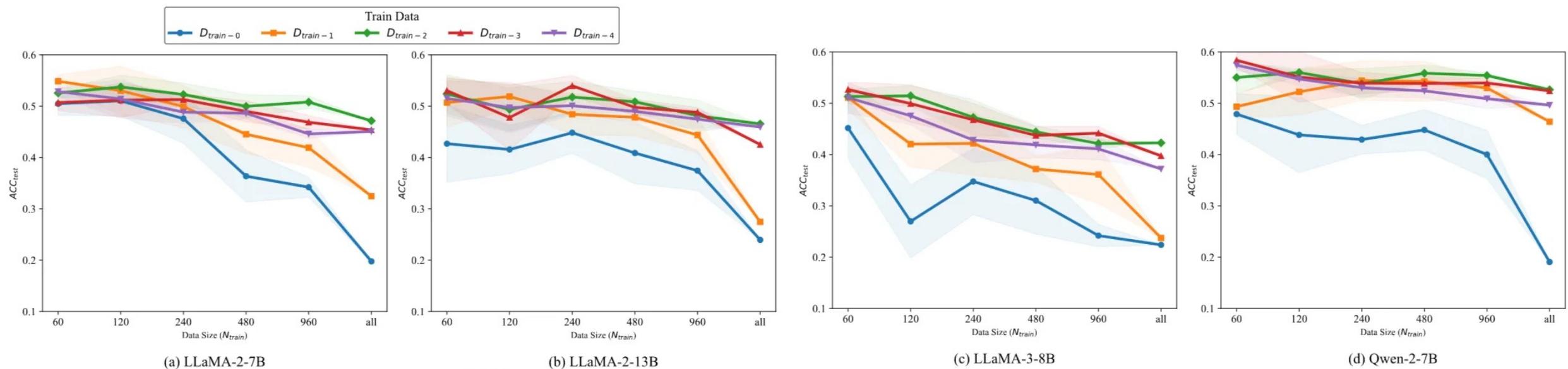
(c) 7B



(d) 13B

## 2. 有监督微调阶段仅需少量数据，并且要跟模型紧密配合

完成某个领域知识问答仅需 60 条训练数据



不同LLMs的数据需求差异巨大

## 2. 有监督微调阶段仅需少量数据

**问题 1：**在 SFT 阶段需要多少数据才能让 LLMs 执行 QA 任务？

**问题 2：**不同的 SFT 数据如何影响 LLMs 在 QA 任务中的表现？

**问题 3：**不同 LLMs 在 SFT 阶段的数据要求有何不同？



## 2. 有监督微调阶段仅需少量数据

模型知识记忆情况判断

knowledge	map(subject, relation)	map(object)
Painblanc <i>subject</i>	<i>Painblanc is located in</i>	France
located in <i>relation</i>	<i>The country where Painblanc is found is</i>	France
France <i>object</i>	<i>Painblanc is a part of</i>	France

$$R_k^M = \frac{\sum_{i=1}^{N_{map}} \sum_{j=1}^{N_{sample}} I(y_i \subseteq M_{base}^j(x_i))}{N_{map} \times N_{sample}}$$

## 2. 有监督微调阶段仅需少量数据

(a) LLaMA-2-7B ( $N_{train} = 60$ )

Train Data	ACC <sub>test-0</sub>	ACC <sub>test-1</sub>	ACC <sub>test-2</sub>	ACC <sub>test-3</sub>	ACC <sub>test-4</sub>	ACC <sub>test</sub>
$D_{train-0}$	<b>4.80</b> $\pm$ 1.11	34.00 $\pm$ 2.93	62.49 $\pm$ 4.83	73.36 $\pm$ 4.13	77.64 $\pm$ 6.02	50.46 $\pm$ 2.28
$D_{train-1}$	4.12 $\pm$ 1.61	<b>38.19</b> $\pm$ 4.54	69.49 $\pm$ 2.23	77.11 $\pm$ 3.43	85.32 $\pm$ 1.62	<b>54.85</b> $\pm$ 1.26
$D_{train-2}$	0.69 $\pm$ 0.35	24.57 $\pm$ 1.70	<b>70.57</b> $\pm$ 2.23	82.39 $\pm$ 1.65	84.56 $\pm$ 0.64	52.56 $\pm$ 0.81
$D_{train-3}$	0.37 $\pm$ 0.41	17.79 $\pm$ 2.33	67.45 $\pm$ 2.34	85.43 $\pm$ 2.13	82.45 $\pm$ 2.84	50.70 $\pm$ 1.57
$D_{train-4}$	0.40 $\pm$ 0.28	19.78 $\pm$ 1.67	69.77 $\pm$ 1.24	<b>87.36</b> $\pm$ 1.45	<b>86.84</b> $\pm$ 0.91	52.83 $\pm$ 0.99

(b) LLaMA-2-13B ( $N_{train} = 60$ )

