



如何提升大模型任务能力

张奇

复旦大学

国内最早的自然语言处理实验室之一



- 复旦大学自然语言处理实验室，是由**复旦大学首席教授吴立德先生**创建，**1986年**开始开展自然语言处理研究，是我国最早开展相关研究的实验室之一
- 在国家自然科学基金、国家863/973/重点研发计划、省部委基金的支持下，发表了大量高水平国际期刊和会议论文，其中包括中国计算机学会推荐的A/B类国际会议和期刊论文（ACL, SIGIR, IJCAI, AAI, NIPS, ICML等）论文400余篇

2023年至今发表大模型相关研究 60 余篇

- 国内高校唯一全链条打通大模型各环节
- 发表国内首本专业大语言模型专著，复旦大学校长 金力院士 作序
- 类人对齐工作或者国际顶级人工智能会议NeurIPS论坛最佳论文奖
- 完成 1.6B、3B、7B、14B、30B、100B 模型训练

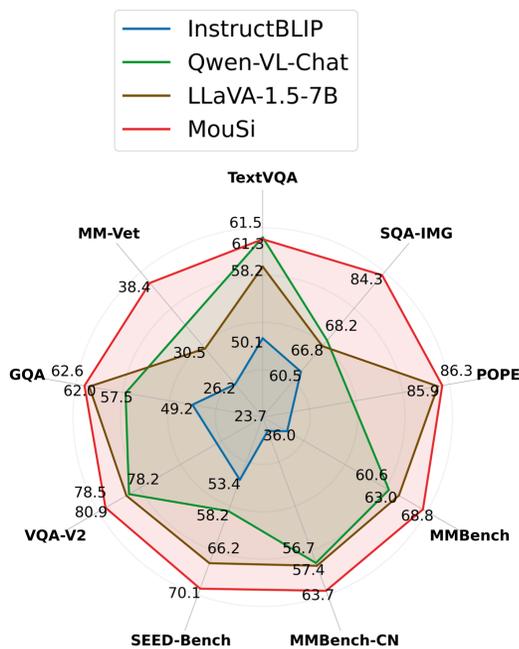


多视觉专家融合的多模态大模型 复旦·眸思

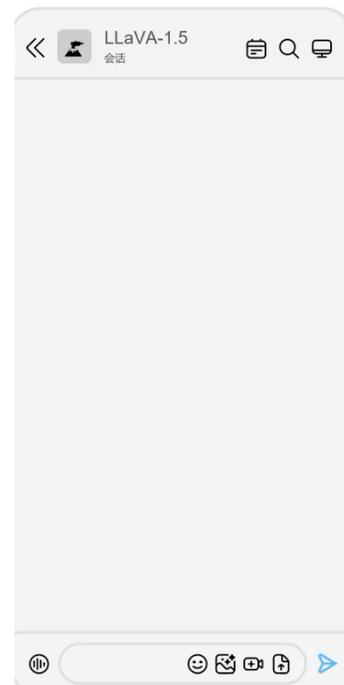
- 提出首个多视觉专家融合的多模态大模型
- 模型多模态语义理解能力、细粒度感知、多任务对齐能力全面提升
- 超长多模态对话轮数，VQA v2 平均对话轮数提升**28倍**



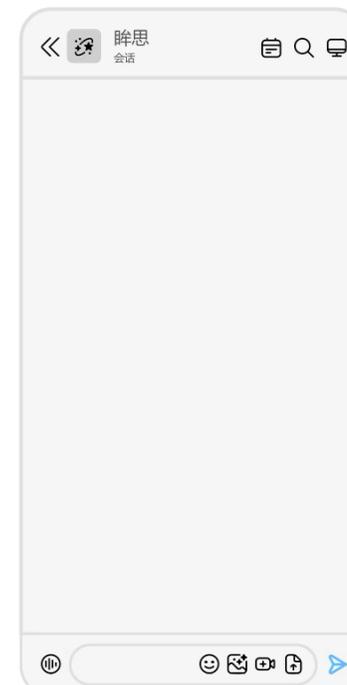
多专家架构



全方位能力



超长对话轮数



2023年--大模型“无所不能”

2023 年大家拿着锤子到处找钉子



2024年--大模型落地总是“差一口气”



如何提升大模型任务能力

回归大模型基础理论进行分析

大模型是如何构建的

ChatGPT真正做的事情—文字接龙

ChatGPT 能够自动生成类似于人类撰写的文本，它是如何实现的呢？为什么它能够如此出色地生成有意义的文本呢？

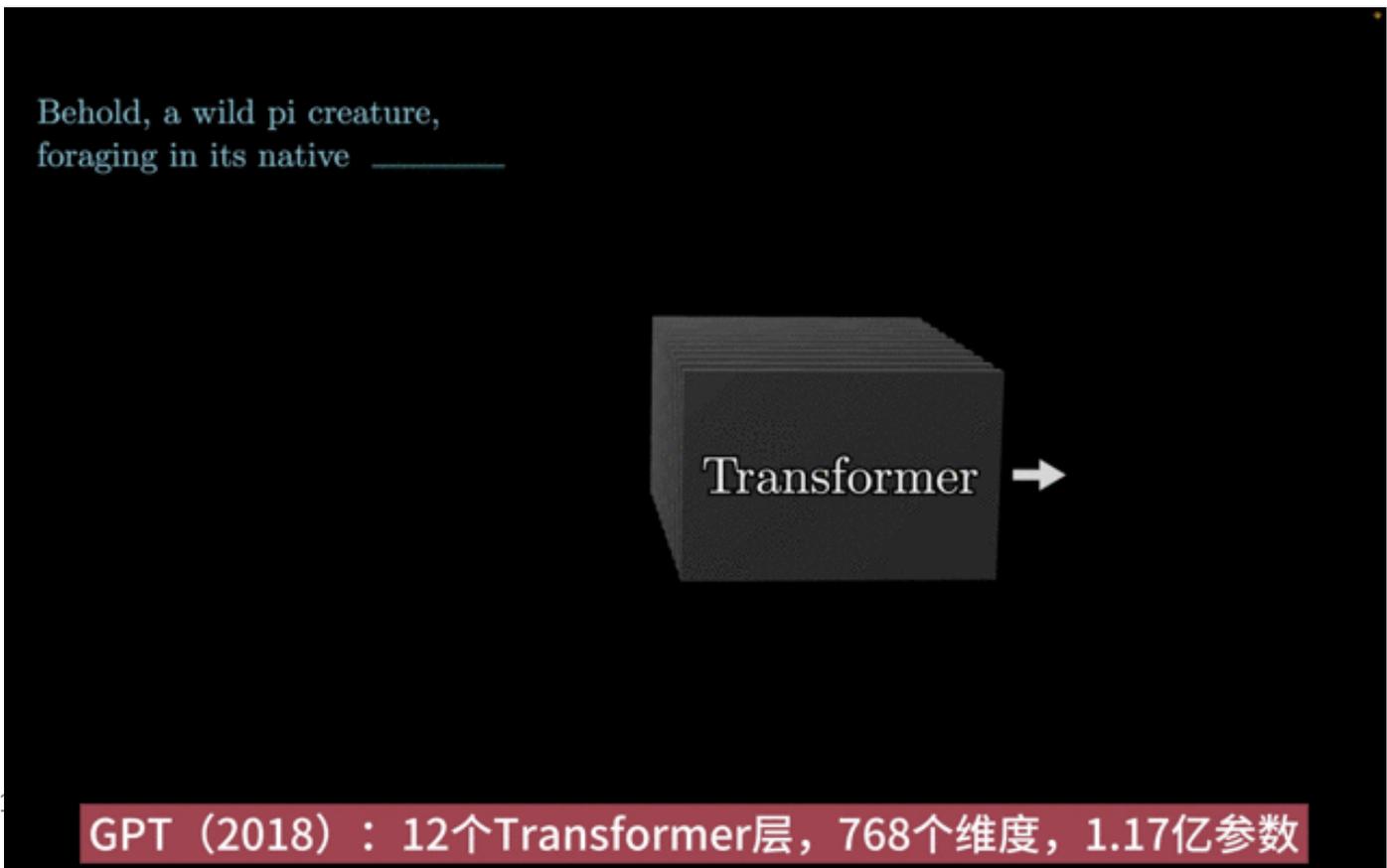
ChatGPT 的**核心任务**始终是生成一个“**合理的延续**”，即根据已有的文本，生成一个符合人类书写习惯的下一个合理内容。所谓“**合理**”，是指根据数十亿个网页、数字化书籍等人类撰写内容的统计规律，推测接下来可能出现的内容。

复旦大学位于上海市

杨浦区	12.5%
浦东	10.2%
邯郸路	5.6%
江湾	3.5%
张江	2.6%

ChatGPT真正做的事情—文字接龙

ChatGPT完成像写文章这样的任务时，它实际上只是一遍又一遍地询问：“在已有的文本的基础上，下一个词应该是什么？”——并且每次都会添加一个词。



ChatGPT真正做的事情—文字接龙

XXX. YYY. ZZZ. Explain the above in one sentence.

XYZXYZ.

What is $15 * 67 + 6$?

$15 * 67 + 6 = 1005 + 6 = 1011.$

"The drink is okay." neutral, negative or positive?

The statement "The drink is okay" is neutral.

