如何提升大模型任务能力

张奇

复旦大学

ChatGPT实现过程

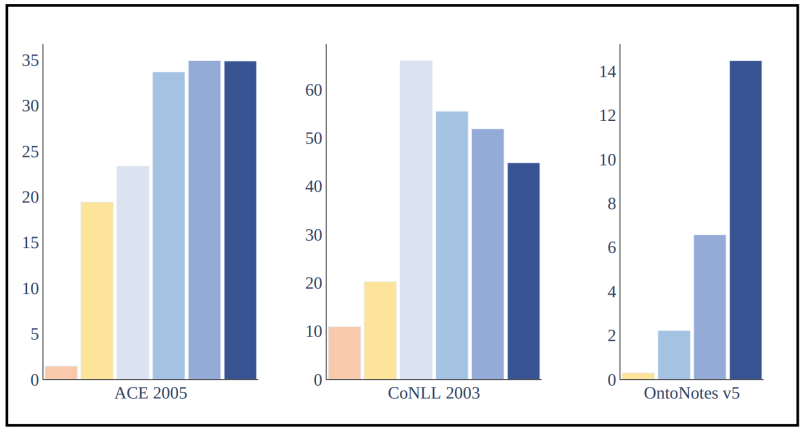
	预训练阶段	指令微调	奖励函数	强化学习
数据集合	原始数据 数千亿单词：图书、百科、网页等	标注用户指令 百万用户指令和对应的答案	标注对比对 百万标注对比对	用户指令 十万用户指令
算法	语言模型预训练	语言模型预训练	二分类模型	强化学习方法
模型	基础模型	SFT 模型	RM 模型	RL 模型
资源需求	1000+GPU 月级别训练时间 GPT 3.0、LLaMa、PaLM	1-100GPU 天级别训练时间 MOSS、ChatGLM6b、Vicuna-13B等	1-100GPU 天级别训练时间	1-100GPU 天级别训练时间 ChatGPT、Claude

大模型的能力是如何获取的

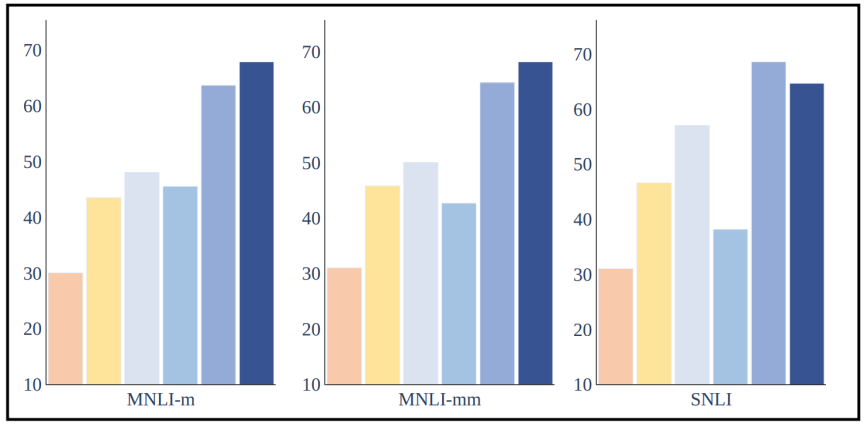
预训练阶段	指令微调	奖励函数	强化学习
原始数据 **数千亿**单词：图书、百科、网页等	标注用户指令 **百万**用户指令和对应的答案	标注对比对 **百万**标注对比对	用户指令 **十万**用户指令
语言模型预训练	语言模型预训练	二分类模型	强化学习方法
基础模型	SFT 模型	RM 模型	RL 模型
知识压缩和表示学习	**能力注入**	**生成式任务能力提升**	

所有的能力都需要精心设计

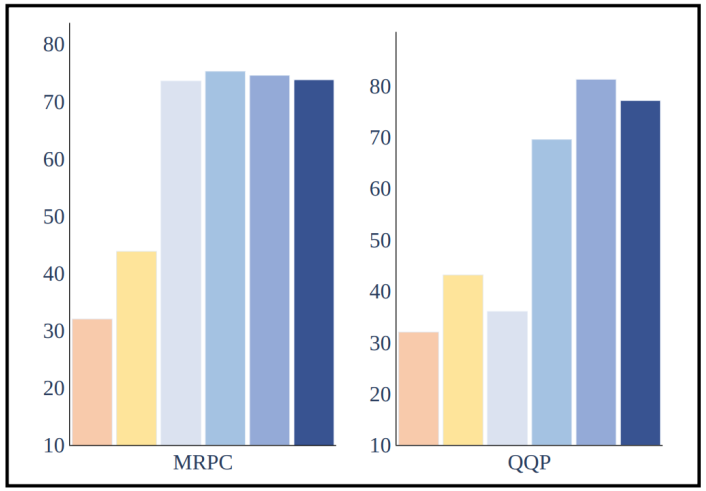
很多任务的能力在一开始并不具备，而是不断叠加上去的



Named Entity Recognition (ENG)



Natural Language Inference



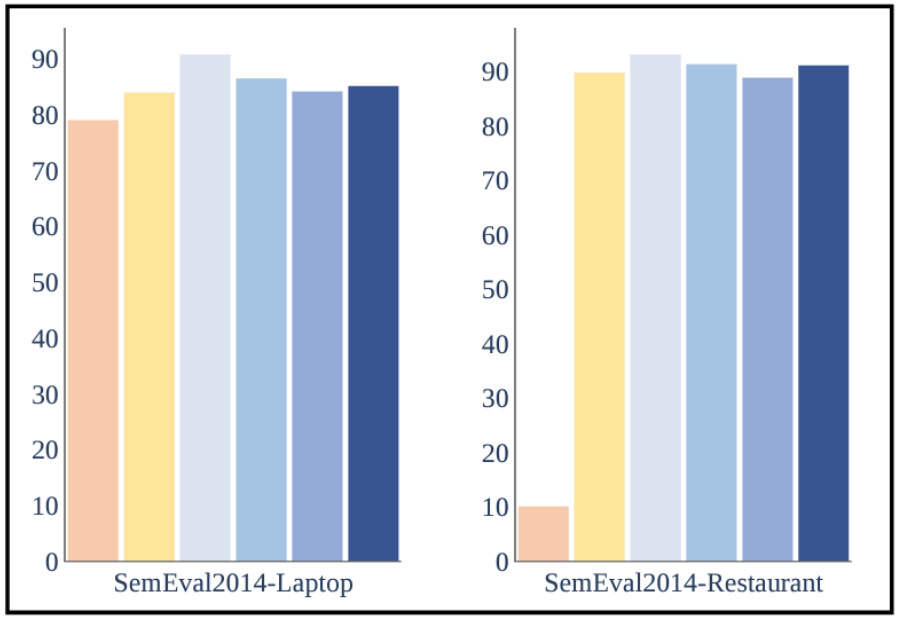
Semantic Matching

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

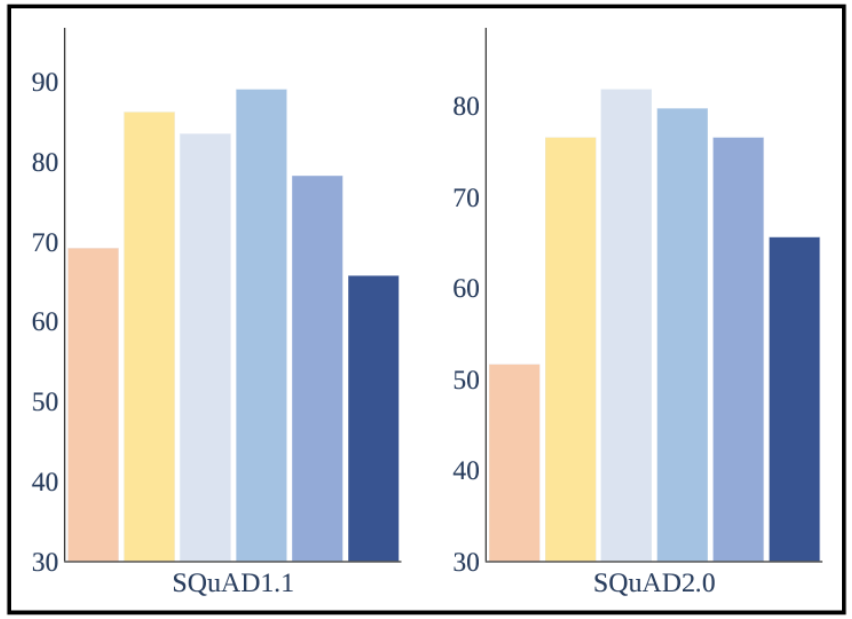
A comprehensive capability analysis of gpt-3 and gpt-3.5 series models, Arxiv 2023