Train Data	ACC <sub>test-0</sub>	ACC <sub>test-1</sub>	ACC <sub>test-2</sub>	ACC <sub>test-3</sub>	ACC <sub>test-4</sub>	ACC <sub>test</sub>
$D_{train-0}$	<b>7.62</b> $\pm$ 1.69	23.66 $\pm$ 4.29	53.17 $\pm$ 9.36	61.96 $\pm$ 11.69	67.00 $\pm$ 14.25	42.68 $\pm$ 7.45
$D_{train-1}$	7.49 $\pm$ 0.54	<b>35.95</b> $\pm$ 5.15	64.22 $\pm$ 5.00	65.98 $\pm$ 5.28	79.84 $\pm$ 9.43	50.69 $\pm$ 4.69
$D_{train-2}$	0.62 $\pm$ 0.50	21.36 $\pm$ 2.93	<b>72.80</b> $\pm$ 5.72	80.65 $\pm$ 5.45	86.47 $\pm$ 5.85	52.38 $\pm$ 3.70
$D_{train-3}$	0.45 $\pm$ 0.21	17.97 $\pm$ 1.81	71.16 $\pm$ 3.63	<b>86.60</b> $\pm$ 3.40	<b>88.87</b> $\pm$ 4.79	<b>53.01</b> $\pm$ 2.54
$D_{train-4}$	0.24 $\pm$ 0.22	16.66 $\pm$ 3.14	68.46 $\pm$ 4.83	83.44 $\pm$ 4.92	88.69 $\pm$ 4.82	51.50 $\pm$ 3.40

(c) LLaMA-3-8B ( $N_{train} = 60$ )

Train Data	ACC <sub>test-0</sub>	ACC <sub>test-1</sub>	ACC <sub>test-2</sub>	ACC <sub>test-3</sub>	ACC <sub>test-4</sub>	ACC <sub>test</sub>
$D_{train-0}$	<b>2.72</b> $\pm$ 1.44	26.05 $\pm$ 8.77	53.82 $\pm$ 10.32	62.90 $\pm$ 8.27	80.36 $\pm$ 9.57	45.17 $\pm$ 5.96
$D_{train-1}$	2.28 $\pm$ 1.23	<b>33.83</b> $\pm$ 7.50	66.35 $\pm$ 3.69	71.97 $\pm$ 4.08	81.02 $\pm$ 3.91	51.09 $\pm$ 2.33
$D_{train-2}$	0.19 $\pm$ 0.08	21.57 $\pm$ 1.71	<b>71.89</b> $\pm$ 1.49	80.23 $\pm$ 1.13	82.65 $\pm$ 5.39	51.31 $\pm$ 1.41
$D_{train-3}$	0.15 $\pm$ 0.03	19.03 $\pm$ 0.91	68.09 $\pm$ 3.64	<b>85.72</b> $\pm$ 2.69	90.30 $\pm$ 1.17	<b>52.66</b> $\pm$ 1.45
$D_{train-4}$	0.17 $\pm$ 0.07	16.39 $\pm$ 1.55	63.76 $\pm$ 6.47	83.58 $\pm$ 4.32	<b>91.54</b> $\pm$ 3.12	51.09 $\pm$ 3.00

(d) Qwen-2-7B ( $N_{train} = 60$ )

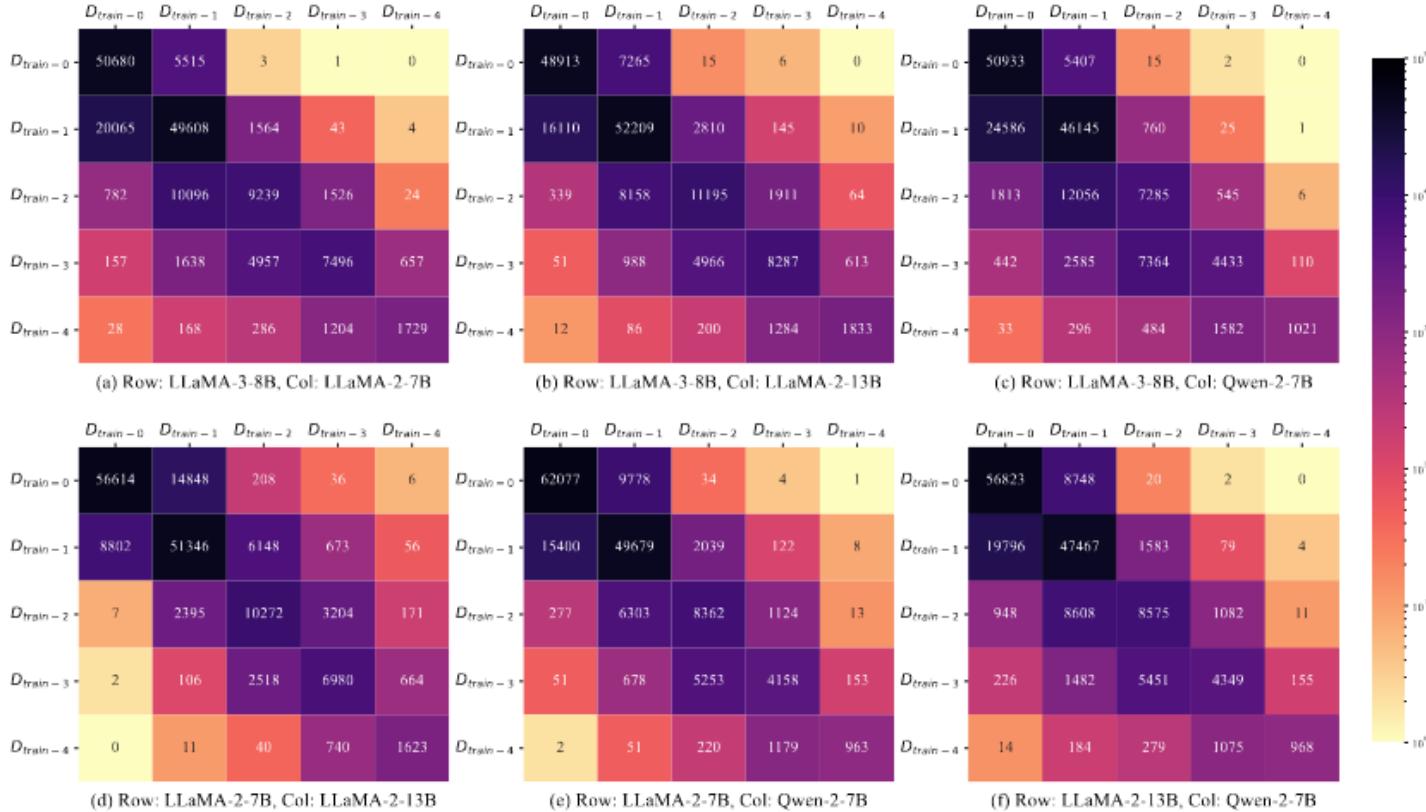
Train Data	ACC <sub>test-0</sub>	ACC <sub>test-1</sub>	ACC <sub>test-2</sub>	ACC <sub>test-3</sub>	ACC <sub>test-4</sub>	ACC <sub>test</sub>
$D_{train-0}$	<b>7.85</b> $\pm$ 0.67	34.88 $\pm$ 1.29	57.84 $\pm$ 4.28	66.17 $\pm$ 6.57	72.70 $\pm$ 8.35	47.89 $\pm$ 3.99
$D_{train-1}$	6.49 $\pm$ 0.61	<b>39.40</b> $\pm$ 2.10	60.68 $\pm$ 4.01	69.50 $\pm$ 5.15	70.57 $\pm$ 2.88	49.33 $\pm$ 2.23
$D_{train-2}$	0.36 $\pm$ 0.09	26.15 $\pm$ 1.76	80.16 $\pm$ 3.59	83.14 $\pm$ 4.35	85.30 $\pm$ 7.44	55.02 $\pm$ 3.12
$D_{train-3}$	0.48 $\pm$ 0.09	21.58 $\pm$ 0.54	<b>82.01</b> $\pm$ 1.89	<b>92.34</b> $\pm$ 1.00	95.48 $\pm$ 4.82	<b>58.38</b> $\pm$ 1.45
$D_{train-4}$	0.21 $\pm$ 0.04	19.55 $\pm$ 1.02	80.48 $\pm$ 1.24	90.84 $\pm$ 1.98	<b>96.00</b> $\pm$ 0.99	57.42 $\pm$ 0.92

