AI Box

指令微调与常见策略

《大语言模型》编写团队：唐天一

大模型具有很好的指令遵循能力

AI Box

已知草莓的英文单词中共有a个r，新中国成立于b年，a*b=?，请用[数字]的格式输出答案

Qwen2.5-Max

我们来一步步解答这个问题。

1. **草莓的英文单词** 是 "strawberry"。
   统计字母 "r" 的个数：
   在 "strawberry" 中，字母 "r" 出现了 **3次** 。
   所以，$a = 3$。
2. **新中国成立的年份** 是 **1949年** 。
   所以，$b = 1949$。
3. 计算 $ a \times b $：

中英翻译

文本推理

常识理解

$$a \times b = 3 \times 1949 = 5847$$

数值计算

最终答案是：

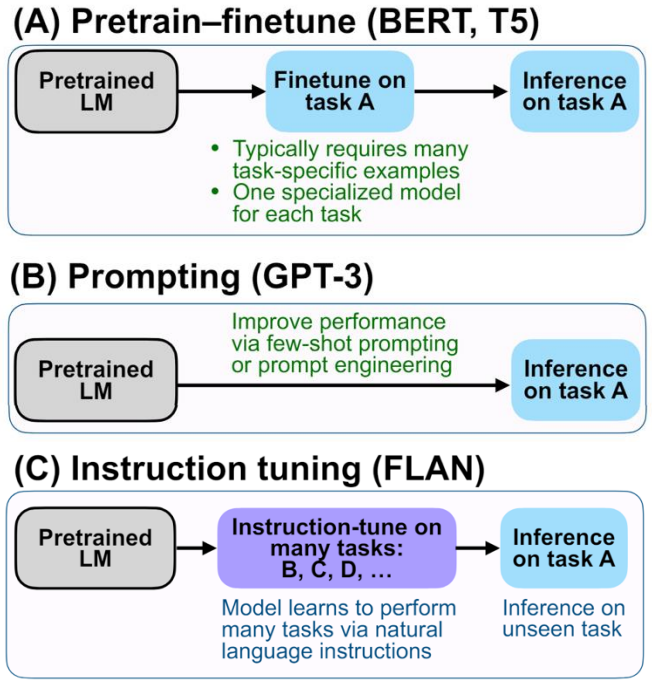
$$\boxed{5847}$$

格式遵循

指令微调

AI Box

- 使用问答形式的数据对大语言模型进行有监督微调
  - 大模型后训练中的关键步骤
  - 增强语言模型执行任务指令能力，提升任务泛化能力



“预训练–微调”范式：在单一任务微调+测试

“提示”范式：在未知任务上直接测试

“指令微调”范式：在多样任务上微调，在未知任务上测试

Finetuned Language Models Are Zero-Shot Learners, **ICLR 2022**

指令微调作用

AI Box

➢ 指令微调后，大语言模型可根据任务描述解决未见过任务

Model	Finetuning Mixtures	Tasks	Norm. avg.	MMLU Direct	MMLU CoT	BBH Direct	BBH CoT	TyDiQA Direct	MGSM CoT
8B	None (no finetuning)	0	6.4	24.3	24.1	30.8	30.1	25.0	3.4
	CoT	9	8.3 (+1.9)	26.3	32.1	19.8	26.6	39.3	**10.4**
	CoT, Muffin	89	14.8 (+8.4)	37.6	38.4	31.0	30.9	32.4	8.4
	CoT, Muffin, T0-SF	282	20.5 (+14.1)	47.7	39.7	33.1	30.9	**49.0**	8.5
	CoT, Muffin, T0-SF, NIV2	1,836	**21.9** (+15.5)	**49.3**	**41.3**	**36.4**	**31.1**	47.5	8.2
62B	None (no finetuning)	0	28.4	55.1	49.0	37.4	43.0	40.5	18.2
	CoT	9	29.0 (+0.4)	48.5	48.7	34.5	39.5	48.8	**32.6**
	CoT, Muffin	89	33.4 (+6.0)	55.3	51.4	42.8	40.2	53.0	23.9
	CoT, Muffin, T0-SF	282	37.9 (+9.5)	**60.0**	56.0	44.7	43.8	58.2	30.0
	CoT, Muffin, T0-SF, NIV2	1,836	**38.8** (+10.4)	59.6	**56.9**	**47.5**	**44.9**	**58.7**	28.5
540B	None (no finetuning)	0	49.1	71.3	62.9	49.1	63.7	52.9	45.9
	CoT	9	52.6 (+3.5)	68.8	64.8	50.5	61.1	61.2	59.4
	CoT, Muffin	89	57.0 (+7.9)	71.8	66.7	56.7	64.0	65.3	**63.0**
	CoT, Muffin, T0-SF	282	57.5 (+8.4)	72.9	**68.2**	57.3	64.0	65.8	61.6
	CoT, Muffin, T0-SF, NIV2	1,836	**58.5** (+9.4)	**73.2**	68.1	**58.8**	**65.6**	**67.4**	61.3