所有的能力都需要精心设计

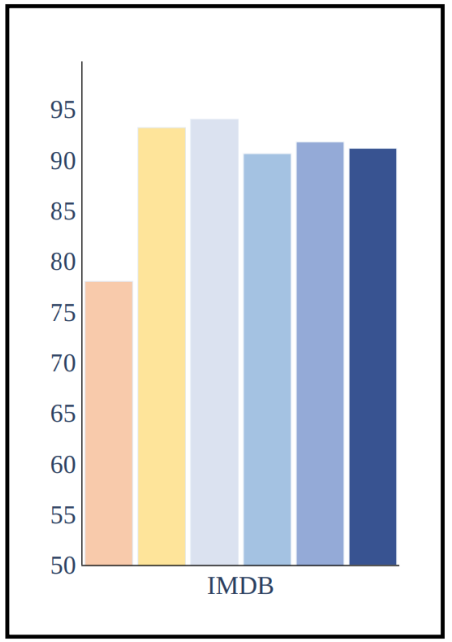
任务大量叠加会造成一些任务能力下降



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**

A comprehensive capability analysis of gpt-3 and gpt-3.5 series models, Arxiv 2023

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

Physics of Language Models: Part 3.3, Knowledge Capacity Scaling Laws, Meta FAIR Labs, arXiv 2024

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

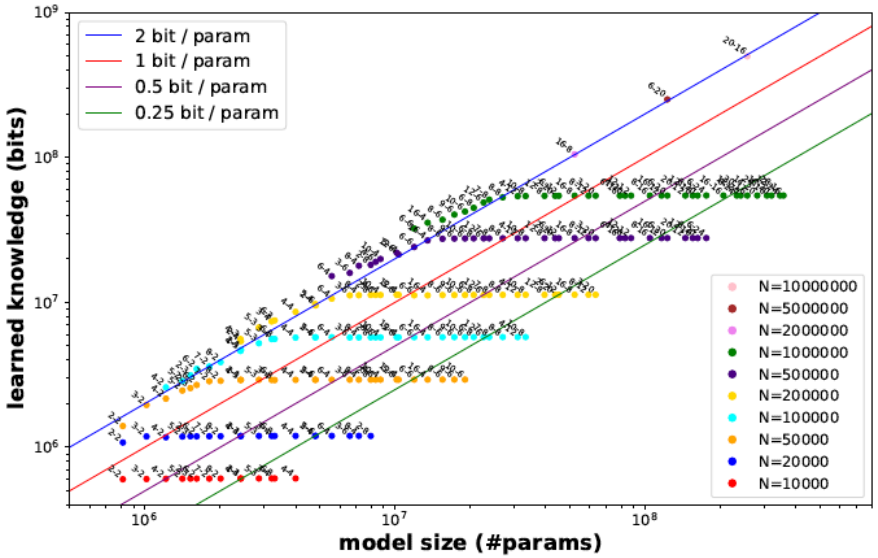
知识记忆不是word-by-word记忆，而是能够通过问答获取答案

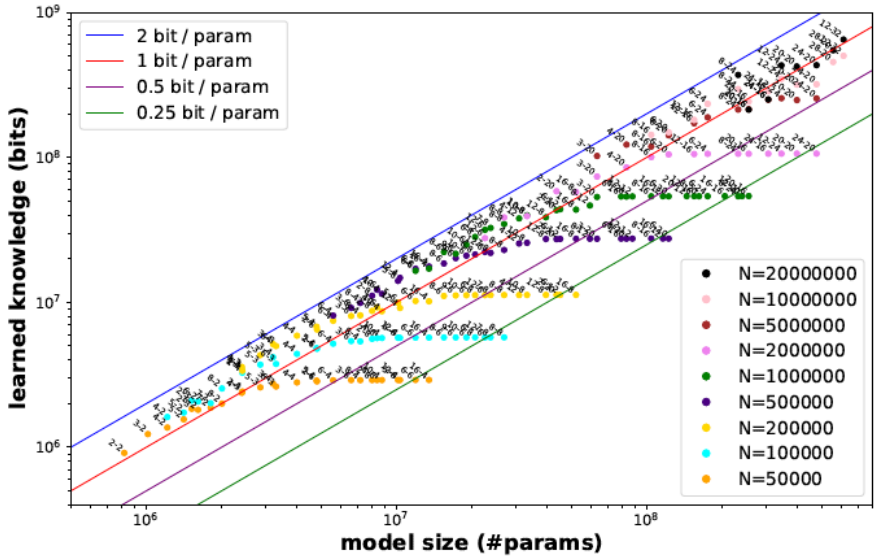
Physics of Language Models: Part 3.3, Knowledge Capacity Scaling Laws, Meta FAIR Labs, arXiv 2024

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**





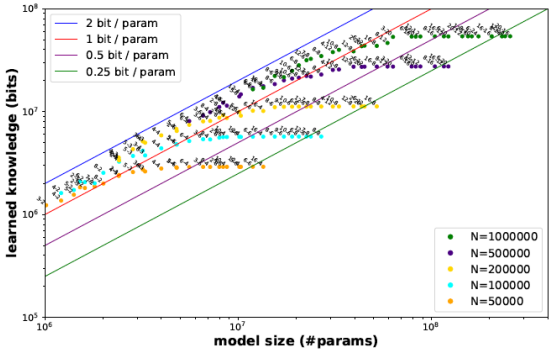
(a) bioS(N) data — **1000 exposures** — peak $R(F) \geq 2$ (b) bioS(N) data — **100 exposures** — peak $R(F) \geq 1$

图标上面数字是l,h参数选择

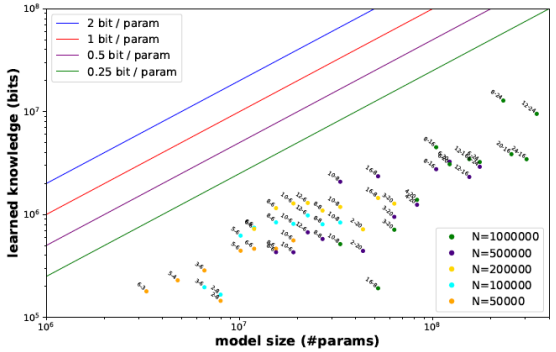
Physics of Language Models: Part 3.3, Knowledge Capacity Scaling Laws, Meta FAIR Labs, arXiv 2024

“垃圾”数据对知识获取有显著影响

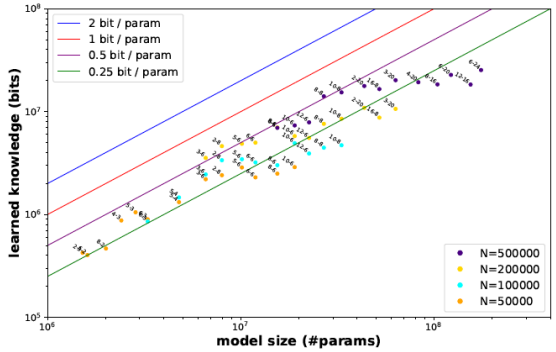
Wikipedia vs. Common Crawl



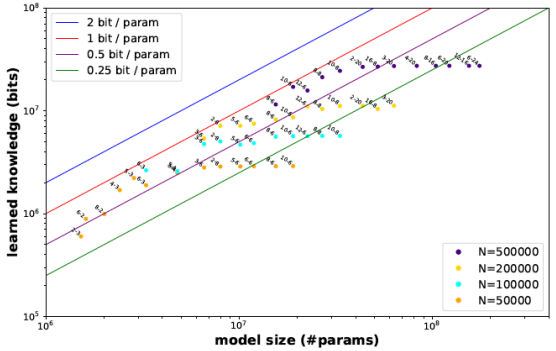
(a) no junk, 100 exposures



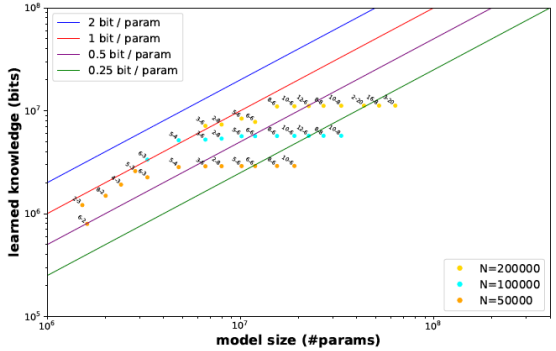
(b) 7/8 junk, 100 exposures



(c) 7/8 junk, 300 exposures



(d) 7/8 junk, 600 exposures



(e) 7/8 junk, 1000 exposures

Physics of Language Models: Part 3.3, Knowledge Capacity Scaling Laws, Meta FAIR Labs, arXiv 2024