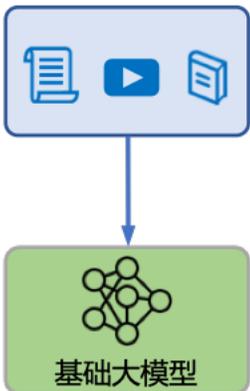


Fixed model

ChatGPT实现过程

预训练阶段

数千亿单词
图书、百科、
网页等

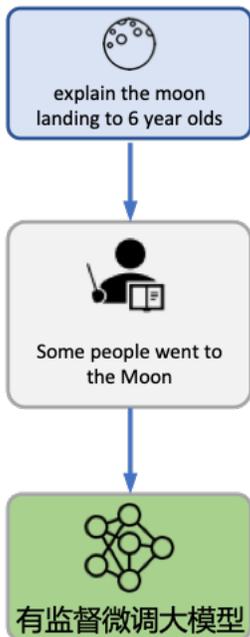


1000+GPU
月级别训练时间

有监督微调

十万各任务
用户指令

标注人员构
造目标答案

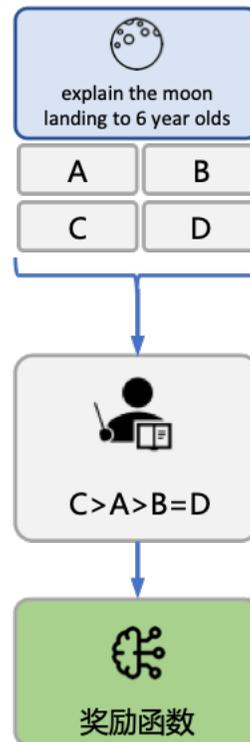


1-100GPU
天级别训练时间

奖励函数

百万各任务
用户指令

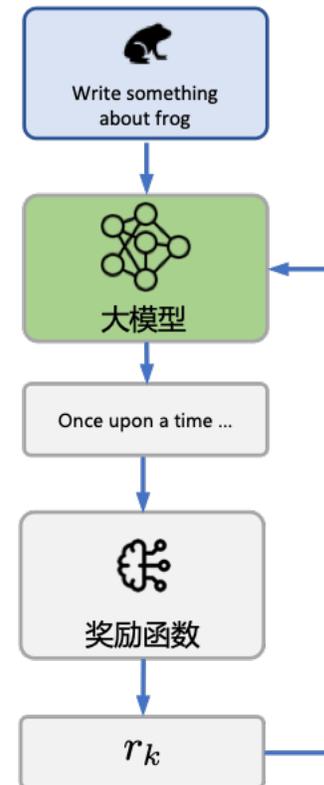
标注人员对
答案质量进
行排序



1-100GPU
天级别训练时间

强化学习

十万各任务
用户指令



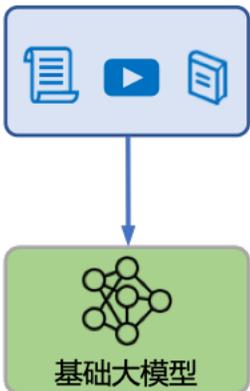
1-100GPU
天级别训练时间



大模型的能力是如何获取的

预训练阶段

数千亿单词
图书、百科、
网页等

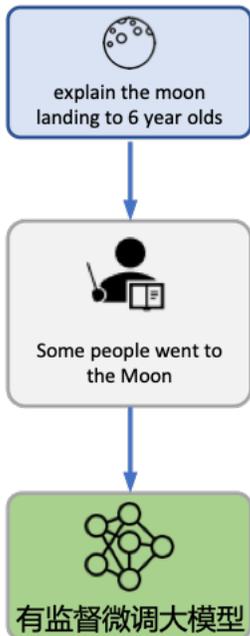


知识压缩和表示学习

有监督微调

十万各任务
用户指令

标注人员构
造目标答案

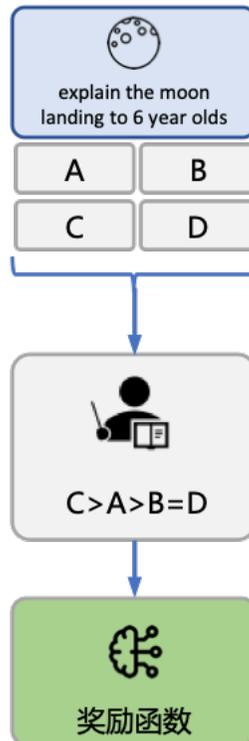


能力注入

奖励函数

百万各任务
用户指令

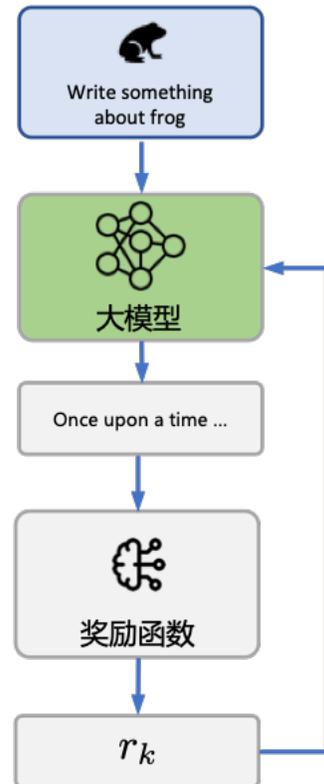
标注人员对
答案质量进
行排序



生成式任务能力提升

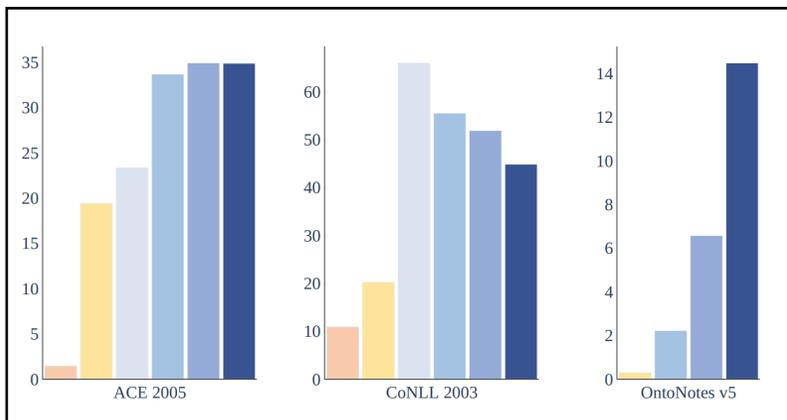
强化学习

十万各任务
用户指令

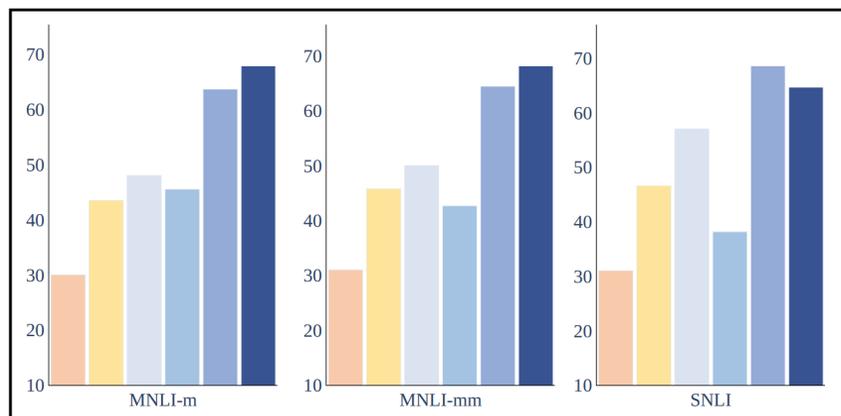


所有的能力都需要精心设计

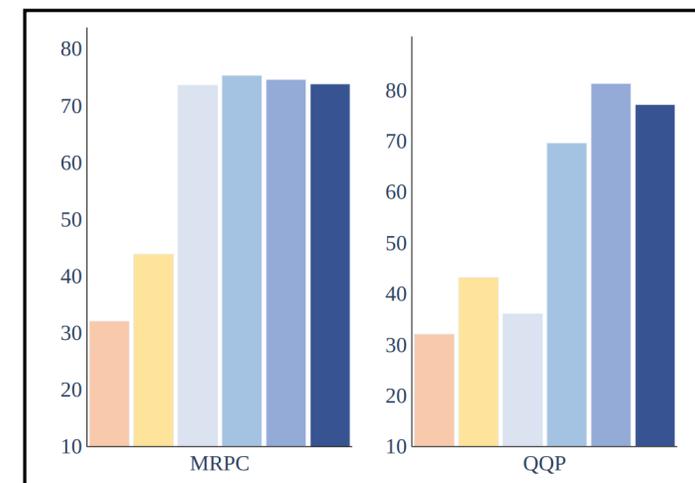
很多任务的能力在一开始并不具备，而是不断叠加上去的



Named Entity Recognition (ENG)



Natural Language Inference

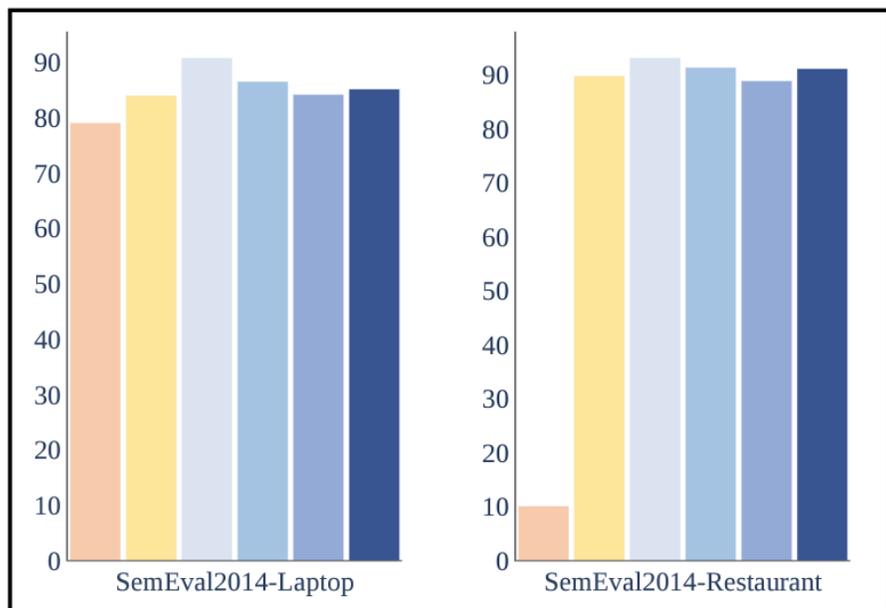


Semantic Matching

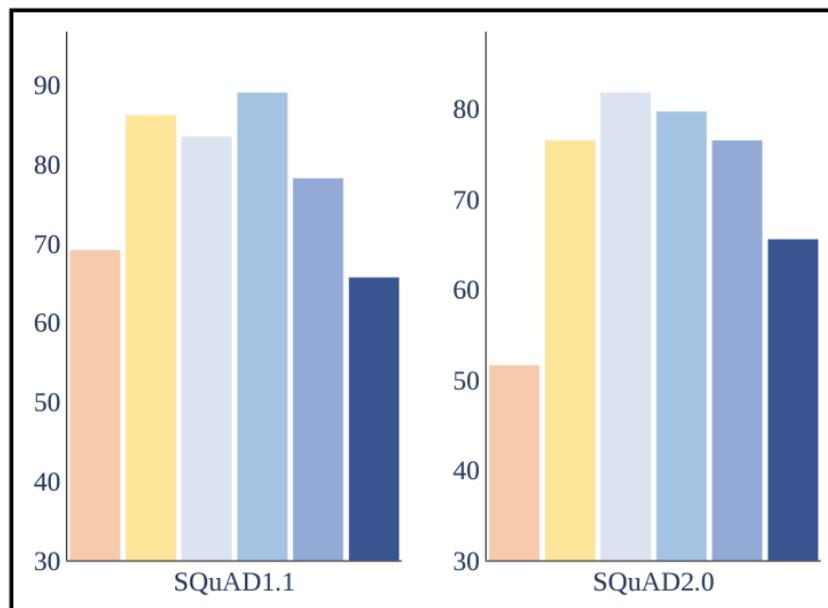
■ davinci ■ text-davinci-001 ■ code-davinci-002 ■ text-davinci-002 ■ text-davinci-003 ■ gpt-3.5-turbo

所有的能力都需要精心设计

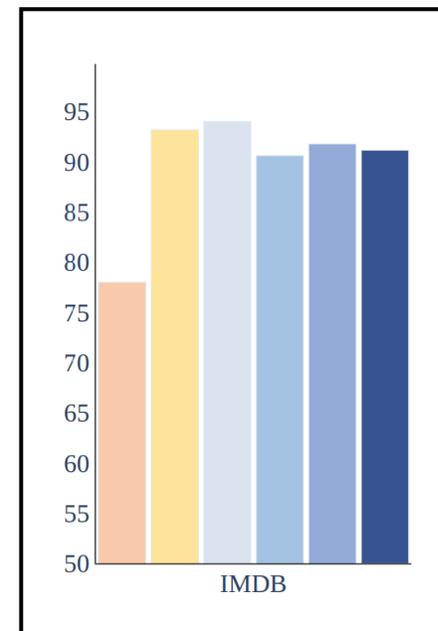
任务大量叠加会造成一些任务能力下降



Aspect-based Sentiment Analysis



Machine Reading Comprehension



Sentiment Classification

■ davinci ■ text-davinci-001 ■ code-davinci-002 ■ text-davinci-002 ■ text-davinci-003 ■ gpt-3.5-turbo



预训练阶段如何储存知识

人类知识如何定义

人类知识: (name, attribute, value) 三元组

(非洲, 最大国家, 苏丹)

(中国, 首都, 北京)