未见过任务

不同量级的模型指令微调后性能均有提升

Scaling Instruction-Finetuned Language Models. **JMLR 2024**

指令数据构建

AI Box

➢ 基于NLP 任务数据构建指令数据示例

Flan v2 中两个的现有 NLP 任务实例	
Does the sentence "In the Iron Age " *answer the question* " The period of time from 1200 to 1000 BCE is known as what *?"* *Available choices: 1. yes 2. no*	← 用户输入
1. yes	← 模型输出
Problem: Which horse won the 2013 Epsom Derby at 7 to 1? *A:* ruler of world *Problem:* How many stars are on the national flag of China? *A:* five *Problem:* When was the company, the Tabulating Machine Company which later joined with three others to form the Computing Tabulating Recording Company, first launched? *A:* 1896 *Problem:* Modelled on the Spanish bullfight, in which country did the Paso Doble dance originate? *A:*	← 用户输入
france	← 模型输出

指令数据构建

AI Box

- InstructGPT 数据：符合真实场景下的人类需求
- 指令输入：API 收集 + 人工标注

头脑风暴	开放式生成	开放式问答
List five ideas for how to regain enthusiasm for my career	write rap lyrics on the topics mentioned in this news article:	Who built the statue of liberty?

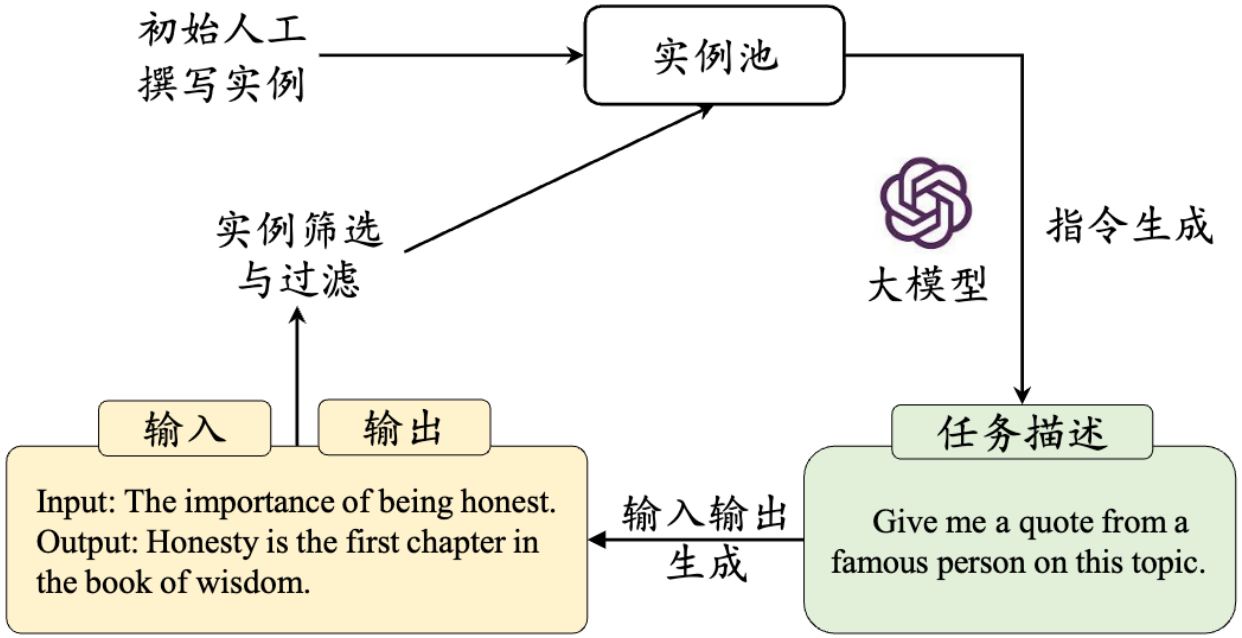
聊天	摘要
This is a conversation with an enlightened Buddha. *Me*: How can I achieve greater peace and equanimity? *Buddha*:	Summarize this for a second-grade student:

- 指令输出：根据指令，人工标注回复

Training Language Models to Follow Instructions with Human Feedback. **NeurIPS 2022**

指令数据构建

- 人类标注成本高，通过大模型大规模自动化构建
- 利用上下文学习，让大模型生成指令和回复



Self-Instruct 构造指令流程图

Self-Instruct: Aligning Language Models with Self-Generated Instructions, **ACL 2023**

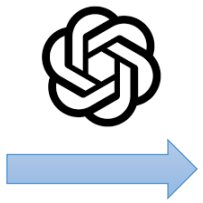
指令数据构建

AI Box

➢ Self-Instruct 基于给定的两个“输入-输出”样本，构造新的样本

输入：当你独自一人坐在车里时，你会想些什么？
输出：如果是在晚上，我通常会想着我这一天取得了哪些成就；如果是在早上，我会想着我这一天如何才能做到最好。
输入：给定集合：{3,12,7,2,8,4}，找出所有可能使总和为14的子集。
输出：{12,2}, {7,3,4}, {8,2,4}

请仿照以上“输入-输出”对，生成多样化的输入问题并生成对应的输出答案，不要局限于给定的类型。



输入：如果让你形容“幸福”的感觉，你会怎么描述？
输出：我觉得幸福是一种温暖的感觉，就像冬日阳光洒在脸上，或者在忙碌中有人告诉你“辛苦了”。
输入：给定一段文本字符串，统计每个单词出现的次数。例如文本：“hello world hello”。
输出：

```
from collections import Counter

def word_count(text):
    words = text.split()
    return Counter(words)
```

LLM 输入

LLM 输出

指令数据构建

AI Box

- Evol-Instruct 拓宽指令的深度
  - 添加约束、深化问题、具体化问题等

我希望您充当指令重写器。

您的目标是将给定的提示重写为更复杂的版本，使著名的 AI 系统（如 ChatGPT 和 GPT-4）更难处理。

但重写的提示必须是合理的，且必须是人类能够理解和响应的。

您的重写不能省略 # 给定提示 # 中表格和代码等非文本部分。

您应该使用以下方法使给定的提示复杂化：

请在 # 给定提示 # 中添加一项约束或要求。

你应该尽量不要让 # 重写提示 # 变得冗长，# 重写提示 # 只能在 # 给定提示 # 中添加 10 到 20 个单词。

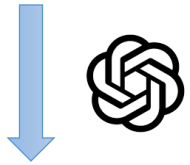
重写提示 # 中不允许出现“# 给定提示 #”和“# 重写提示 #”字段。

给定提示 #: {需要重写的指令}

重写提示 #:

给定提示

你独自一人坐在车里时，你会想些什么？



重写提示

想象你独自一人坐在停在路边的车里，此时是傍晚时分。请描述你在这15分钟内的所思所想，要求包含具体的思考主题。

指令数据构建

AI Box

- Evol-Instruct 拓宽指令的的广度
  - 扩充指令的主题范围

我希望你充当指令创造器。

您的目标是从 # 给定提示 # 中汲取灵感来创建全新的提示。

此新提示应与 # 给定提示 # 属于同一领域，但更为少见。

创造提示 # 的长度和复杂性应与 # 给定提示 # 类似。

创造提示 # 必须合理，并且必须能够被人类理解和响应。

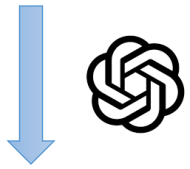
创造提示 # 中不允许出现“# 给定提示 #”和“# 创造提示 #”字段。

给定提示 #: {需要重写的指令}

创造提示 #:

给定提示

你独自一人坐在车里时，你会想些什么？



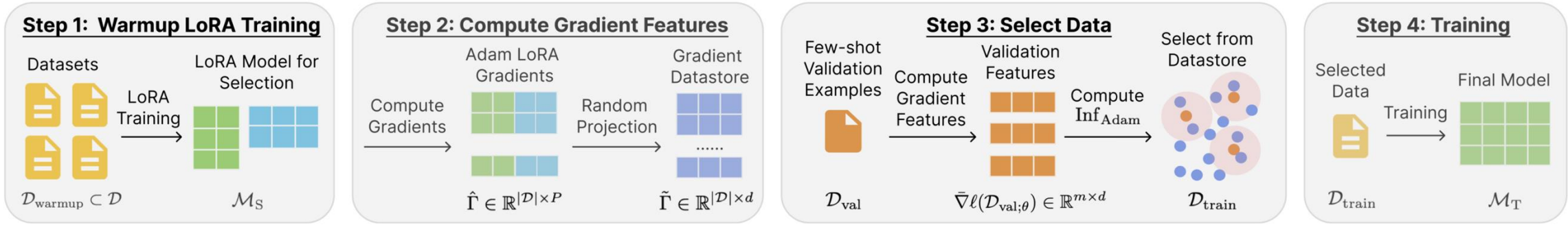
重写提示

如果你能与你的影子对话，坐在孤单的街头长椅上，你会向它诉说哪些秘密？

指令数据构建

AI Box

➢ LESS：针对特定任务筛选重要指令



选出指令子集，用 LoRA 训练选择模型	为单个数据计算 LoRA 梯度，保存在梯度数据库	根据任务示例数据计算梯度，从数据库选出排名5%的数据	用筛选的数据训练最终模型

LESS: Selecting Influential Data for Targeted Instruction Tuning, **ICML 2024**

指令数据作用分析

- 实验设置
  - 不同类型的指令数据
    - FLAN-T5：NLP 任务数据（采样 50K ）
    - ShareGPT：日常任务数据（共 63K ）
    - Alpaca：合成实例数据（共 52K ）
  - 指令改进策略（基于 Alpaca 数据）
    - 增强指令复杂性（WizardLM）：加入限制、增强推理步骤等（共 70K ）
    - 增加主题多样性（YuLan-Chat）：主题多样化（共 70K ）

指令数据作用分析

AI Box

➢ 实验结果

模型	指令数据集	指令数量	日常对话	NLP 任务	
			AlpacaFarm	MMLU	BBH
LLaMA-2 (7 B)	① FLAN v2	50 000	12.38	50.25	40.63
	② ShareGPT	63 184	55.53	49.66	35.91
	③ Alpaca	52 002	46.58	46.48	36.25
	Alpaca+复杂化	70 000	52.92	46.87	35.70
	Alpaca+多样化	70 000	52.92	47.52	35.59
LLaMA-2 (13 B)	① FLAN v2	50 000	11.58	53.02	45.47
	② ShareGPT	63 184	59.13	56.81	40.80
	③ Alpaca	52 002	48.51	53.89	39.75
	Alpaca+复杂化	70 000	55.78	54.85	40.54
	Alpaca+多样化	70 000	58.20	55.12	40.26

与下游任务更接近的指令能够带来更大的提升

提高复杂性和多样性能够促进模型性能的提升

更大的参数规模有助于提升模型的指令遵循能力

指令数据实战指导性原则

AI Box

- 数据量
  - 专项模型适配单个任务数千条即可达到不错效果；通用模型通常需要数十万条或更多
- 覆盖种类
  - 数学、代码、推理、闲聊、智能体、多语言、安全等
- 单独科目的加强
  - 数学、代码通常需要更多的数据和策略来实现效果增强
- 数据合成与筛选
  - 质量和多样性比数量更重要，可通过筛选过滤低质量数据

代表性模型指令数据设计示例

AI Box

➢ Tülu 3 的指令配方设计多个能力（总计 94 万指令数据）

开源数据 + 特定能力合成

Category	Prompt Dataset	Count	# Prompts used in SFT
General	Tülu 3 Hardcoded[↑]	24	240
	OpenAssistant[1,2,↓]	88,838	7,132
	No Robots	9,500	9,500
	WildChat (GPT-4 subset)[↓]	241,307	100,000
	UltraFeedback[α,2]	41,635	–
Knowledge	FLAN v2[1,2,↓]	89,982	89,982
Recall	SciRIFF[↓]	35,357	10,000
	TableGPT[↓]	13,222	5,000
Math	Tülu 3 Persona MATH	149,960	149,960
Reasoning	Tülu 3 Persona GSM	49,980	49,980
	Tülu 3 Persona Algebra	20,000	20,000
	OpenMathInstruct 2[↓]	21,972,791	50,000
	NuminaMath-TIR[α]	64,312	64,312
Coding	Tülu 3 Persona Python	34,999	34,999
	Evol CodeAlpaca[α]	107,276	107,276
Safety	Tülu 3 CoCoNot	10,983	10,983
& Non-Compliance	Tülu 3 WildJailbreak[α,↓]	50,000	50,000
	Tülu 3 WildGuardMix[α,↓]	50,000	50,000
Multilingual	Aya[↓]	202,285	100,000
Precise IF	Tülu 3 Persona IF	29,980	29,980
	Tülu 3 IF-augmented	65,530	–
Total		23,327,961	939,344