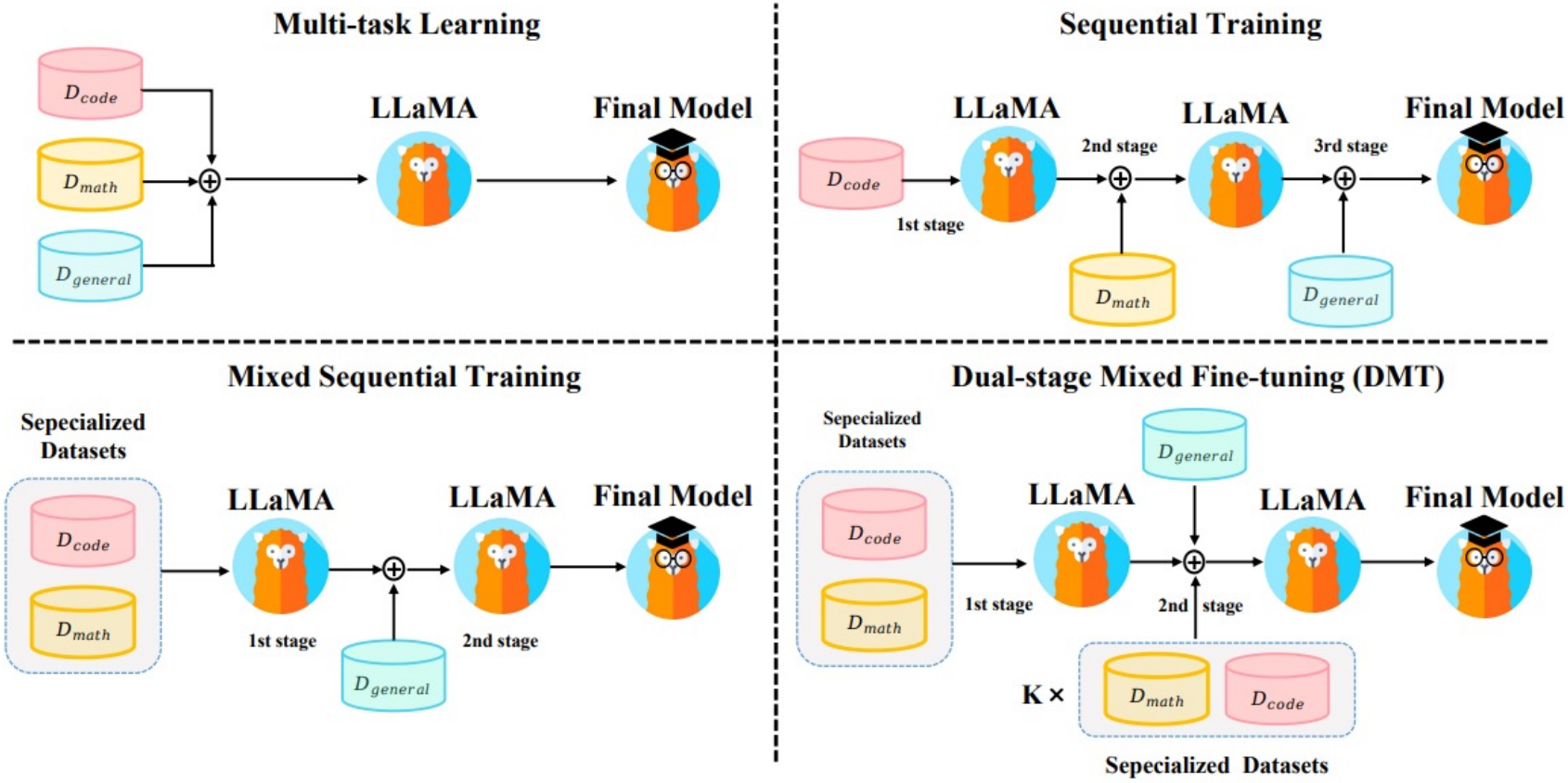
预训练阶段结论

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

Physics of Language Models: Part 3.1, Knowledge Storage and Extraction, Meta FAIR Labs, arXiv 2023
Physics of Language Models: Part 3.2, Knowledge manipulation, Meta FAIR Labs, arXiv 2023
Physics of Language Models: Part 3.3, Knowledge Capacity Scaling Laws, Meta FAIR Labs, arXiv 2024

有监督微调阶段如何训练?

有监督微调的四种方式



How Abilities in Large Language Models are Affected by Supervised Fine-tuning Data Composition, arXiv 2024

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?

How Abilities in Large Language Models are Affected by Supervised Fine-tuning Data Composition, arXiv 2024

实验设置

SFT 数据集 $\{D_1, D_2, \ldots, D_k\}$，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