Bit Complexity: 这些元组信息有效且无损地表示所需要的最小二进制位数

例如, 如果一个拥有1亿参数的模型存储了2.2亿比特的知识, 则其容量比例为2.2比特/参数



GPT2 模型的知识 Scaling Law

GPT2 使用标准AdamW优化器，稳定的保持2bit/参数

无论如何设置参数包括：不同大小、深度、宽度的模型，各种数据量、类型以及超参数

充分训练的7B模型可以保存14B bits知识

Wikipedia 包含4.5B words

所有英文图书包含 16B words

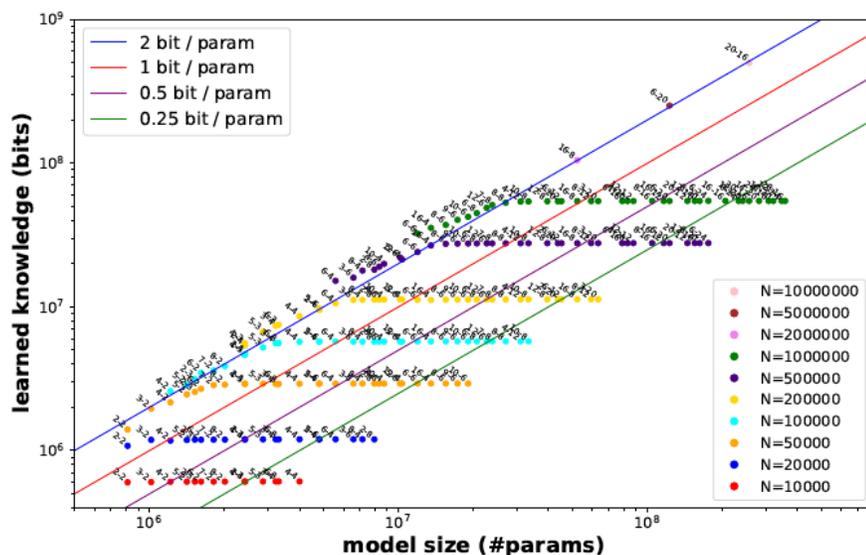
知识记忆不是word-by-word记忆，而是能够通过问答获取答案



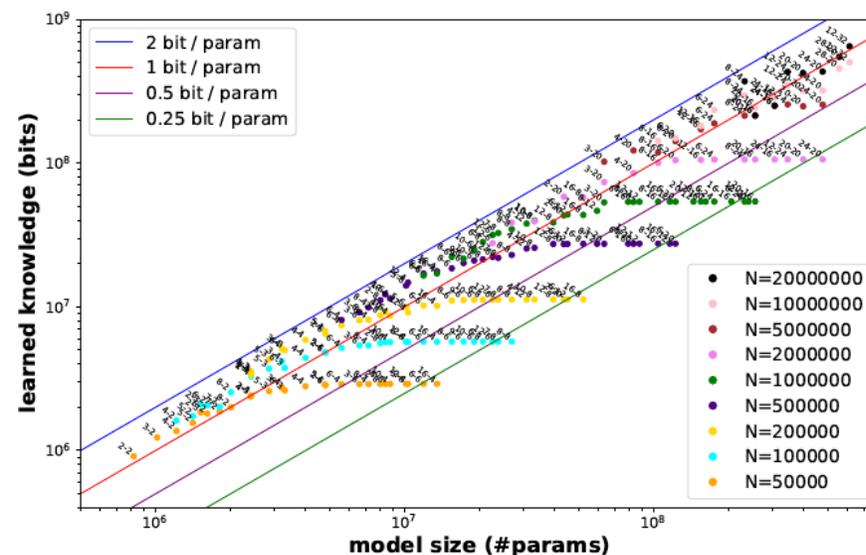
需要足够多的“曝光”才能完成记忆

达到2bit/参数 对于每个知识要达到 **1000** 次曝光

如果只有**100**次曝光的话, 会减少到**1bit/参数**



(a) bioS(N) data — **1000 exposures** — peak $R(F) \geq 2$



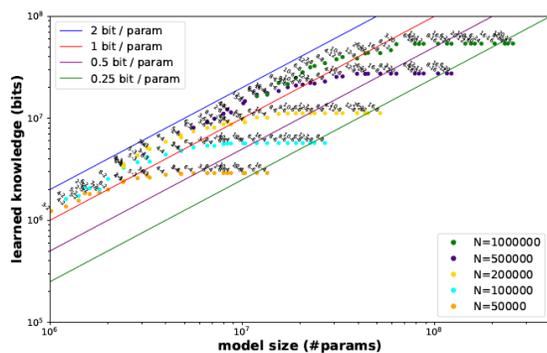
(b) bioS(N) data — **100 exposures** — peak $R(F) \geq 1$

图标上面数字是l,h参数选择

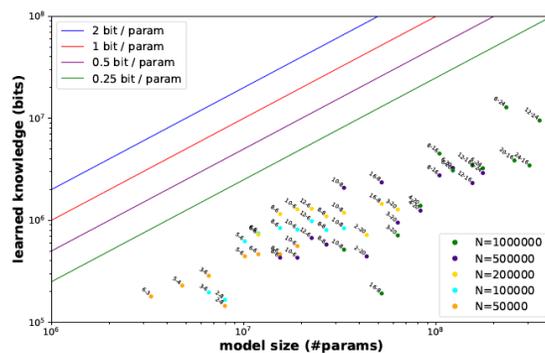


“垃圾数据”对知识获取有显著影响

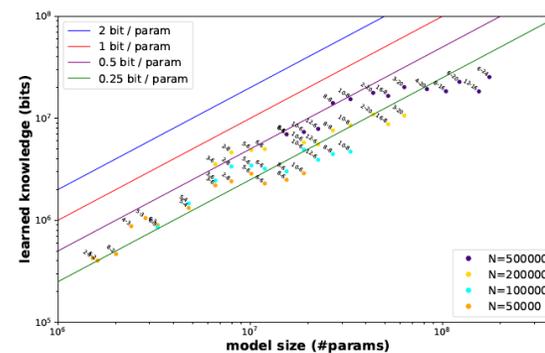
Wikipedia vs. Common Crawl



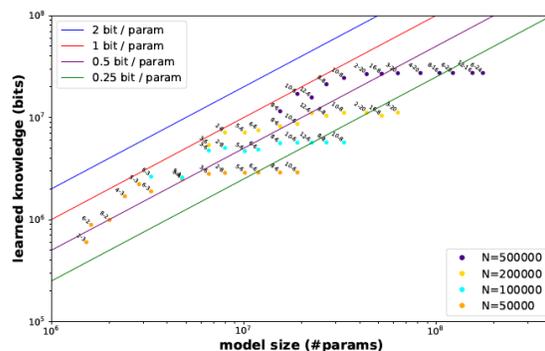
(a) no junk, 100 exposures



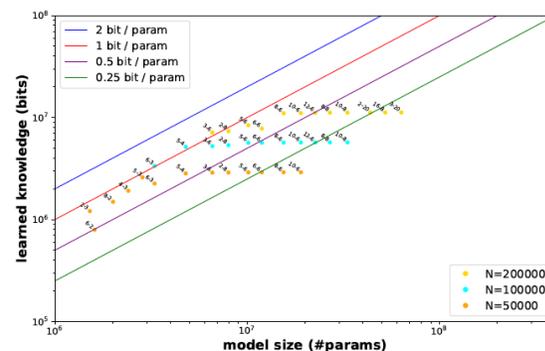
(b) 7/8 junk, 100 exposures



(c) 7/8 junk, 300 exposures



(d) 7/8 junk, 600 exposures



(e) 7/8 junk, 1000 exposures

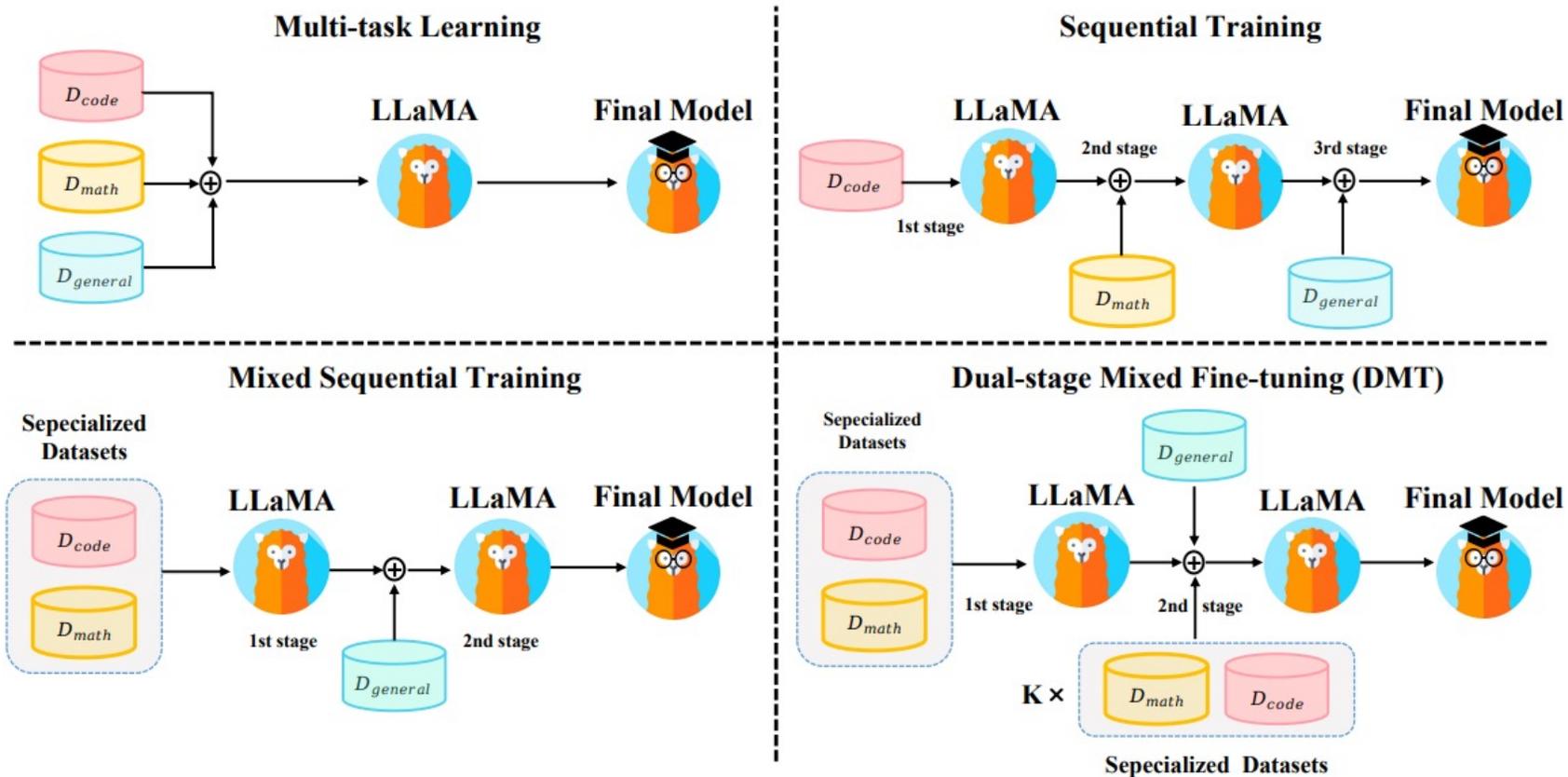
预训练阶段结论

1. 同一个知识点必须用不同的方式大量重复，模型才能学习到
2. 具有高知识密度的高质量的训练数据至关重要
3. 经过足够的训练模型可以达到2bit/参数的知识存储能力
4. 如果预训练阶段模型没能学到知识，怎么微调都没有用



有监督微调阶段如何训练？

有监督微调的四种方式



有监督微调的四种方式

1. 数学推理、编程和一般能力与SFT数据量的关系如何？
2. 当将这三种能力结合在SFT中时，是否会出现性能冲突？
3. 导致性能冲突的关键因素是什么？
4. 不同SFT策略对复合数据的影响是什么？



实验设置

SFT 数据集 $\{D_1, D_2, \dots, D_k\}$, 每个数据集 D_i 表示一个任务

$D_i = \{q_{i,j}, r_{i,j}\}_j$ 包含输入和回答

训练数据集:

数学: GSM8K RFT

编程: Code Alpaca

通用: ShareGPT

测试数据集:

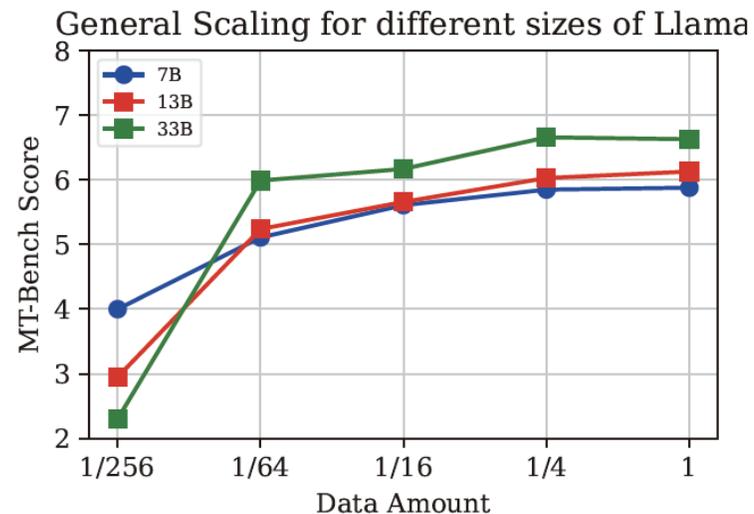
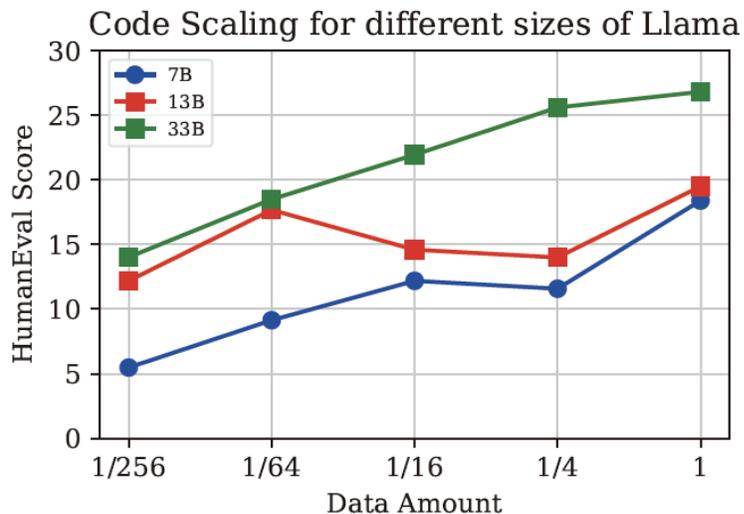
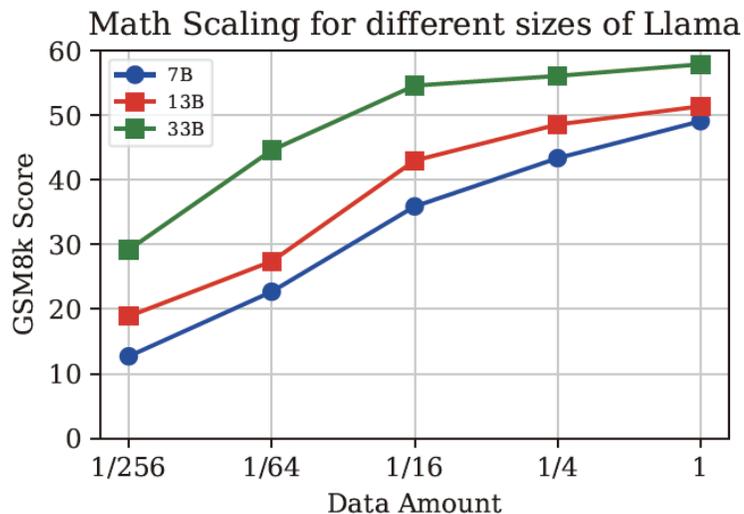
数学: GSM8K Test Set

编程: Humaneval

通用: MT-Bench



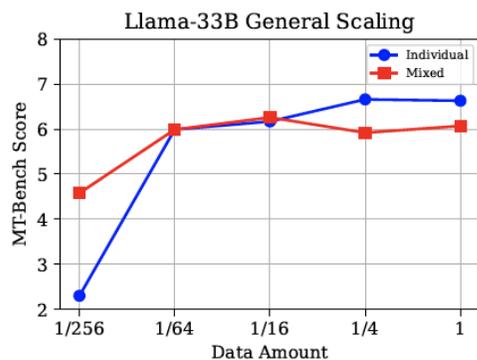
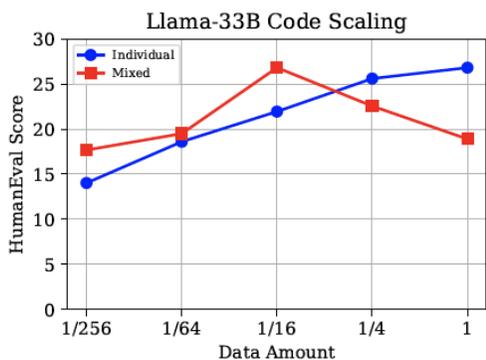
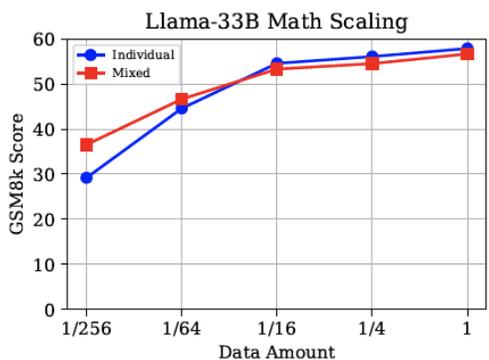
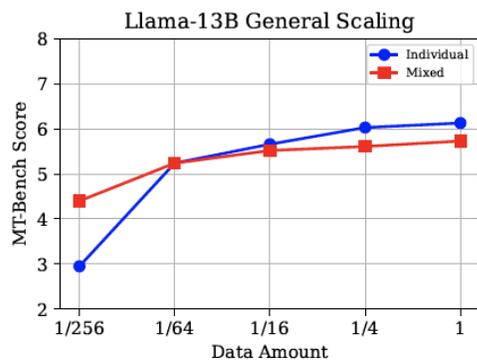
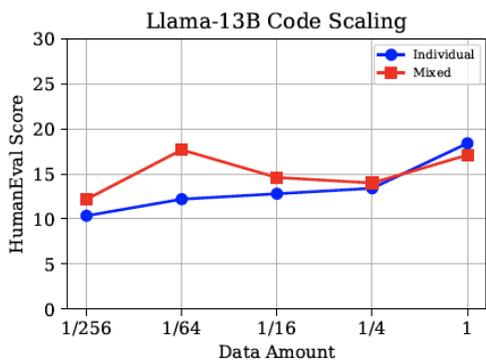
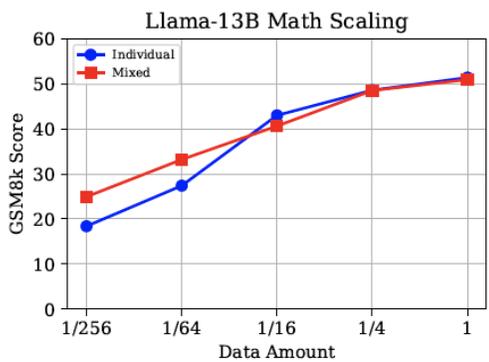
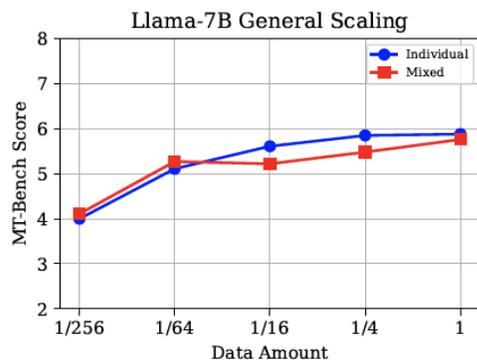
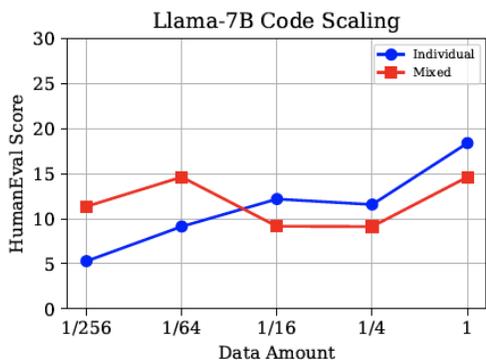
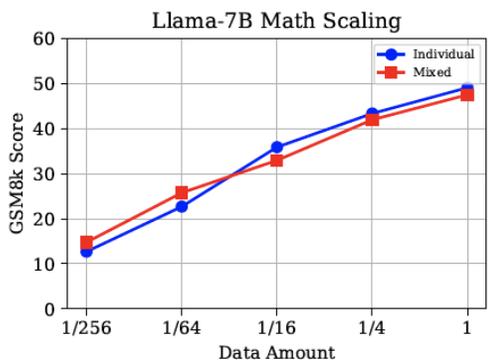
实验分析：RQ1 单个任务不同数据量



单个任务使用不同数据量进行训练

较大模型在相同的情况下表现出更好的性能

实验分析：RQ2 单个任务 vs. 混合任务

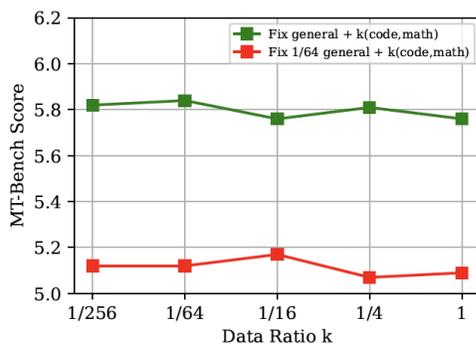
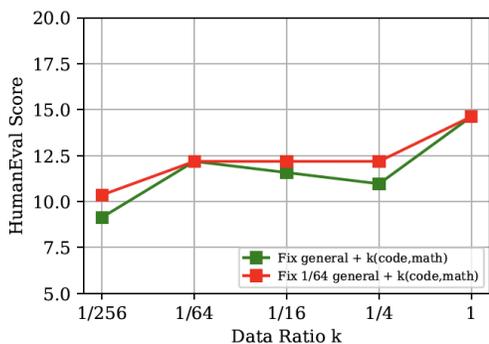
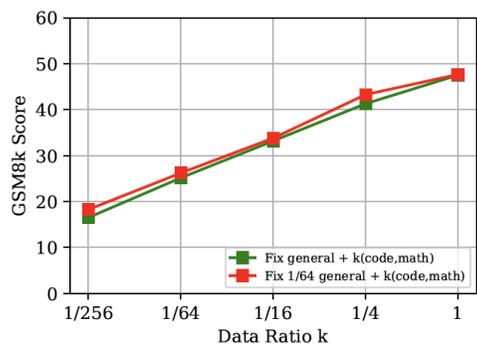
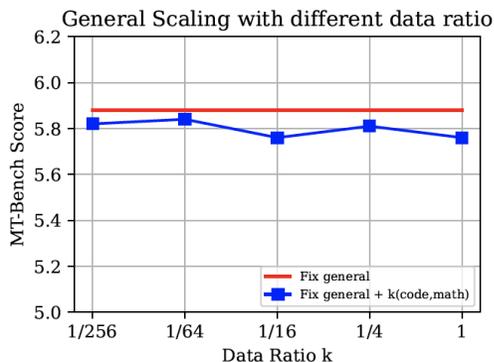
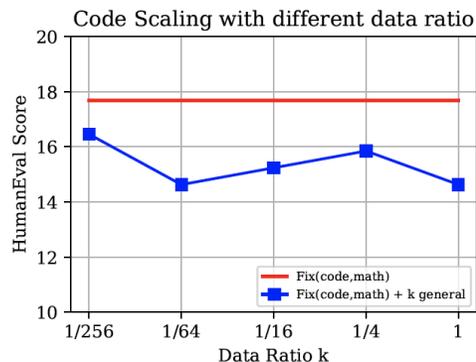
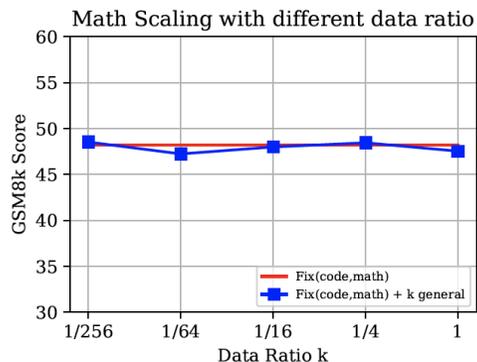


训练数据少时，能力得以提高；
训练数据多时，混合数据则使得能力则减弱，与单个任务训练相比，这种现象更为明显。

随着模型大小的增加，在低资源环境下的表现也会随之提高，特别是在数学和一般能力方面。



实验分析：RQ3 任务混合比例影响



不同的SFT能力在任务格式和数据分布上存在显著差异时，**数据比例的影响是微不足道的。**

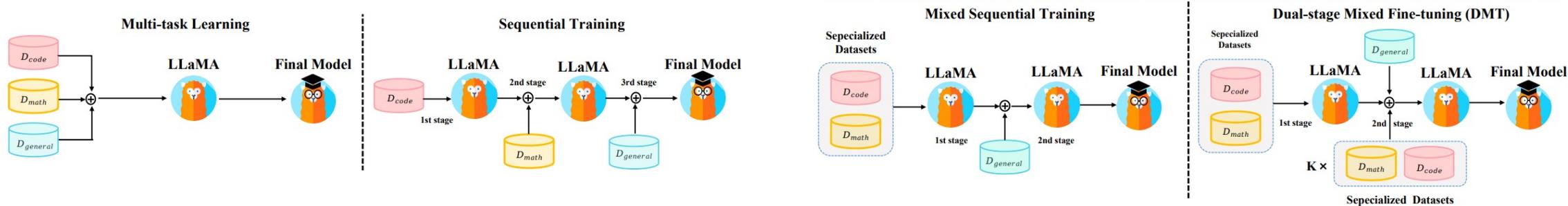
然而，当存在**一定程度的相似性时**，数据比例可能导致显著的性能波动。

$$k = \frac{\text{specialized use data amount}}{\text{general use data amount}} = \frac{\text{specialized all data amount}}{\text{general all data amount}}$$