**发现 1：**无论使用何种数据进行微调，LLM 都能对在预训练中记忆较好的知识提供更准确的答案。

**发现 2：**使用特定记忆水平的数据进行训练可提高 LLM 在该知识水平上的表现。

**发现 3：**总体而言，更有效的策略是使用记忆水平较高的数据进行 SFT。

## 2. 有监督微调阶段仅需少量数据

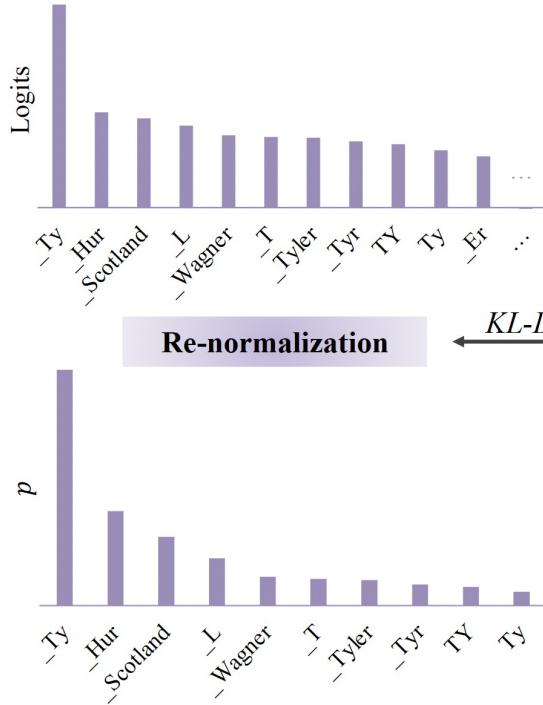


发现 4：不同 LLMs 的预训练语料库的差异导致知识分布的显著不同

### 3. 预训练模型参数与SFT模型关联

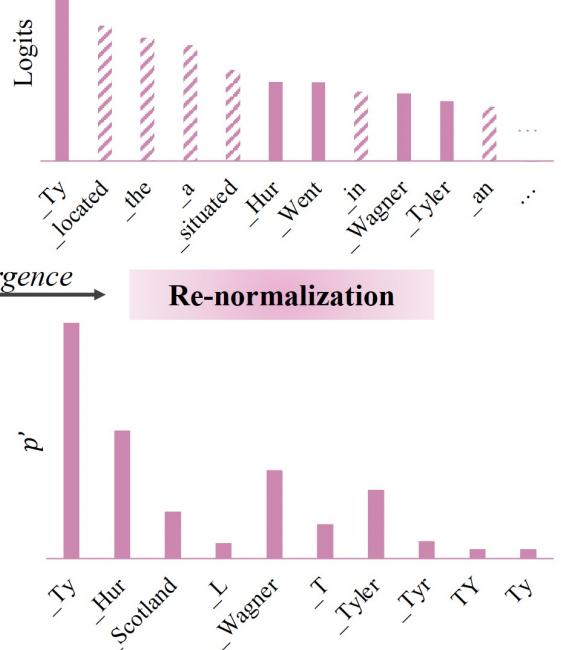
**Finetuned LLM**

<question> What is the capital of Bon Homme County? <answer> \_\_\_\_



**Pretrained LLM**

The capital of Bon Homme County is \_\_\_\_



Re-normalization

$\xleftarrow{KL\text{-Divergence}}$

$\xrightarrow{\text{Re-normalization}}$

Legend:  $D_{train-0}^M$  (red circle),  $D_{train-1}^M$  (purple square),  $D_{train-2}^M$  (green diamond),  $D_{train-3}^M$  (pink triangle),  $D_{train-4}^M$  (blue triangle)