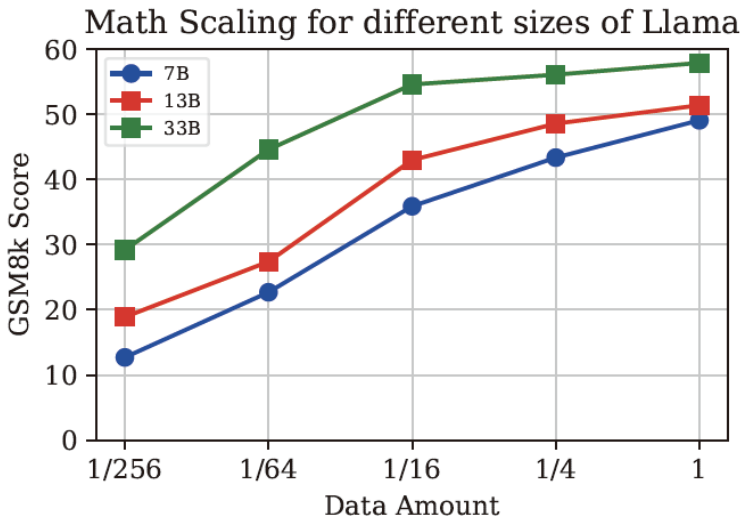
测试数据集：

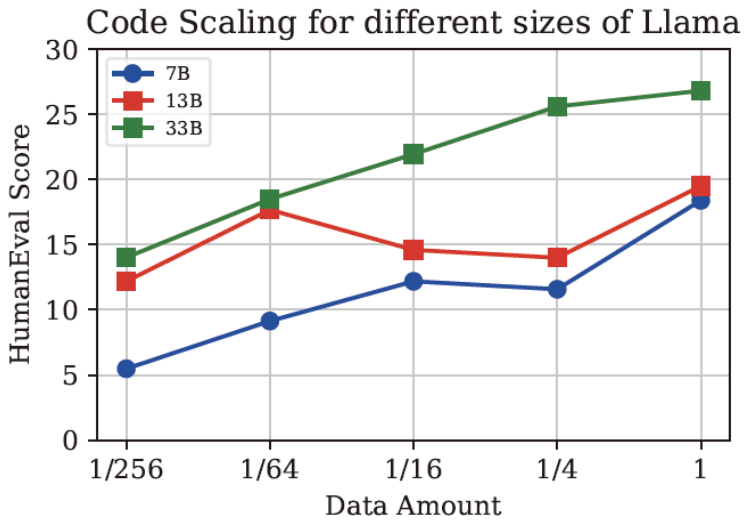
数学：GSM8K Test Set

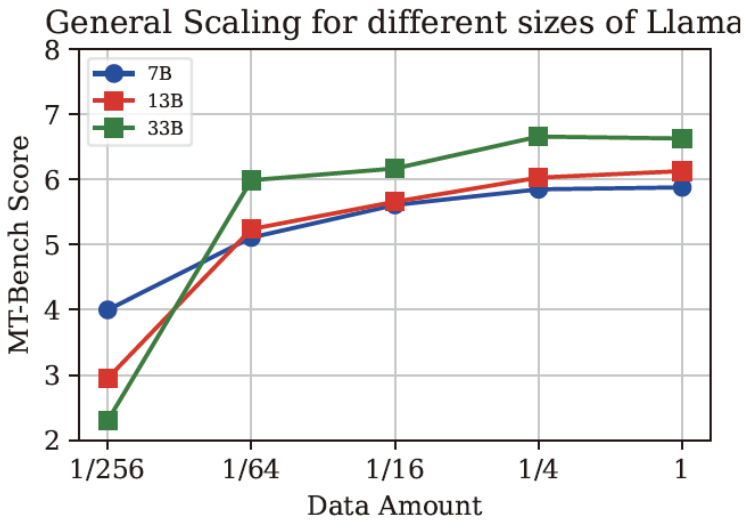
编程：Humaneval

通用：MT-Bench

实验分析：RQ1 单个任务不同数据量





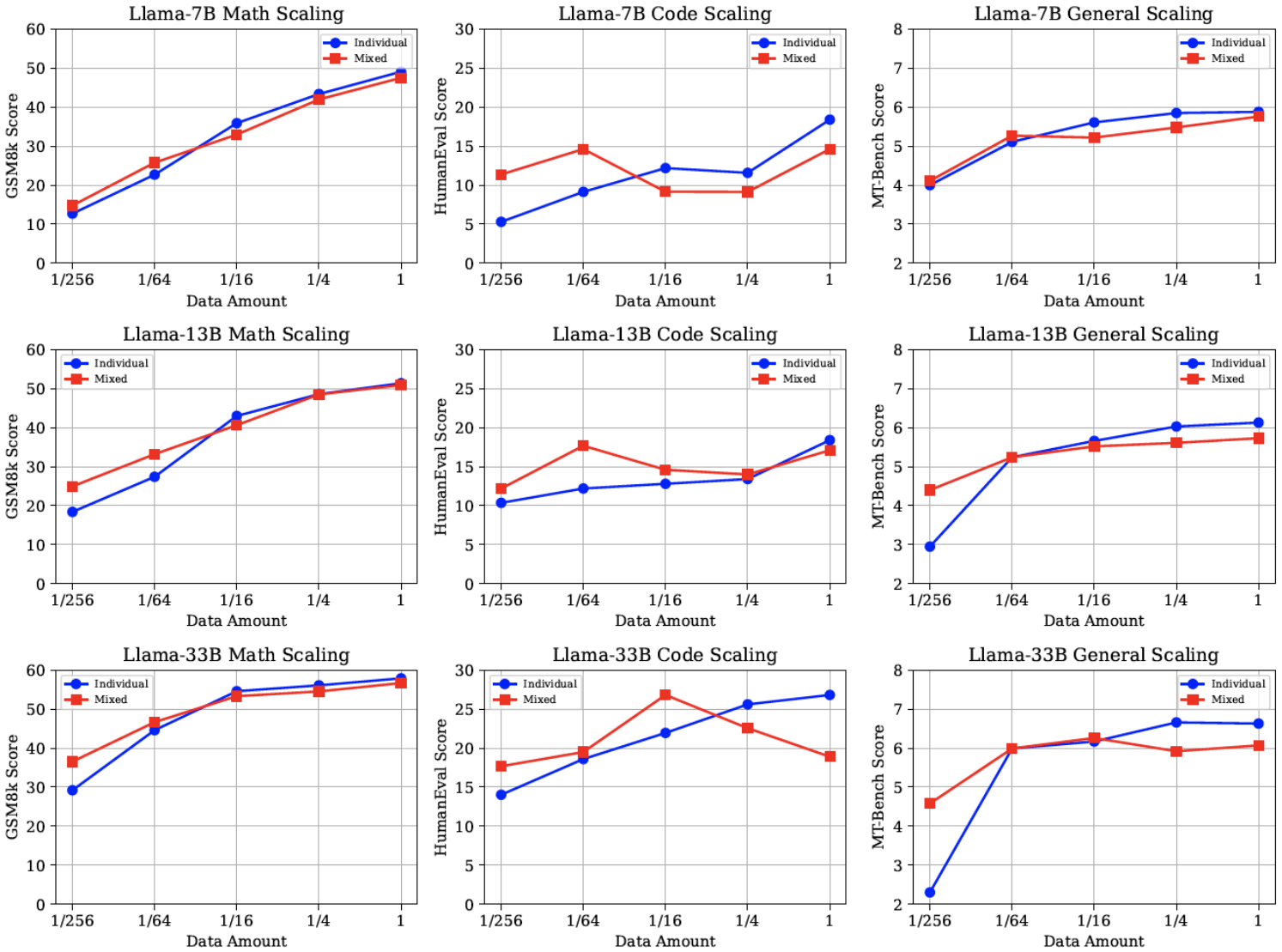


单个任务使用不同数据量进行训练

较大模型在相同的情况下表现出更好的性能

How Abilities in Large Language Models are Affected by Supervised Fine-tuning Data Composition, arXiv 2024

实验分析：RQ2 单个任务 vs. 混合任务

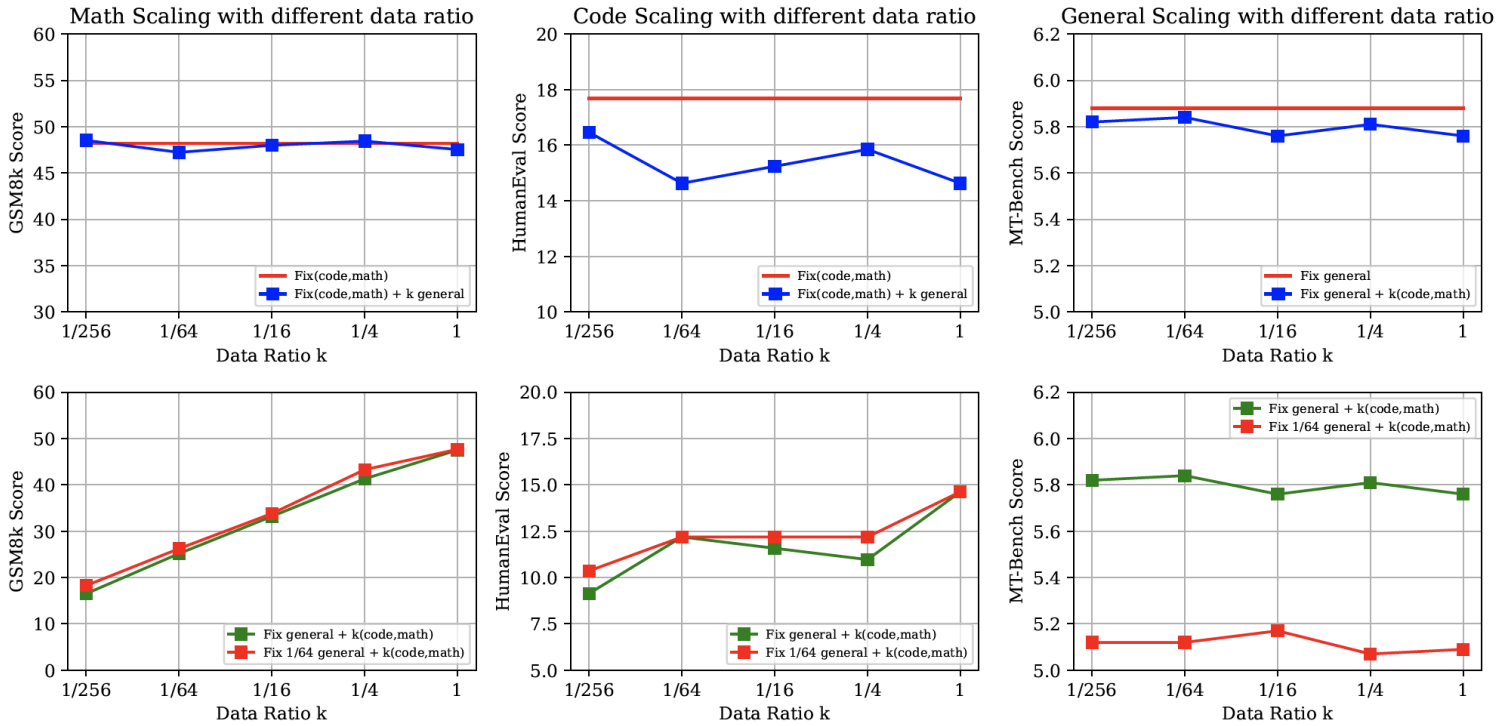


训练数据少时，能力得以提高；训练数据多时，混合数据则使得能力则减弱，与单个任务训练相比，这种现象更为明显。

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。

How Abilities in Large Language Models are Affected by Supervised Fine-tuning Data Composition, arXiv 2024