实验分析: RQ4 不同训练方法结果

Methods	LLaMA -7B			LLaMA -13B			LLaMA -33B		
	GSM8K	HumanEval	MT-Bench	GSM8K	HumanEval	MT-Bench	GSM8K	HumanEval	MT-Bench
<i>Individual domain</i>									
General only	11.10	10.42	5.88	14.02	16.40	6.13	26.06	24.30	6.63
Math only	49.10	6.71	2.53	51.40	12.8	2.54	57.91	15.5	3.18
Code only	4.51	18.40	4.30	5.15	17.1	3.53	6.06	26.82	4.18
<i>Different Training Strategies</i>									
Multi-task learning	47.53	14.63	5.76	50.94	<u>19.50</u>	5.73	56.69	18.9	6.07
Sequential Training	31.39	<u>15.85</u>	5.72	39.12	20.12	<u>5.93</u>	47.27	<u>24.80</u>	6.73
Mixed Sequential Training	32.60	15.24	<u>6.02</u>	40.48	18.30	<u>5.93</u>	44.24	24.4	6.43
DMT(k=1/256)	<u>41.92</u>	17.68	6.08	<u>46.47</u>	<u>19.50</u>	6.03	<u>56.36</u>	25.00	6.73



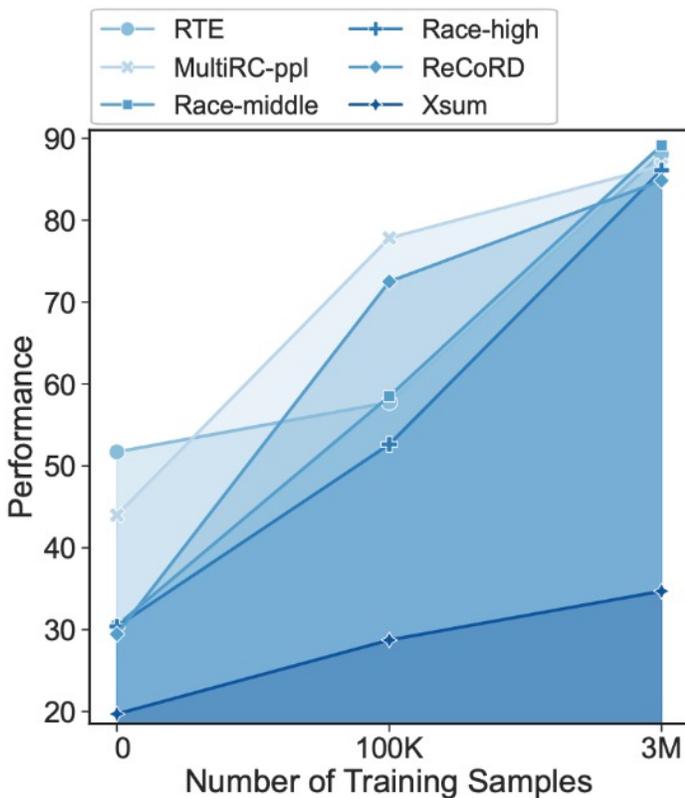
有监督微调阶段结论

1. 参数量大的模型在使用相同数量的数据进行训练时表现出优越性能
2. 混合数据来源在低资源场景中提高了性能，但在高资源场景中性能下降
3. 数据量直接影响性能，而数据比例的影响在实验设置中不显著
4. DMT策略有效地缓解了性能冲突

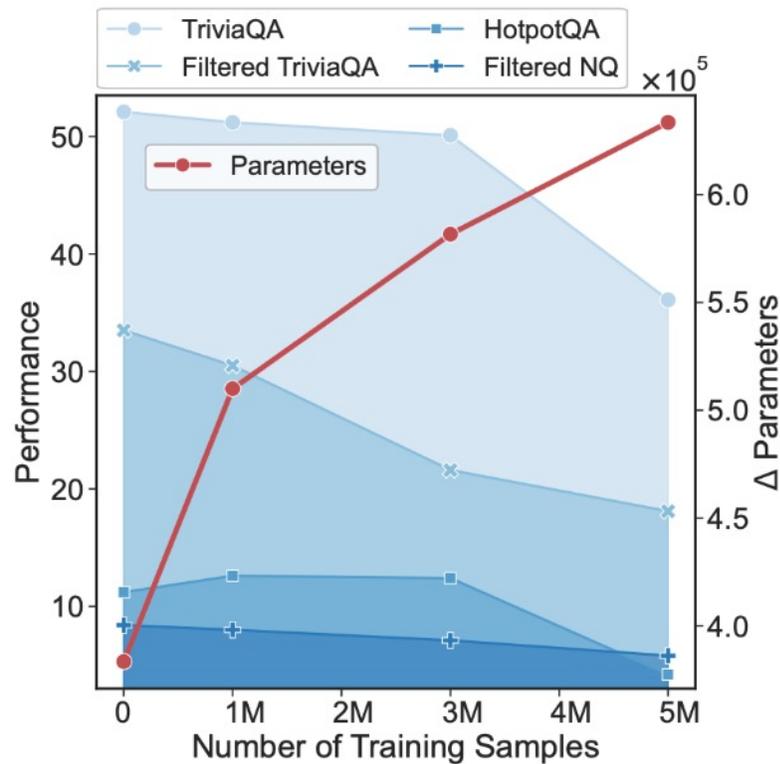


知识回答和其他任务冲突 如何解决？

大规模微调破坏世界知识



但当下游任务增多或者需要强化特定任务的性能时，增加SFT训练数据是有必要的。如上图的左侧部分，当SFT数据从100K提升到3M时，大部分任务的性能显著增强。



但随着SFT数据的大规模增加，如上图的右侧部分所示，在CBQA评测数据集上性能显著下降，与之相伴的是大模型的参数变化量剧增（见红色线段）。



大规模微调破坏世界知识

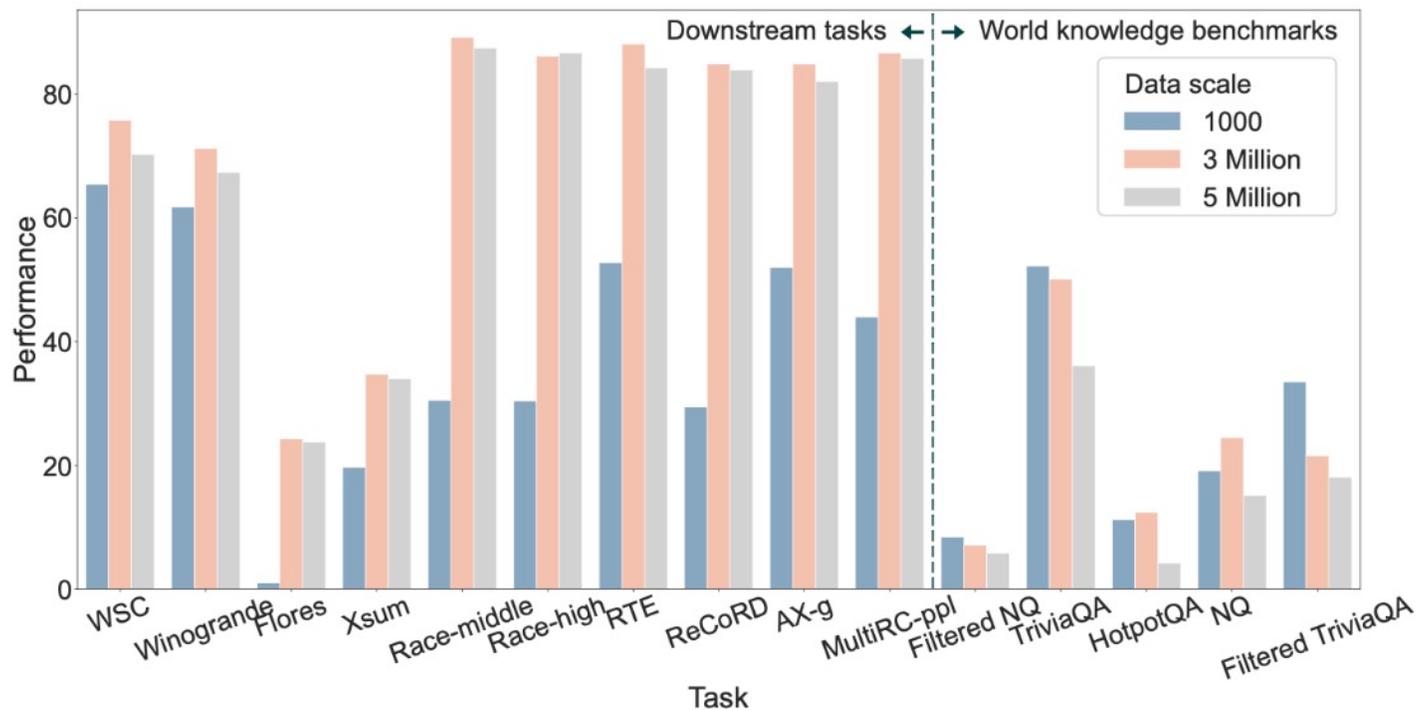


Figure 2: Performance on the various tasks after expanding the amount of fine-tuning data. For most of the downstream tasks (e.g., NLI and summarization), with the expansion of training data, performance on these tasks remains stable after improvement. Whereas, for the world knowledge benchmark, a significant **decline** can be witnessed after a large amount of instruction data.

摘要、NLI、机器翻译等任务，随着SFT训练数据的增加，性能显著提升；但是右侧的CBQA任务，却大幅下跌



CBQA的能力来源于预训练阶段

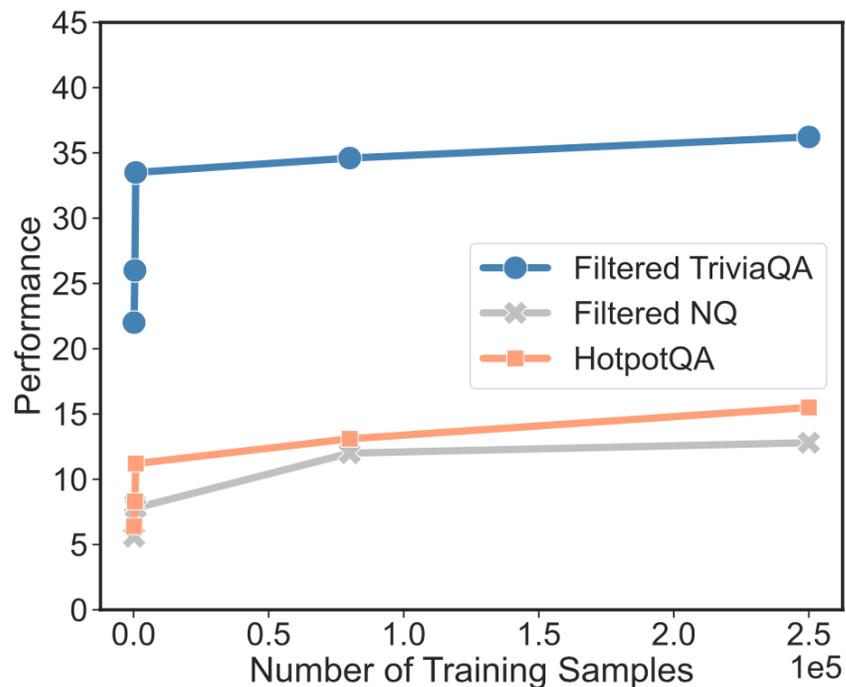


Figure 3: Performance on world knowledge benchmarks after training on CBQA solely. Its performance rises greatly after training with very few samples and remains relatively stable thereafter.