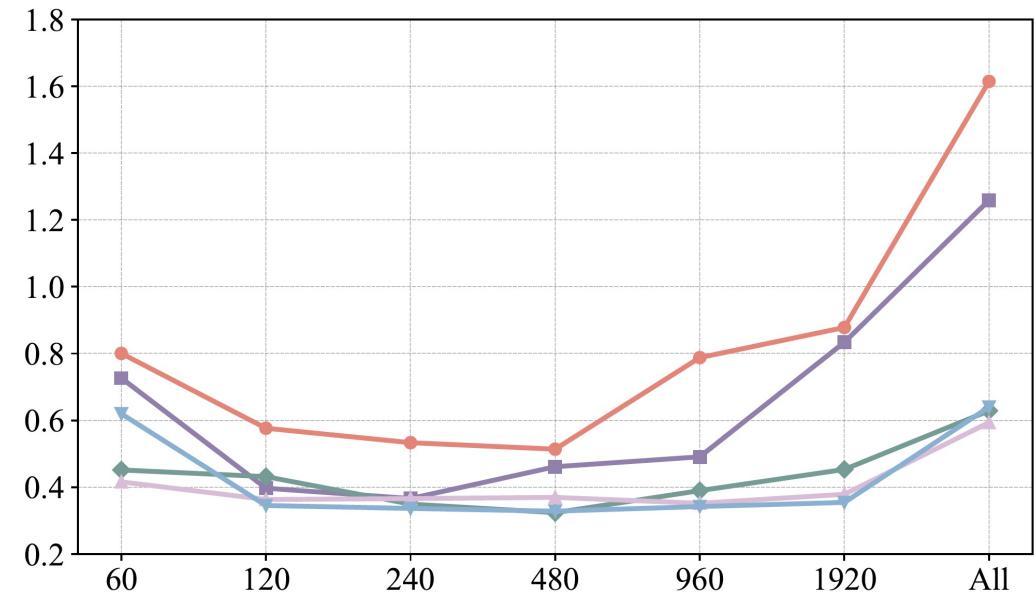


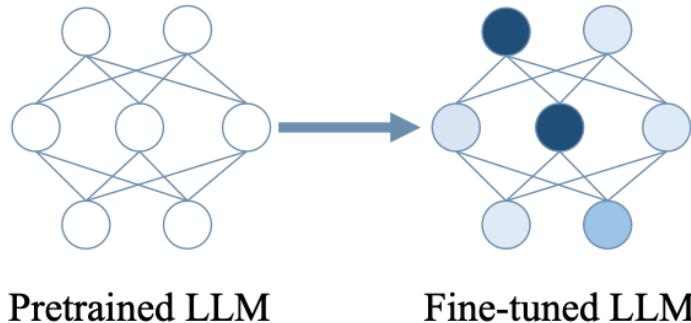
Figure 5: KL divergence of logits distribution between LLaMA-3-8B fine-tuned with different datasets and the pre-trained one.