实验分析：RQ3 任务混合比例影响



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

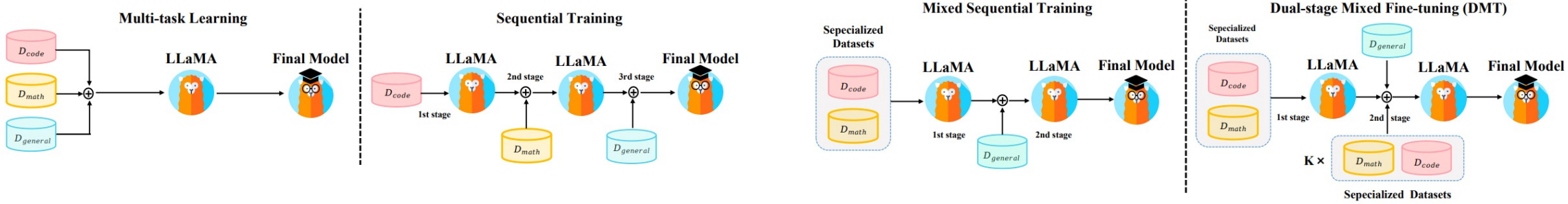
不同的SFT能力在任务格式和数据分布上存在显著差异时，**数据比例的影响是微不足道的**。

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。

How Abilities in Large Language Models are Affected by Supervised Fine-tuning Data Composition, arXiv 2024

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
Individual domain									
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
Different Training Strategies									
Multi-task learning	**47.53**	14.63	5.76	**50.94**	19.50	5.73	**56.69**	18.9	6.07
Sequential Training	31.39	15.85	5.72	39.12	**20.12**	5.93	47.27	24.80	**6.73**
Mixed Sequential Training	32.60	15.24	6.02	40.48	18.30	5.93	44.24	24.4	6.43
DMT(k=1/256)	41.92	**17.68**	**6.08**	46.47	19.50	**6.03**	56.36	**25.00**	**6.73**



How Abilities in Large Language Models are Affected by Supervised Fine-tuning Data Composition, arXiv 2024

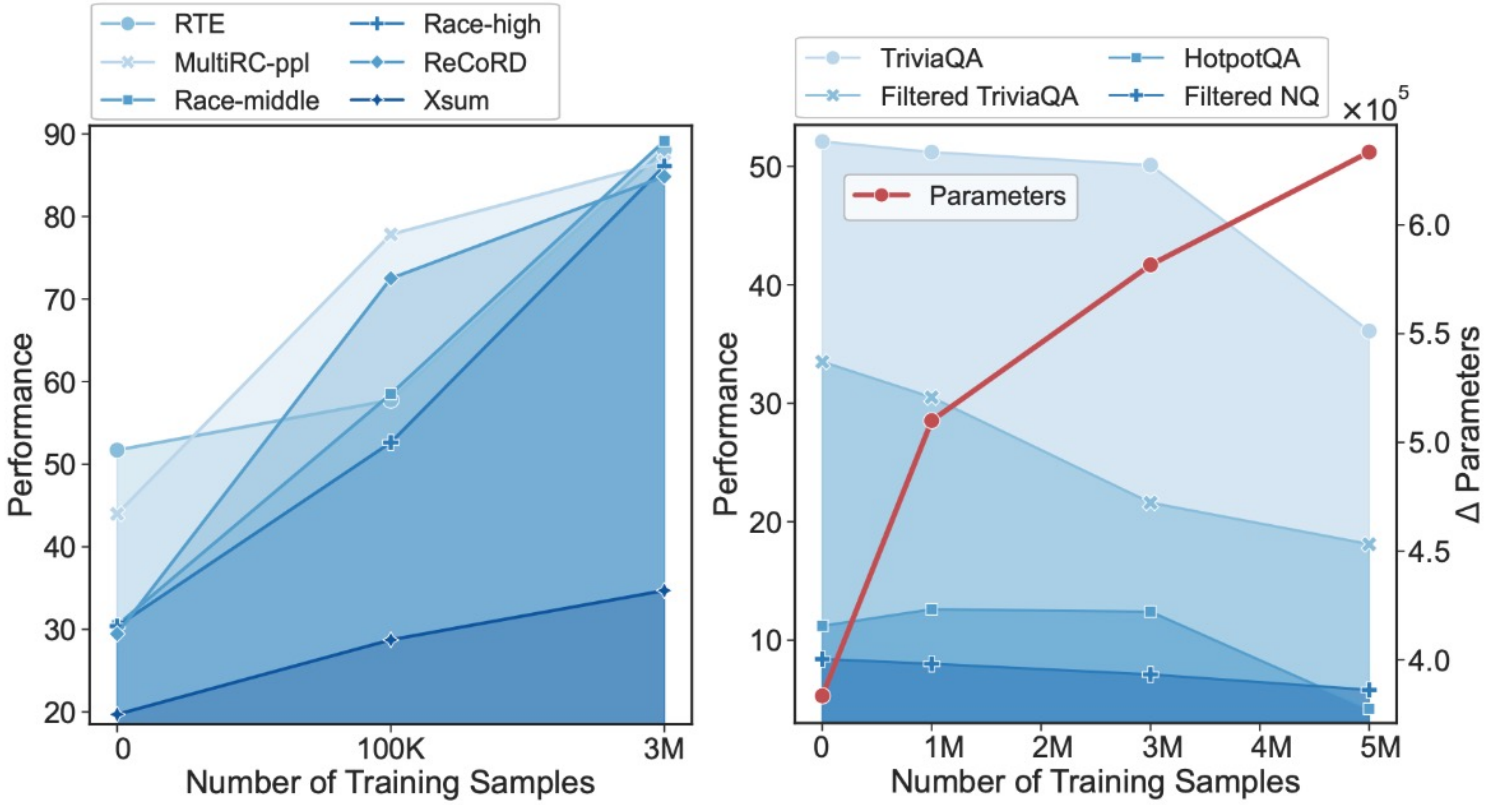
有监督微调阶段结论

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

How Abilities in Large Language Models are Affected by Supervised Fine-tuning Data Composition, arXiv 2024

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）。

Loramoe: Revolutionizing mixture of experts for maintaining world knowledge in language model alignment, arXiv 2023