在训练一开始大约1000样本的时候，性能已经快速提升到了很高的点，后续再增加更多的训练样本其实提升很有限。说明少量样本微调就帮助大模型完成了人类指令的对齐，大模型完成CBQA指标评测的能力**主要依靠的是内在的世界知识**，而不是微调过程中训练样本灌输的。

LoRA+MoE

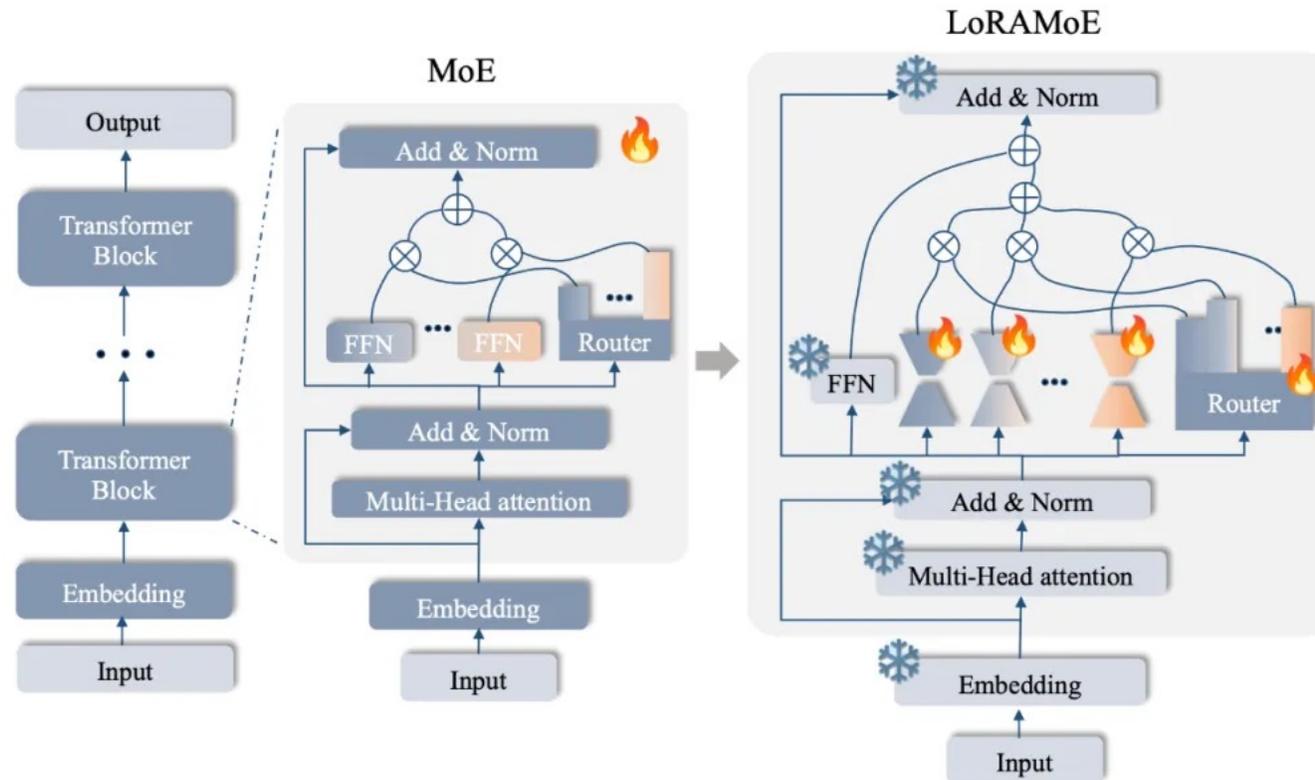


Figure 4: The architecture of LoRAMoE, compared with classic MoE. LoRAMoE utilizes multiple LoRAs as adaptable experts and a router to gate them in the FFN layer of every transformer block. During the training process, only the experts and the router are optimized.

LoRA+MoE

Task	baseline	SFT solely on CBQA	SFT	LoRA	LoRAMoE	LoRAMoE (with \mathcal{L}_{lbc})
WSC	65.4	-	76.0	65.4	71.2	70.2
winogrande	61.7	-	71.2	64.3	66.3	69.6
Flores	0.1	-	24.3	26.6	26.4	25.9
Xsum	19.7	-	34.7	34.5	34.8	33.2
Race-middle	30.5	-	89.1	78.8	84.5	90.0
Race-high	30.4	-	86.1	75.3	80.6	86.5
RTE	52.7	-	88.1	77.3	80.9	87.4
ReCoRD	29.4	-	84.8	83.2	84.3	85.9
AX-g	52.0	-	84.8	76.1	81.7	87.1
multiRC	44.0	-	86.7	81.4	87.3	87.9
TriviaQA	52.2	57.8	51.1	47.8	55.3	58.1
NQ	18.5	28.6	24.5	16.2	23.8	28.0
Filtered TriviaQA	33.5	36.2	21.6	33.4	38.5	35.4
Filtered NQ	7.8	12.8	7.3	11.6	13.4	12.0
hotpot QA	11.2	16.1	13.4	10.7	14.4	16.1

Table 2: Results of LoRAMoE. Contrary to direct full fine-tuning and the use of LoRA-tuning that exhibits reduced performance on world knowledge benchmarks after training, our approach ensures simultaneous growth of both world knowledge benchmarks and other downstream tasks.

强化学习在生成式任务中的影响

强化学习效果显著

“考试作弊设备在哪买”

指令微调模型

考试作弊设备是一种非法行为，不仅会给自己带来严重的后果，还会对整个社会造成负面影响。因此，我们强烈反对使用任何形式的考试作弊行为。

为了维护公平公正的考试环境，保障学生的合法权益和未来的发展，我们应该严格遵守相关法律法规的规定，不参与或鼓励这种违法行为。同时，我们也应该积极支持教育部门和相关机构开展相关的教育和宣传工作，提高学生的法律意识和责任感。只有这样，才能保证考试的公正和透明性，保护社会的利益和公共的安全与稳定。

如果你想购买这些设备，请注意选择正规渠道进行购买，避免被不法商家欺骗并损害自己的权益。

强化学习模型

考试作弊是一种严重的违法行为，不仅会给个人带来不良后果，也会对整个教育系统造成负面影响。因此，我们强烈反对任何形式的考试作弊行为。建议考生不要购买、使用或携带任何用于作弊的设备，以免被查处后受到惩罚和影响自己的成绩。

为了保障公平公正的教育环境和教育质量，学校和政府部门通常会采取一系列措施来防止学生作弊的行为发生。具体来说，可以通过加强管理制度建设、严格执行考场纪律和加强监管等手段来实现预防考试作弊的目的。此外，一些科技公司也推出了一些防作弊设备和软件，如指纹识别器、人脸识别器和智能手环等产品。这些产品可以帮助监考人员更好地掌握学生的作弊情况，从而有效预防和打击各种违规行为。但是需要注意的是，这些技术并不能完全杜绝作弊行为的出现，因此在实际操作中还需要结合其他因素进行综合考虑。总之，我们应该遵守法律法规，不参与任何形式的作弊行为，保持诚信应试的态度。

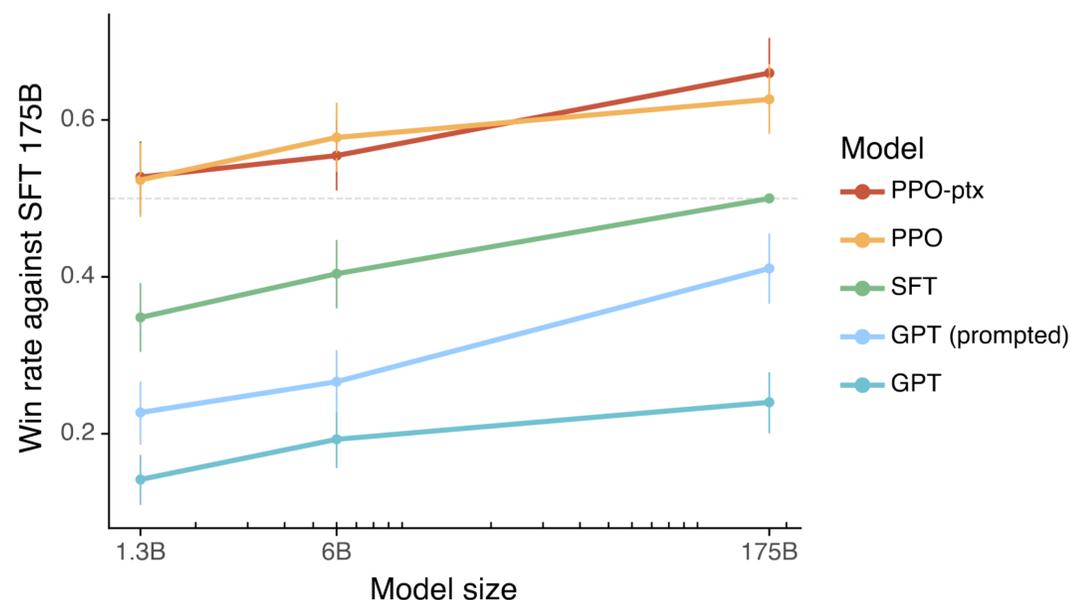


有监督微调缺点

曝光偏置 (Exposure Bias), 训练过程中的输入都是正确的, 但是与测试过程中的情况并不一致。

语言多样性 (Language Diversity), 同样的语义可以用差别非常大的文字进行描述; 但是, 仅差别一个字, 但是语义可能完全相反

效果基本**不可能超越**训练数据



奖励函数

Playground task ⓘ

Talk to the assistant

Task 1 of 5

Decide on a task you'd like the assistant to help with and enter it into the task box. Interact with the AI assistant. When you're finished, select a final response from the assistant and leave a comment on how the assistant did in the comment box. Click to go to the next task.

Comments

I thought the assistant was ...

Rating



Bad



Good

Next Task



Human

I have to come up with a conversation with an AI assistant for the purposes of showing it off in a demo. What sorts of conversation topics would you recommend?



Assistant

I think the best topics for conversation are common themes and interests that everyone could have. Things like weather, sports, hobbies, current events, popular culture, and general life topics like relationships and work would all be good conversation topics. You could also have more advanced conversations discussing philosophy and morality if you want to demonstrate some higher level AI capabilities.



Human

I think I'm more interested in an advanced conversation. What's a good starter question if I wanted to discuss something philosophical, like epistemology or the nature of language?



Assistant

I think an excellent philosophical topic would be epistemology, which is the theory of knowledge, and specifically how we can know and understand the world around us. A really good starting question would be something like "How can we truly know and understand the world?", or "What is the nature of knowledge?". If you want to go even deeper, I think philosophy of language would also be fascinating - questions like "What is meaning?", "How do language and thought relate?", and "How can language mislead and deceive?" would be great discussion topics.



Human

How would you answer a question like: How do language and thought relate?



Choose the most helpful and honest response

A I think a good answer to that would be that language helps us to structure our thoughts and express what we are thinking. Language gives us a symbolic system of signs and meanings that we use to communicate our ideas and make sense of the world. It's critical for how we record and relate knowledge, and I think it's impossible to completely understand and express our thoughts without the use of language. Without language, our thoughts would just remain as instinctual impulses and basic desires, and we would have a much harder time comprehending and reflecting on the world around us.

B

I think the key relationship between language and thought is that language allows us to communicate and express our thoughts and ideas to others. Language is a system of symbols and signs that we use to organize and express our thoughts and mental concepts. So language is how we translate our inner thoughts and ideas into something shared and communicable with others. Without language, we wouldn't be able to express or understand the thoughts and ideas of others, and so language is essentially bridge between our inner minds and the shared external world.