使用记忆水平较低的数据进行 SFT，会大幅度改变模型参数

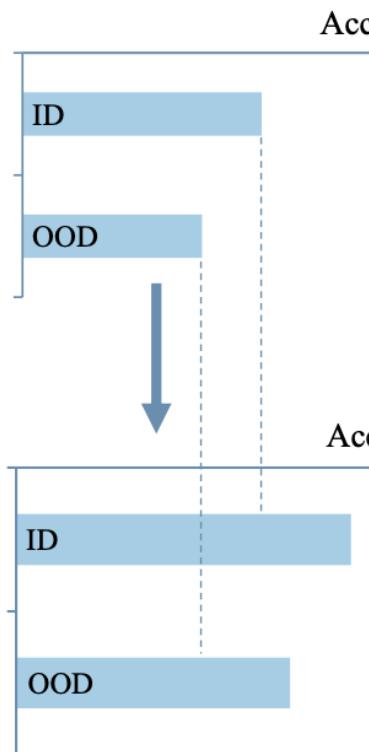
### 3. 预训练模型参数与SFT模型关联

尽可能少的改动预训练参数

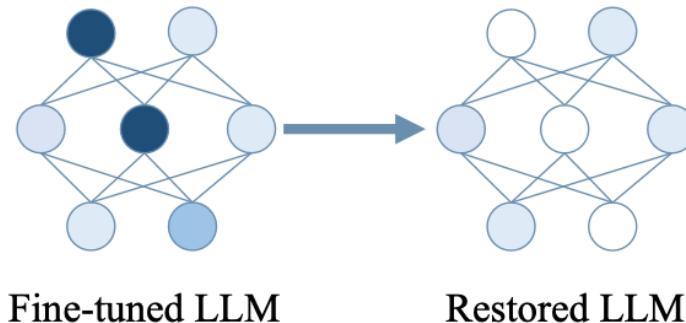
#### Supervised Fine-Tuning



#### Performance



#### Parameter Restoration

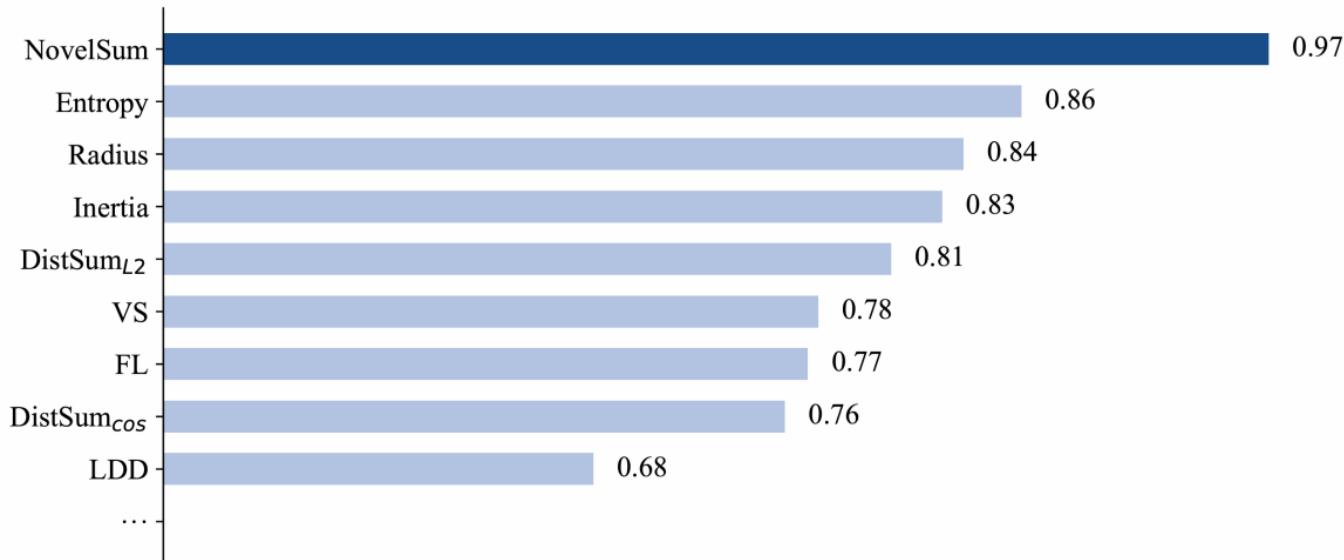


Restore	$\mathcal{D}_{\text{train}-0}^M$	$\mathcal{D}_{\text{train}-1}^M$	$\mathcal{D}_{\text{train}-2}^M$	$\mathcal{D}_{\text{train}-3}^M$	$\mathcal{D}_{\text{train}-4}^M$
<i>Number of Training Data: 240</i>					
0	55.33	57.96	59.32	59.12	53.97
1%	55.76	58.17	59.62	59.24	54.30
3%	56.64	58.52	59.77	59.40	54.31
5%	57.22	58.68	59.89	59.63	54.44
10%	58.32	59.45	60.40	59.83	54.69
20%	59.07	59.81	59.88	59.91	46.45
40%	59.77	33.40	42.44	11.20	23.83
60%	1.68	2.20	3.65	2.56	1.65
<i>Number of Training Data: 1920</i>					
0	44.96	52.43	58.80	57.70	55.22
1%	46.73	53.72	59.85	58.68	55.88
3%	48.53	55.01	60.56	59.23	56.76
5%	49.85	55.96	61.10	59.65	57.34
10%	52.10	57.14	61.67	60.02	58.24
20%	54.81	58.33	62.21	58.93	58.66
40%	55.44	22.06	59.97	6.92	56.50
60%	1.48	1.12	1.62	0.51	0.60

Performance of LLaMA-3-8B after restoring different scales of parameters across various fine-tuning datasets.

## 4. 多样性对SFT训练的影响

如何衡量多样性?