大规模微调破坏世界知识

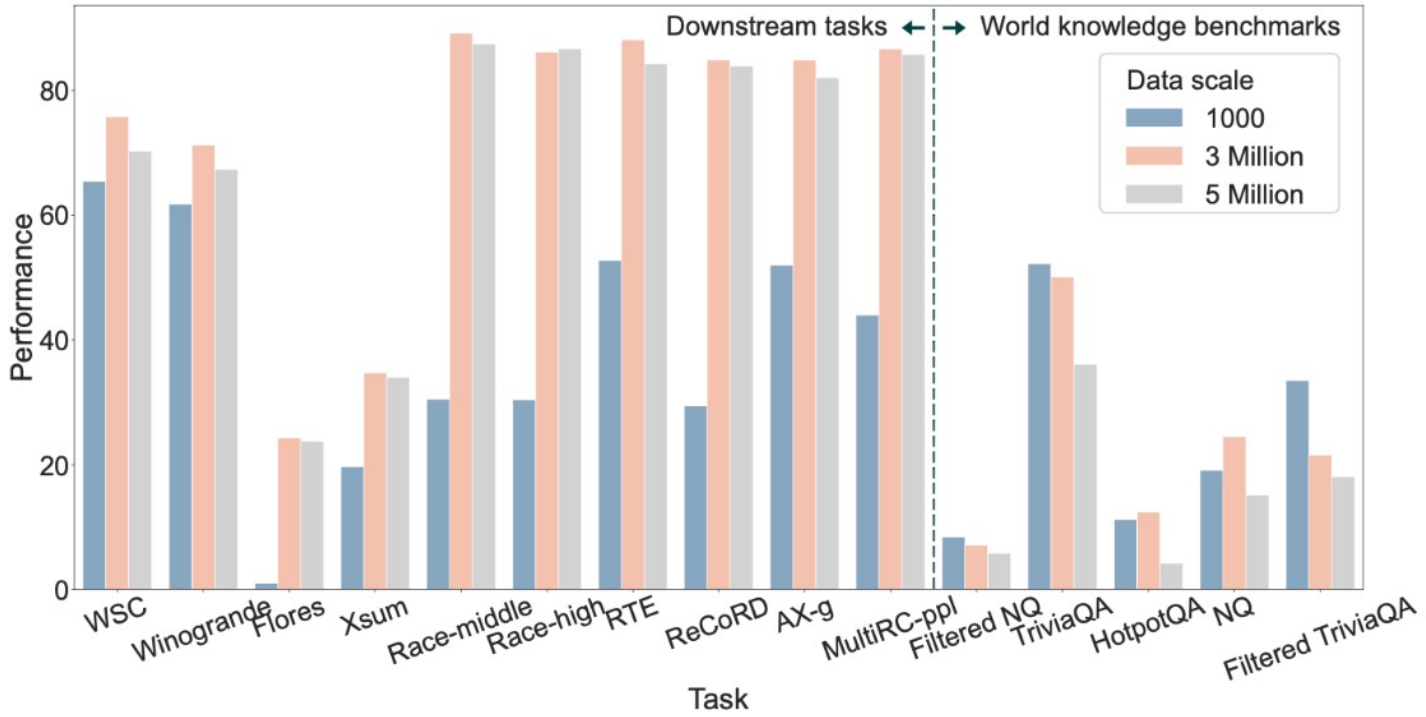


Figure 2: Performance on the various tasks after expanding the amount of fine-tuning data. For most of the downstream tasks (e.g., NLI and summarization), with the expansion of training data, performance on these tasks remains stable after improvement. Whereas, for the world knowledge benchmark, a significant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

Loramoe: Revolutionizing mixture of experts for maintaining world knowledge in language model alignment, arXiv 2023

CBQA的能力来源于预训练阶段

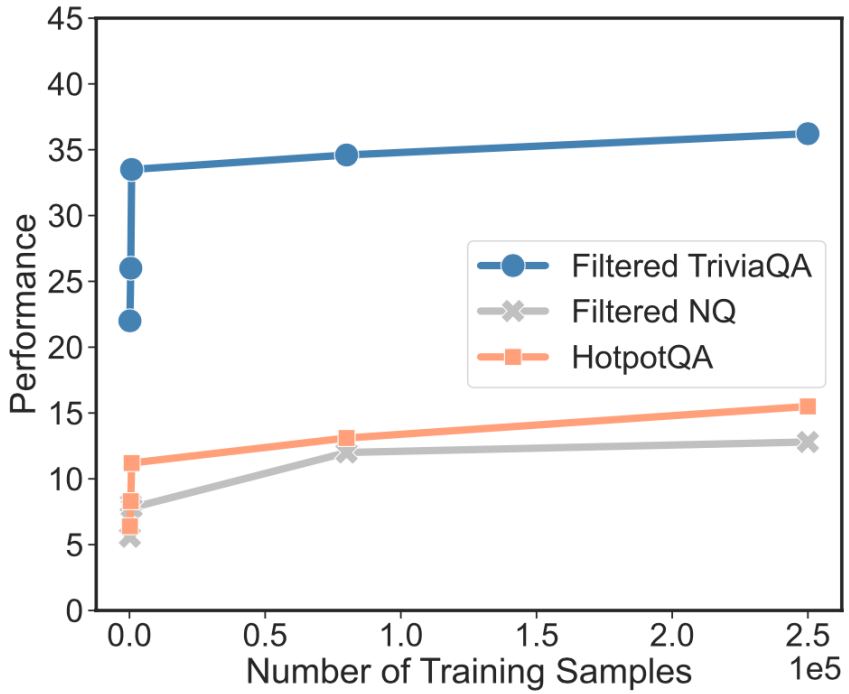


Figure 3: Performance on world knowledge benchmarks after training on CBQA solely. Its performance rises greatly after training with very few samples and remains relatively stable thereafter.

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。

Loramoe: Revolutionizing mixture of experts for maintaining world knowledge in language model alignment, arXiv 2023

LoRA+MoE

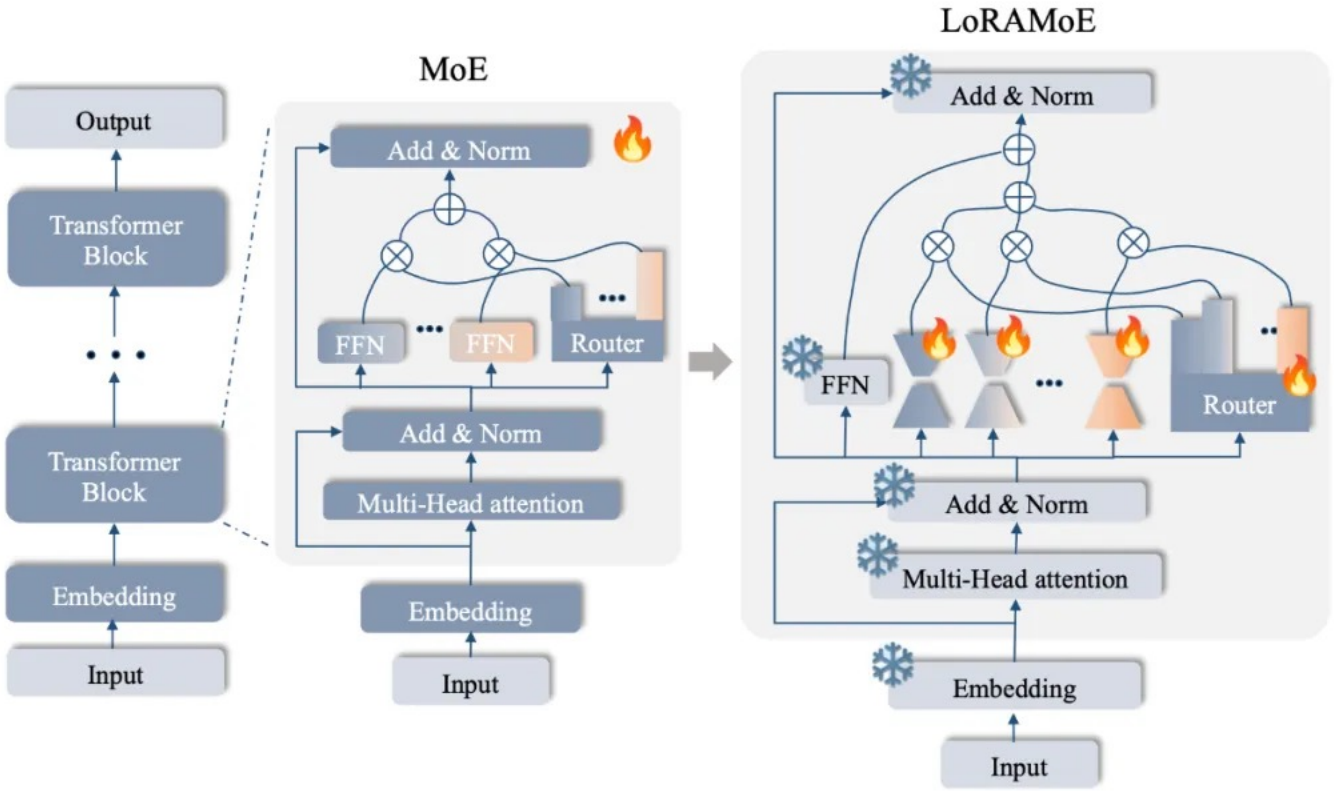


Figure 4: The architecture of LoRAMoE, compared with classic MoE. LoRAMoE utilizes multiple LoRAs as adaptable experts and a router to gate them in the FFN layer of every transformer block. During the training process, only the experts and the router are optimized.

Loramoe: Revolutionizing mixture of experts for maintaining world knowledge in language model alignment, arXiv 2023

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.