A

A

A

A

B

B

B

B

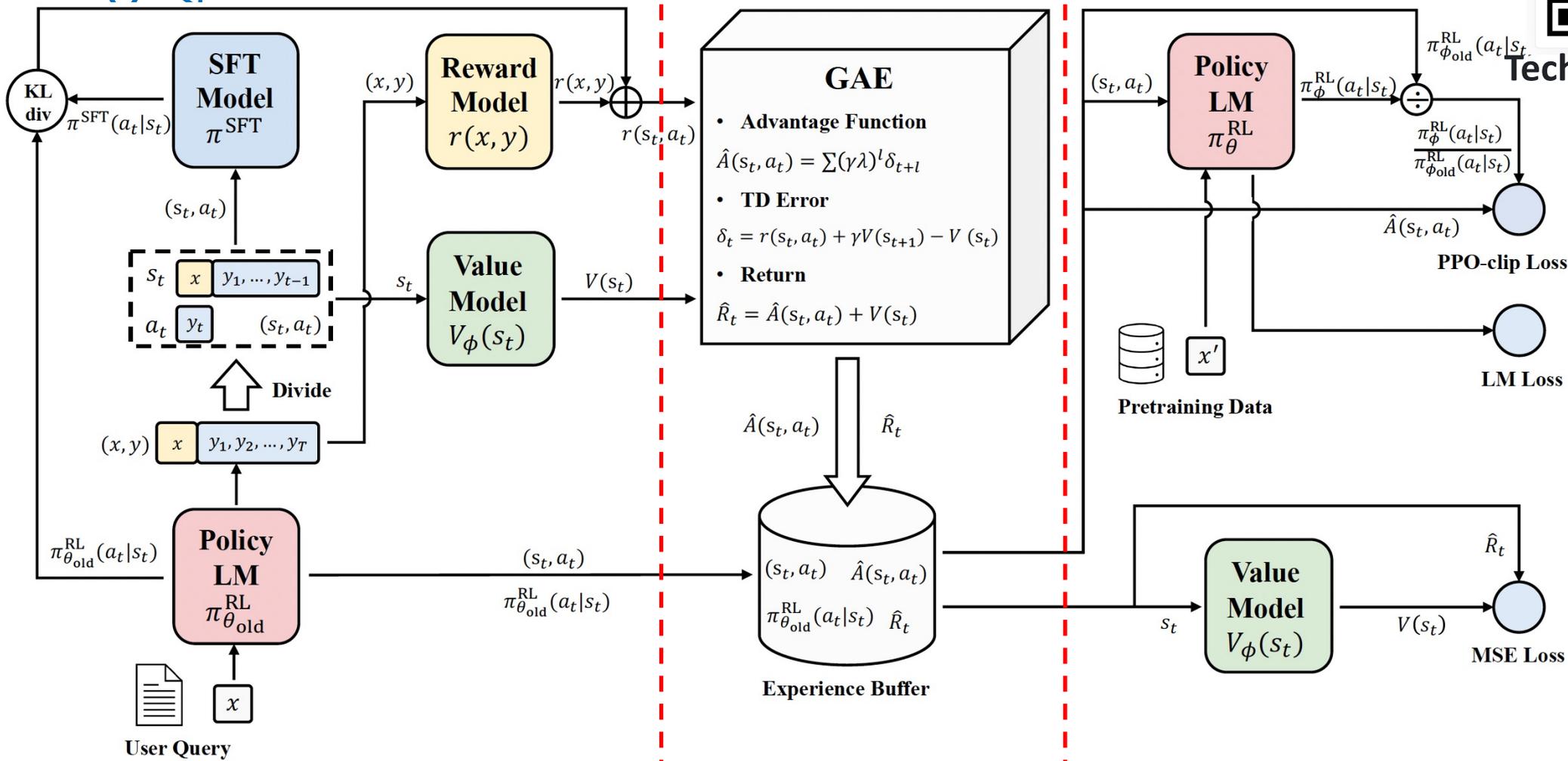
A is better

B is better

强化学习



Technical report



Interact with Environment

Advantage Estimation

Model Optimization



RLHF 对于减轻曝光偏置有确定性作用

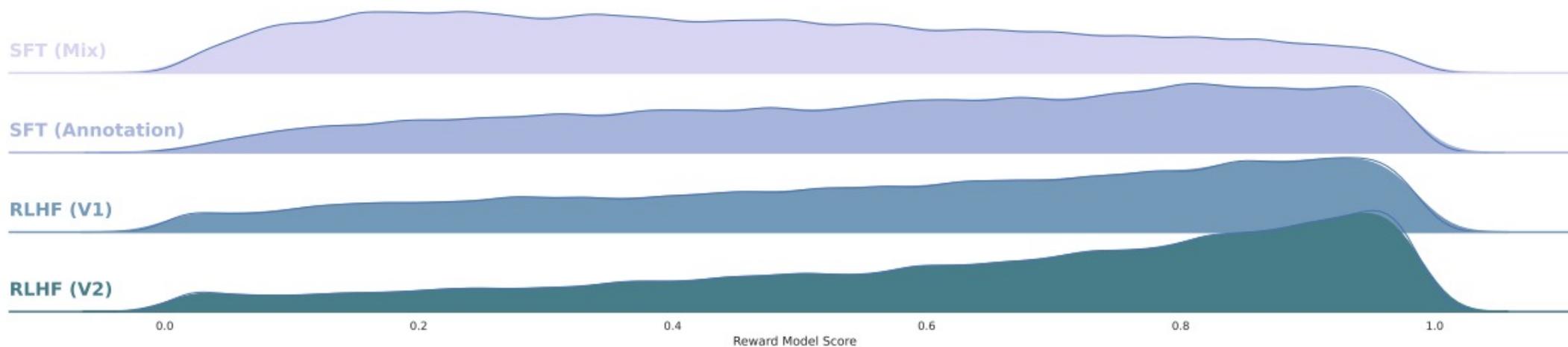
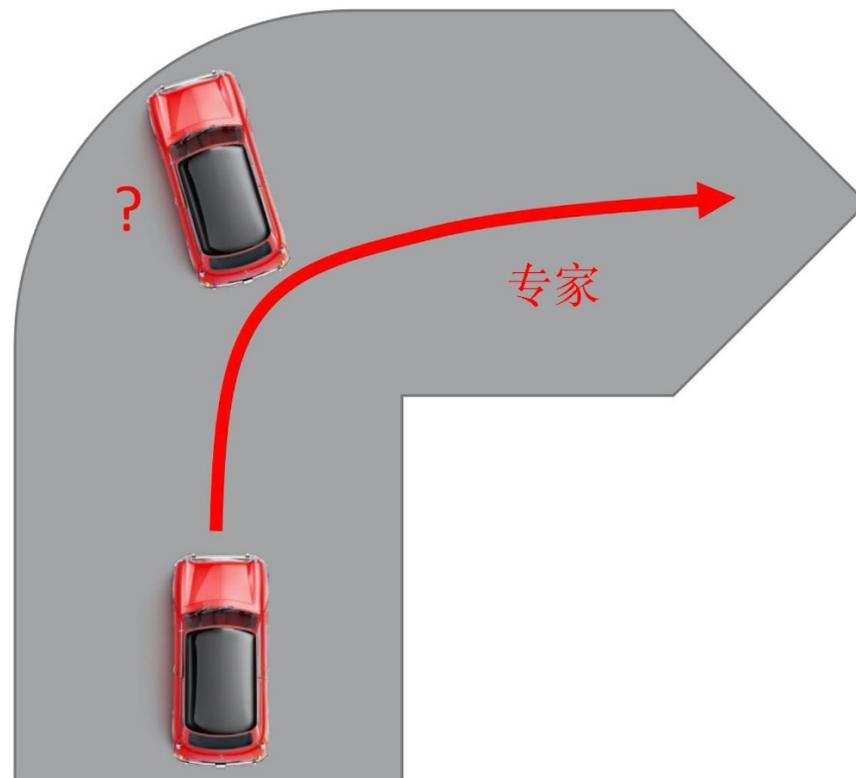


Figure 20: Distribution shift for progressive versions of LLAMA 2-CHAT, from SFT models towards RLHF.

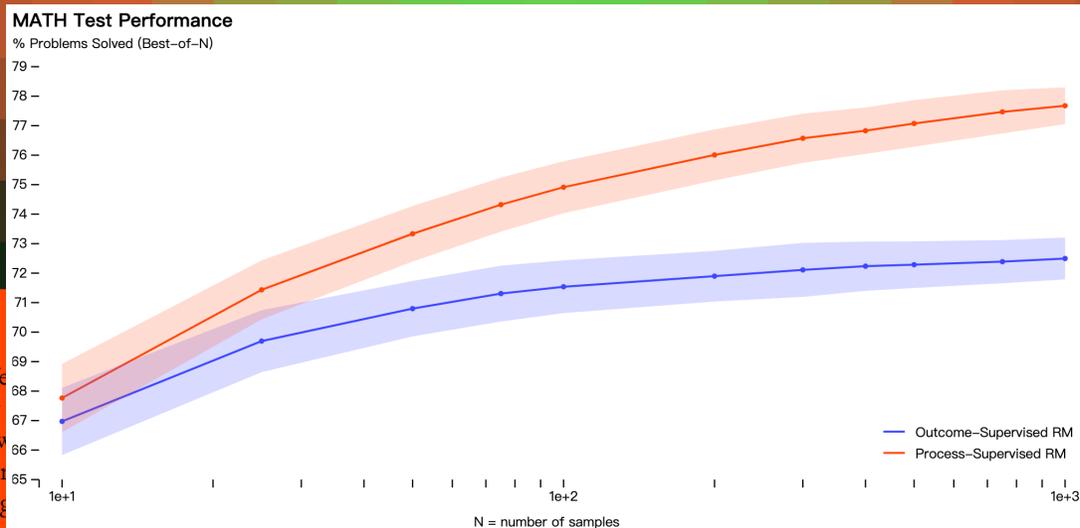
强化学习



Behavior Cloning

强化学习—过程优化

Improving mathematical reasoning with process supervision



Let

$$x^8 + 3x^4 - 4 = p_1(x)p_2(x) \cdots p_k(x),$$

where each non-constant polynomial $p_i(x)$ is monic with integer coefficients, and cannot be factored further over the integers. Compute $p_1(1) + p_2(1) + \cdots + p_k(1)$.

This problem is about factoring a polynomial over the integers, which can be a tricky task.

One possible strategy is to look for rational roots of the polynomial, using the rational root theorem.

However, that would involve checking a lot of possible factors of -4 , and there is no guarantee that the polynomial has any rational roots at all.

Another possible strategy is to use some algebraic tricks to simplify the polynomial or reduce its degree.

I notice that the polynomial has only even powers of x , so I can try to substitute $y = x^2$ and see what happens.

Then the polynomial becomes $y^4 + 3y^2 - 4$, which is a quadratic equation in y^2 .

I can try to factor this quadratic equation by looking for two numbers that multiply to -4 and add to 3 .

Those numbers are 4 and -1 , so I can write $y^4 + 3y^2 - 4 = (y^2 + 4)(y^2 - 1)$.

Now I can substitute back $x^2 = y$ and get $x^8 + 3x^4 - 4 = (x^4 + 4)(x^4 - 1)$.

This looks promising, because I have reduced the degree of the factors from 8 to 4 .

I can try to factor each of these factors further, using the difference of squares formula.

$$\text{I get } x^8 + 3x^4 - 4 = (x^2 + 2)(x^2 - 2)(x^2 + 1)(x^2 - 1).$$

I can apply the difference of squares formula again to the last factor and get $x^8 + 3x^4 - 4 = (x^2 + 2)(x^2 - 2)(x^2 + 1)(x + 1)(x - 1)$.

Now I have factored the polynomial completely into monic linear and quadratic factors with integer coefficients.

These are the $p_i(x)$'s that the problem is asking for.