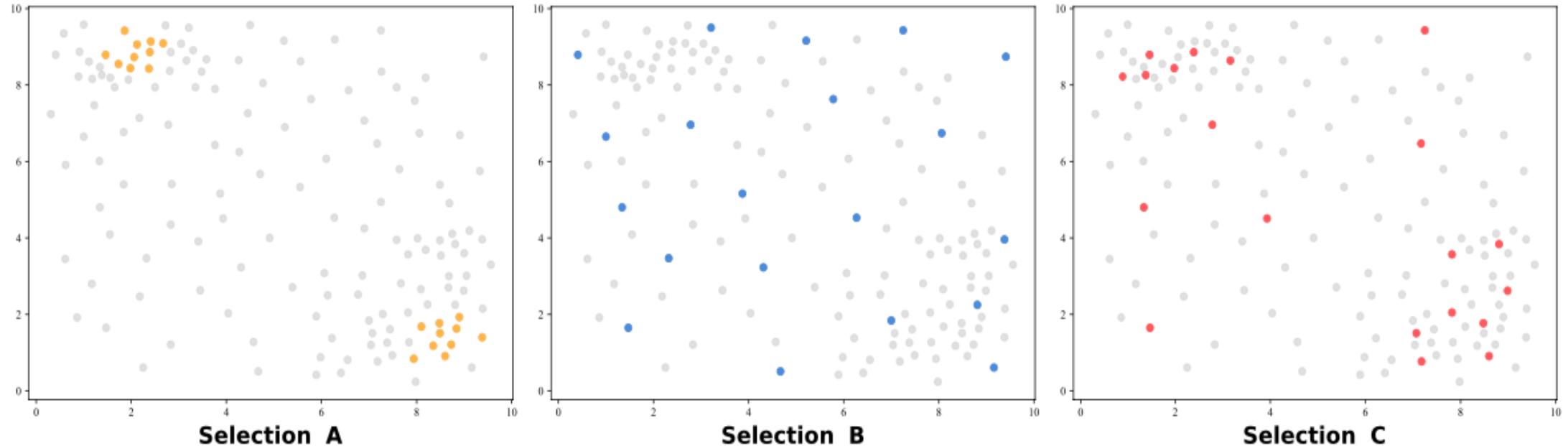
不同多样性评价指标与模型效果的相关性

Methods	MT-bench	AlpacaEval	Aggregated $\mathcal{P}$
Random	6.18	75.47	1.20
Repr Filter	6.17	72.57	1.05
QDIT	6.21	75.91	1.25
K-Center-Greedy	6.33	75.30	1.31
K-means	6.33	75.46	1.32
NovelSelect	<b>6.47</b>	<b>78.07</b>	<b>1.55</b>

Table 4: Comparisons of different diversity-oriented data selection strategies on IT performance.  $\mathcal{P}$  aggregates the performance based on z-scores (Eq. 5).

## 4. 多样性对SFT训练的影响

如何衡量多样性?



在二维空间中模拟数据选择：选择A模拟具有冗余的数据集，选择B优化样本之间的距离，而选择C同时考虑距离和密度

## 5. 推理能力的来源是强化学习吗？

相同的方法为什么Qwen可以，Llama不行？

“Countdown” 游戏作为测试，使用给定的数字和基本运算（加减乘除）达到指定的目标数

例如，给定25、30、3、4四个数字，目标是32，解决方案可以是： $(30 - 25 + 3) \times 4$

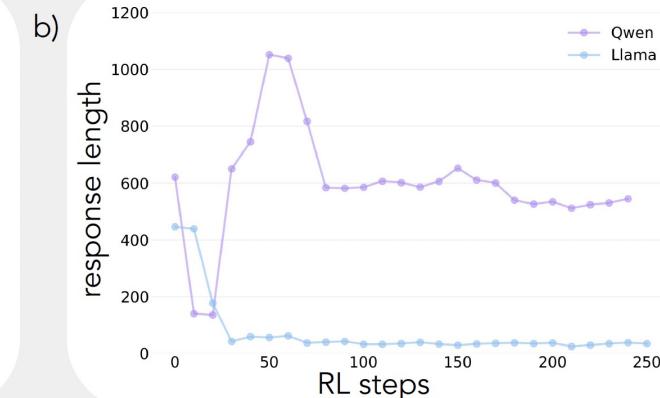
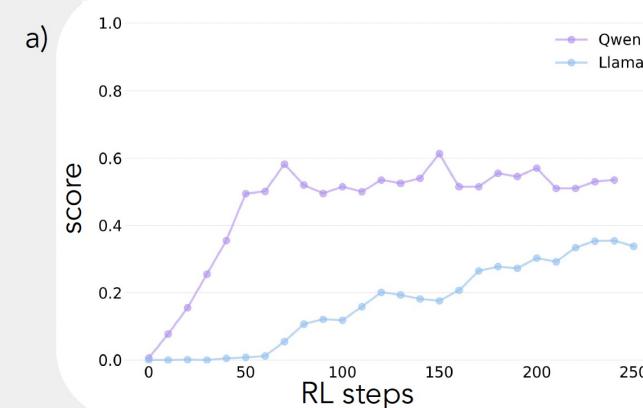
### A tale of two models: Qwen 2.5 3B and Llama 3.2 3B



Let's start with the sum of the largest two numbers and then subtract the smallest two:  $84 + 83 - 34 - 72$ . This gives us  $167 - 76$  which equals 91. That's not 39. Let's try another combination:  $84 + 83 - 72 - 34$ . This gives us  $167 - 106$  which equals 61. That's still not 39. Let's try  $84 + 72 - 83 - 34$ . This gives us  $156 - 117$  which equals 39. This is the correct equation.



84 is the difference between 108 and 34.  
<answer>  $(84 - 34) / 108$  </answer>