Loramoe: Revolutionizing mixture of experts for maintaining world knowledge in language model alignment, arXiv 2023

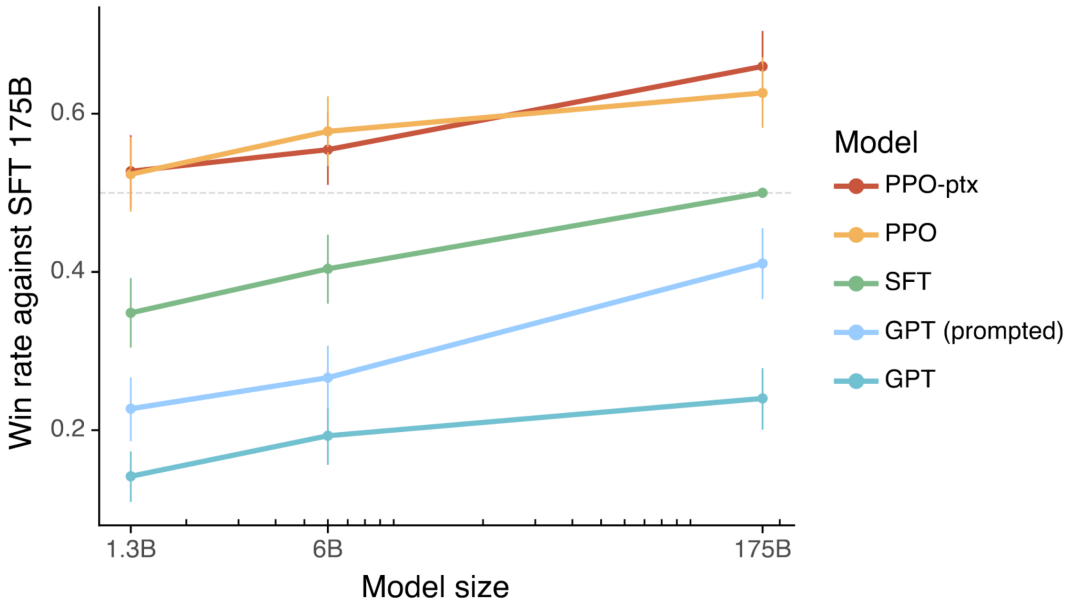
强化学习在生成式任务中的影响

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



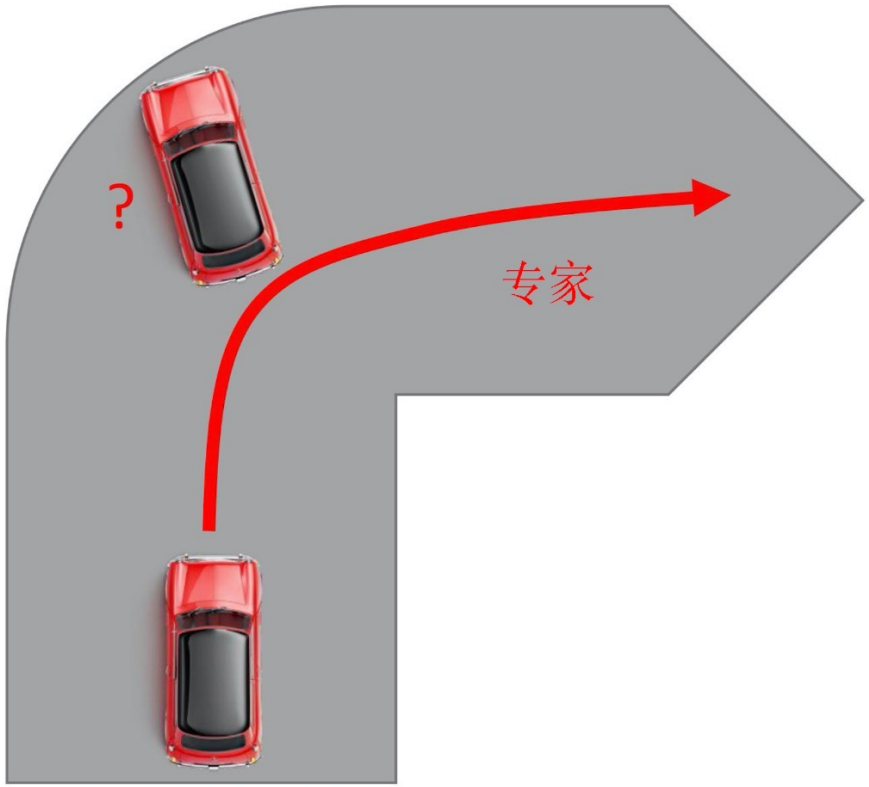
https://www.alignmentforum.org/posts/Rs9ukRphwg3pJeYRF/why-do-we-need-rlhf-imitation-inverse-rl-and-the-role-of

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.

Llama 2: Open Foundation and Fine-Tuned Chat Models, 2023

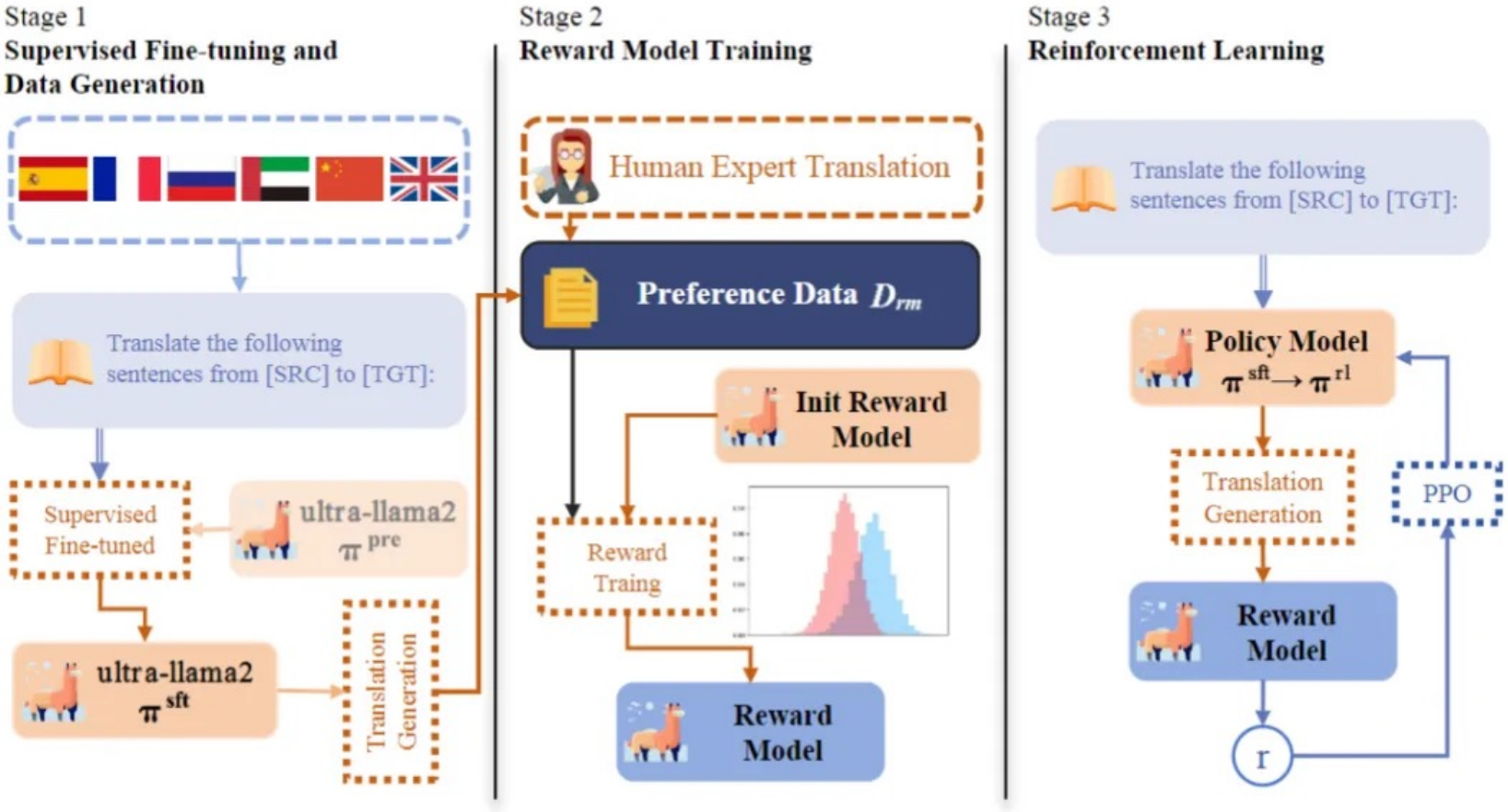
强化学习



Behavior Cloning

https://www.alignmentforum.org/posts/Rs9ukRphwg3pJeYRF/why-do-we-need-rlhf-imitation-inverse-rl-and-the-role-of