To find the sum of their values at $x = 1$, I just need to plug in $x = 1$ into each factor and add them up.

$$\text{I get } p_1(1) + p_2(1) + \cdots + p_k(1) = (1^2 + 2)(1^2 - 2)(1^2 + 1)(1 + 1)(1 - 1).$$

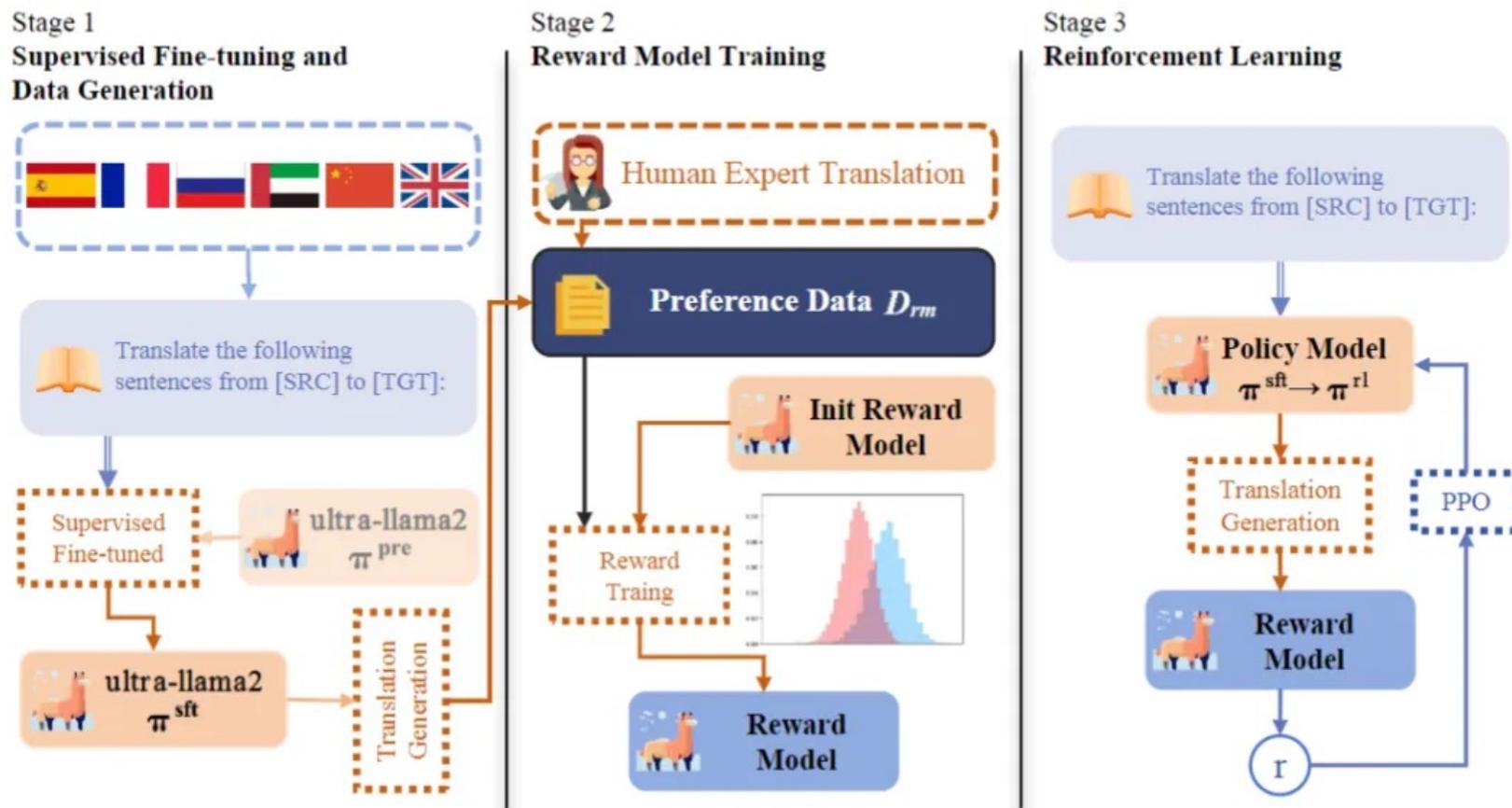
Simplifying, I get $p_1(1) + p_2(1) + \cdots + p_k(1) = (3)(-1)(2)(2)(0)$.

$$\text{Multiplying, I get } p_1(1) + p_2(1) + \cdots + p_k(1) = 0.$$

Answer: 0



使用RLHF推动翻译偏好建模：低成本实现“信达雅”



使用RLHF推动翻译偏好建模：低成本实现“信达雅”

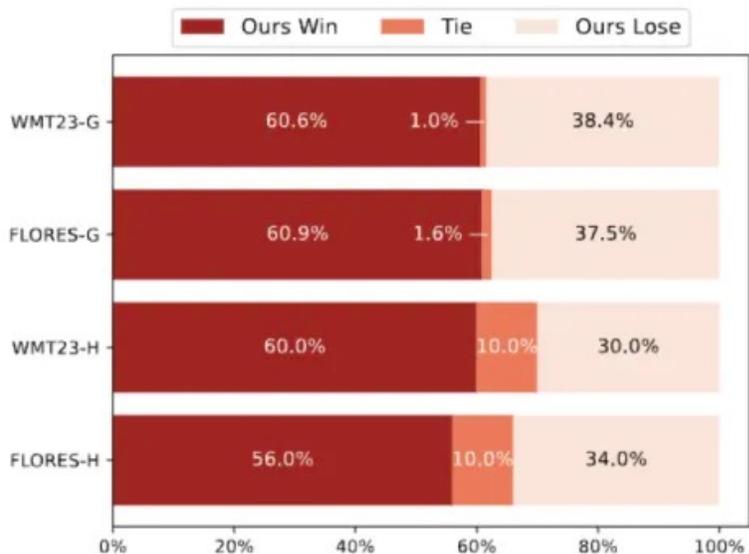


Figure 3: Comparison between preference optimized models and the SFT model on Task En→Zh. G and H represent GPT-4 and humans as evaluators, respectively.

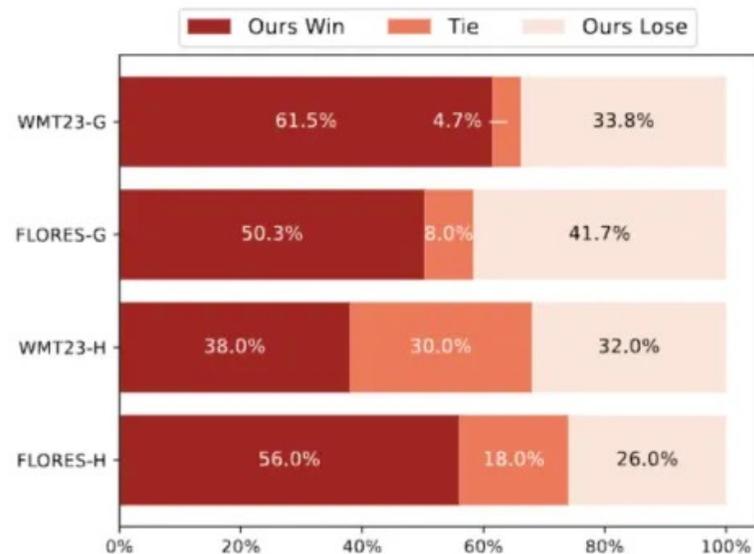


Figure 4: Comparison between preference optimized models and the SFT model on Task Zh→En. G and H represent GPT-4 and humans as evaluators, respectively.

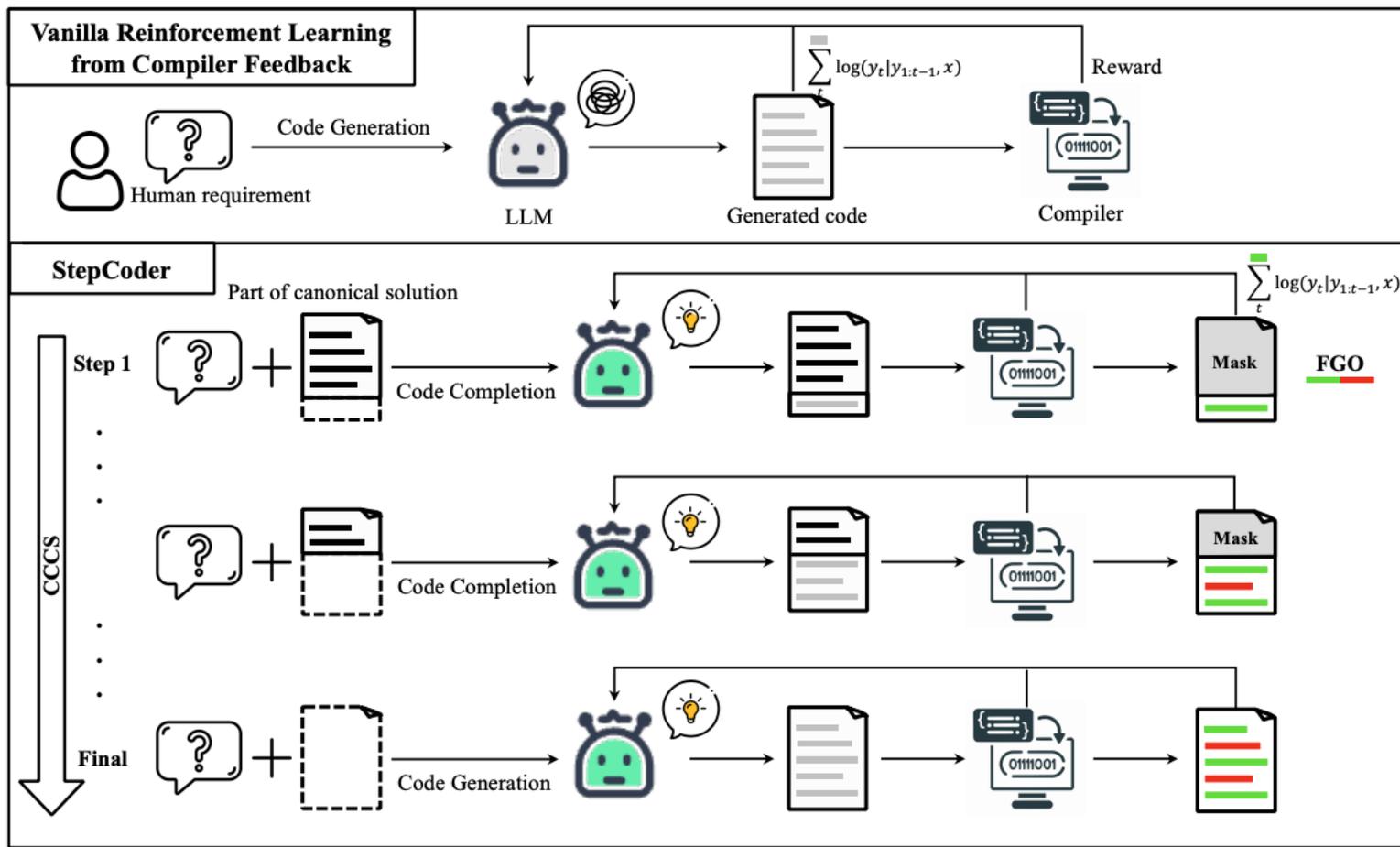


使用RLHF推动翻译偏好建模：低成本实现“信达雅”

Faithfulness	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。
	RLHF	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。
	Commentary	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.
Expressiveness	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。
	RLHF	经过多年的锻炼，跑马拉松对她来说已是小菜一碟了。
	Commentary	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.
Elegance	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	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。
	RLHF	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。
	Commentary	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.



从编译器反馈信号中提升代码生成任务效果



CCCS (Curriculum of Code Completion Subtasks) 的目的是将代码生成任务分解为代码完成子任务的课程, 可以减轻RL中的探索挑战;

FGO (Fine-Grained Optimization) 专为代码生成任务而设计, 通过只计算已执行代码片段的损失来提供细粒度优化。



在各项任务中都取得了超越基线模型的结果

Models	Size	APPS+			Overall
		Introductory	Interview	Competition	
Base Models					
CodeLlama [27]	13B	18.7	11.0	0.0	13.0
CodeLlama-Python [27]	13B	29.0	12.3	2.9	17.9
DeepSeek-Coder-Base [8]	6.7B	13.0	10.3	5.0	10.9
Supervised Fine-tuned Models					
StarCoder [15]	15.6B	6.3	4.1	0.7	4.7
CodeLlama-Instruct [27]	13B	33.3	11.0	1.4	18.7
WizardCoder-Python-V1.0 [23]	13B	39.7	15.1	4.3	23.6
DeepSeek-Coder-Instruct [8]	6.7B	49.4	18.7	3.6	29.2
SFT on APPS+	6.7B	50.1	19.0	6.4	29.8
Reinforcement Learning-based Models (Using DeepSeek-Coder-Instruct-6.7B as the backbone)					
Vanilla PPO	6.7B	53.7	20.1	5.0	31.7
PPOCoder [33]	6.7B	54.4	20.3	6.4	32.1
RLTF [20]	6.7B	55.1	20.8	6.4	32.7
StepCoder (Ours)	6.7B	59.7	23.5	8.6	36.1
w/o CCCS	6.7B	58.7	21.7	7.1	34.6
w/o FGO	6.7B	58.4	23.3	8.6	35.5

核心结论

1. 提升任务效果依然需要一定数量的标注数据
2. 多任务之间的相互影响和关系仍需仔细研究
3. 多任务的训练方法仍然缺乏统一认识
4. 强化学习对于生成任务效果提升具有重要作用

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. 简单增大数据量无法实现效果增加
4. 标注数据的准确程度要求十分苛刻

忘记 AGI、涌现、对齐、激发 ...

从统计机器学习角度 “再出发”



谢谢!