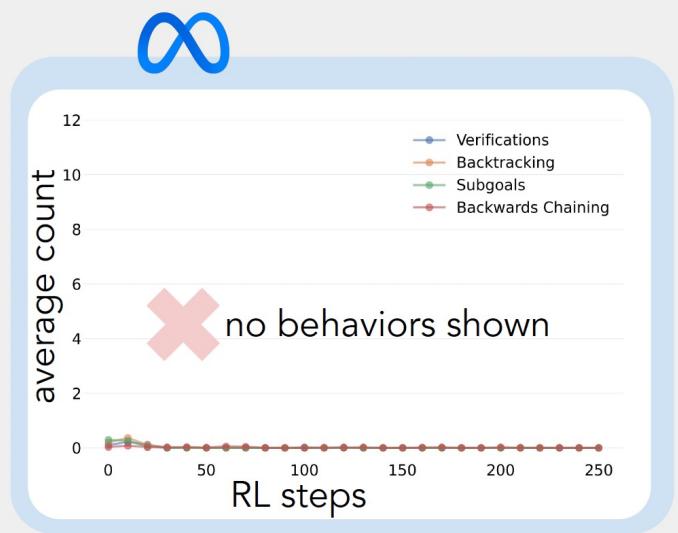
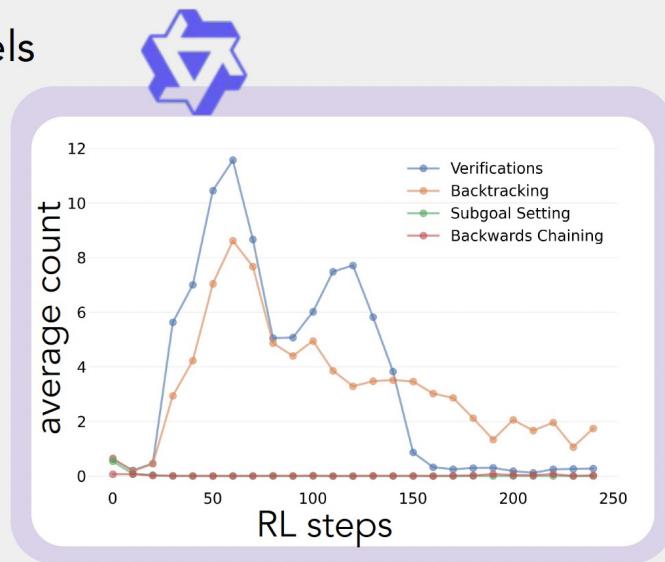
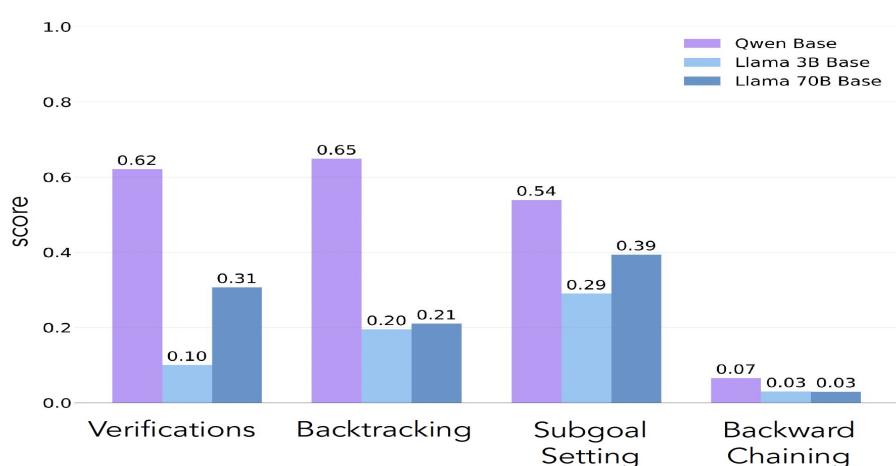
Qwen模型通过强化学习 (RL) 训练后表现出显著进步，而Llama模型却几乎停滞不前

## 5. 推理能力的来源是强化学习吗?

高效思考的四个关键行为：

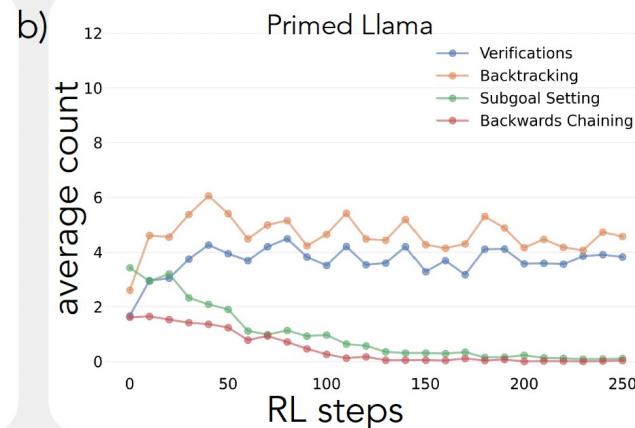
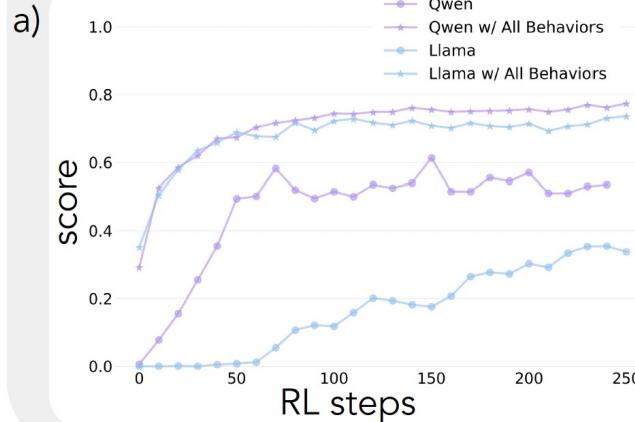
- (1) 验证能力 (Verification) : 系统性地检查中间结果和步骤是否正确, 如"让我们验证这个结果..."
- (2) 回溯能力 (Backtracking) : 当发现错误时, 能够明确地修改和调整方法, 如"这个方法行不通, 因为..."
- (3) 子目标设定 (Subgoal Setting) : 将复杂问题分解为可管理的步骤, 如"要解决这个问题, 我们首先需要..."
- (4) 逆向链式推理 (Backward Chaining) : 从目标反向推导解决方案, 如"要达到75的目标, 我们需要一个能被...整除的数"