使用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

使用RLHF推动翻译偏好建模：低成本实现“信达雅”

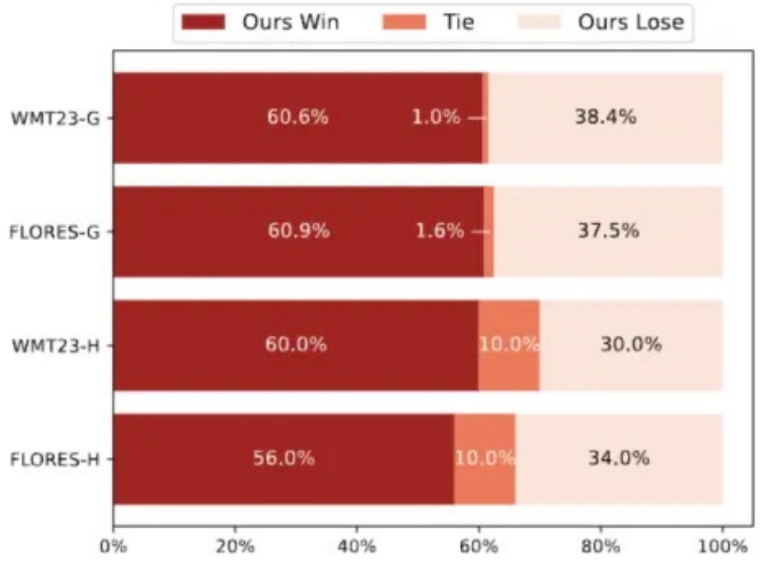


Figure 3: Comparison between preference optimized models and the SFT model on Task En→Zh. G and H represent GPT-4 and humans as evaluators, respectively.

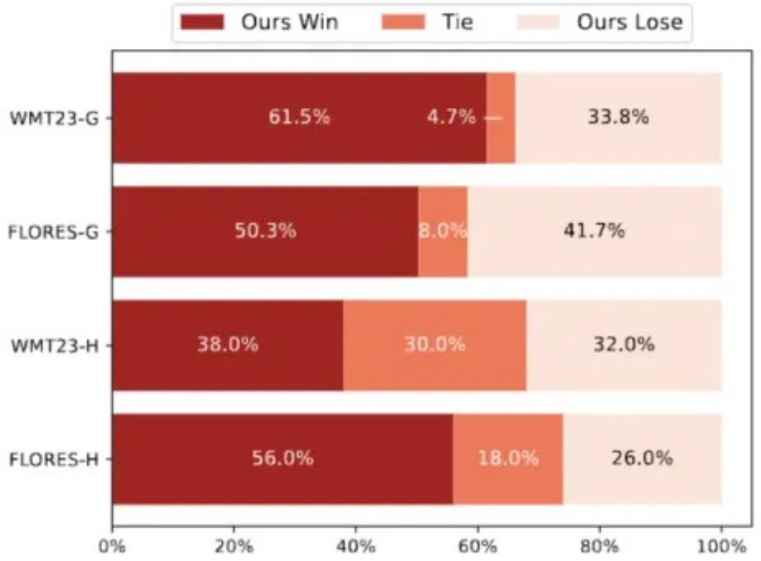


Figure 4: Comparison between preference optimized models and the SFT model on Task Zh→En. G and H represent GPT-4 and humans as evaluators, respectively.

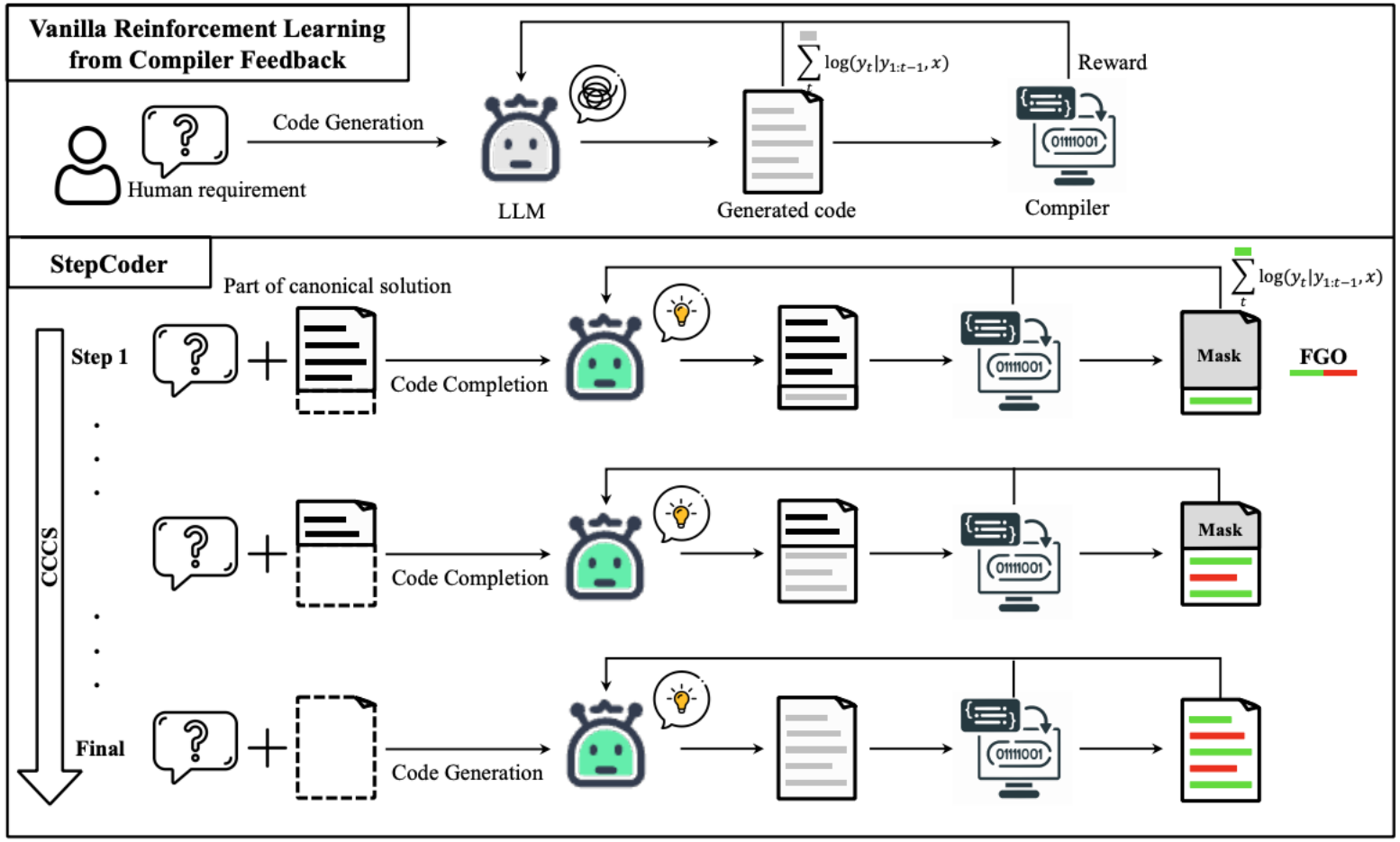
Advancing Translation Preference Modeling with RLHF: A Step Towards Cost-Effective Solution, arXiv, 2024

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.

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

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。

StepCoder: Improve Code Generation with Reinforcement Learning from Compiler Feedback, arXiv, 2024

在各项任务中都取得了超越基线模型的结果

Models	Size	APPS+ Introductory	APPS+ Interview	APPS+ Competition	APPS+ Overall
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

StepCoder: Improve Code Generation with Reinforcement Learning from Compiler Feedback, arXiv, 2024

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”


FUDAN UNIVERSITY
1905


FUDAN UNIVERSITY
NLP
NATURAL LANGUAGE PROCESSING

谢谢！