A contrast in behaviors explored by the two models

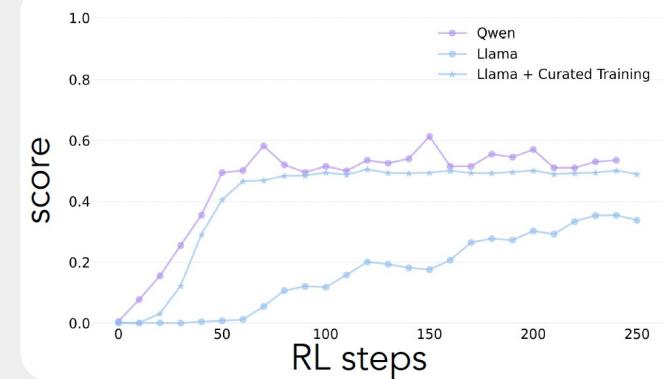


## 5. 推理能力的来源是强化学习吗?

Priming with behaviors reduces performance gap



We can curate a continued pre-training set so that Llama shows similar improvements to Qwen



(1) **行为启发 (Priming)**：使用包含这些认知行为的示例来启发Llama模型时，它在随后的强化学习中表现出显著改善，甚至能够匹配Qwen的表现轨迹。

(2) **错误示例也有效**：即使用带有错误答案但展示正确思考模式的示例来启发模型，也能取得类似的效果。这表明**认知行为的存在，而非正确答案的获取，才是自我提升的关键因素**。

(3) **预训练数据强化**：通过从OpenWebMath数据中筛选并强化这些认知行为的内容，成功地使Llama模型获得了与Qwen相当的自我提升能力。



## 6. Aha Moment 真的是“涌现”？吗？

**Question(a):** Find the least positive integer  $n$  such that  $1/\sin 45^\circ \sin 46^\circ + 1/\sin 47^\circ \sin 48^\circ + \dots + 1/\sin 133^\circ \sin 134^\circ = 1/\sin n^\circ$ .

**Response**

First, I notice that the left-hand side of the equation has the form  $\sum_{k=1}^n \frac{1}{\sin a_k \sin a_{k+1}}$ , where  $a_1 = 45^\circ, a_2 = 47^\circ, \dots, a_{90} = 133^\circ$ . I also notice that  $\sin a_{k+1} = \sin(a_k + 2^\circ)$ .

...

Simplifying, I get  $\sin(3a) = 3 \sin a - 4 \sin^3 a$ .

Aha! I can use this to get  $\sin(a + 2b) = \sin((a + b) + b) = 3 \sin(a + b) \dots$

This looks messy, but I notice that it has the form  $\sin a \cos b + \cos a \sin b$  again. This gives me an idea. Maybe I can use this identity repeatedly to simplify the left-hand side of the equation.

...

**Question(b):** An element is randomly chosen from among the first 15 rows ...

**Response**

In Pascal's Triangle, every row starts and ends with 1, ...

...

This can be calculated as: \*awkward silence\* Wait, I'm overthinking. Let's try again. The number of elements in the first  $n$  rows of Pascal's Triangle is  $(n + 1)!/[2! * (n - 1)!]$ .

...

**Question(c):** Denali and Nate work for a dog walking business and are paid ...

**Response**

To solve this problem, let's start by defining the variables...

...

Again, since  $x \neq 0$ , this equation doesn't give us a valid solution either. Let's check if we made an error. We should verify the problem conditions again. The correct approach is to ...

...

Figure 4: Cases showing that DeepSeek-V3-Base already exhibits “Aha moment” even before RL tuning.

### DeepSeek-V3-Base 就展现出了 “Aha moment”

## 6. Aha Moment 真的是“涌现”？吗？

Base model + Template	AIME24	AMC	MATH500	Minerva	OlympiadBench	Avg.
Qwen2.5-Math-1.5B						
(4-shot prompting)	0.0	20.0	50.4	12.1	15.9	19.7
R1 template	0.0	9.6	21.2	6.6	2.2	7.9
Qwen template	<b>20.0</b>	32.5	33.0	12.5	22.8	24.2
No template	16.7	<b>43.4</b>	<b>61.8</b>	<b>15.1</b>	<b>28.4</b>	<b>33.1</b>
Qwen2.5-Math-7B						
(4-shot prompting)	3.3	22.5	61.6	10.7	20.9	23.8
R1 template	0.0	0.0	0.0	0.0	0.1	0.0
Qwen template	<b>16.7</b>	38.6	50.6	9.9	16.6	26.5
No template	0.2	<b>45.8</b>	<b>69.0</b>	<b>21.3</b>	<b>34.7</b>	<b>38.2</b>

Qwen2.5-Math 模型可能在将问题-答案文本连接后进行预训练，这导致在不使用模板时达到最佳性能。

# 大模型能力来源总结

1. 一切能力都来源预训练 + 后训练
2. 预训练使得模型记住知识 + 学习到语义分布表示
3. SFT使得模型分布微小变化，预训练能力 “激活”
4. RL用于所有生成式任务，延迟奖励

# 几点感想

1. 大模型可以很快速的在很多任务上做到**70分**
2. 基于大模型在任何任务上完成**90分**都十分困难
3. 不要神话和拟人化大模型，依然是**统计机器学习**
4. 积极拥抱AI，**场景选择是关键**



谢谢！