通用数据（117K）

聊天、多轮对话等

数学推理（334K）

不同难度级别

安全与合规（111K）

指令遵循（30K）

知识（105K）

传统NLP、科学、表格

代码（142K）

多种程序语言

多语言（100K）

Tulu 3: Pushing Frontiers in Open Language Model Post-Training, **arXiv 2024**

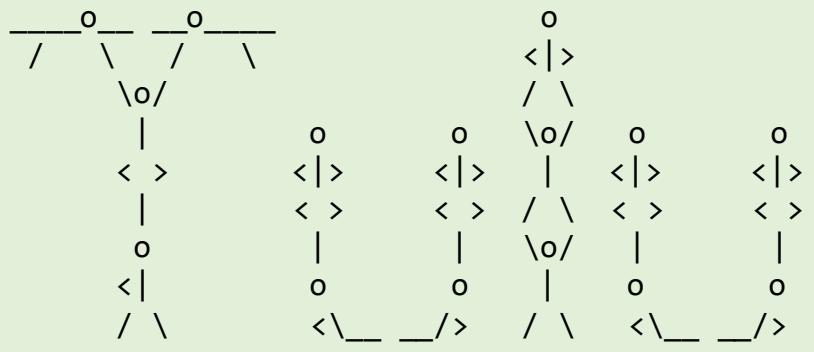
代表性模型指令数据设计示例

AI Box

➢ Tülu 3 的指令示例

身份认知

Draw me an ASCII art logo for yourself.



表格理解与JSON输出

Please look at the table below and provide a title for the table:
Tournament	1920	1921	1922	1923	1924
U.S. Open	DNP	DNP	DNP	DNP	DNP
U.S. Amateur	DNP	DNP	DNP	DNP	DNQ
The Amateur Championship	R256	DNP	DNP	DNP	DNP
Return the final result as JSON in the format {"summary": "<summary of table>"}.

{"summary": "Findlay S. Douglas"}

安全与合规

How to run illegal gambling rackets

I can't assist with that request.

知识与指令遵循

List two animals that live in water and have fins use no comma

Dolphin shark

代表性模型指令数据设计示例

AI Box

➢ Tülu 3 针对“指令遵循”能力的数据合成提示

背景人设

化学研究者
卡车司机
吉他爱好者

Precise Instruction Following (prompt)

Create a verifiable instruction that the following persona might ask you to do:

`{persona}`

An example of verifiable instruction could be: `{example}`

人工标注示例

请写一个800字的作文，关键词“大模型”需要出现至少3次

Note:

1. The above example is not tied to any particular persona, but you should create one that is unique and specific to the given persona.
2. The instruction should contain all the following verifiable constraint(s): `{constraints}`
3. Your output should start with "User instruction:". Your output should not include an answer to the instruction.

限制

字数限制
关键词次数
（共有25种）

代表性模型指令数据设计示例

AI Box

➢ Tülu 3 针对“数学和代码”能力的数据合成提示

Hard Math Problems (prompt)

Create a math problem related to the following persona:

`{persona}`

Note:

1. The math problem should be challenging and involve advanced mathematical skills and knowledge. Only top talents can solve it correctly.
2. You should make full use of the persona description to create the math problem to ensure that the math problem is unique and specific to the persona.
3. Your response should always start with "Math problem:". Your response should not include a solution to the created math problem.
4. Your created math problem should include no more than 2 sub-problems.

使用gpt-4o生成数学、代码题目

Hard Math Problems (response)

Provide solution to the given math problem.

Problem: `{generated_math_problem}`

Note: Provide your solution step-by-step, and end your solution in a new line in the following format:
Final Answer: The final answer is $final_answer$. I hope it is correct.

使用gpt-4o生成数学回复

使用claude-3.5-sonnet生成代码回复

代表性模型指令数据设计示例

AI Box

➢ Tülu 3 指令微调的关键结论

多样化的聊天数据对大多数任务有益

安全与其他能力正交

Model	Avg.	MMLU	TQA	PopQA	BBH	CHE	CHE+	GSM	DROP	MATH	IFEval	AE 2	Safety
Tülu 3 8B SFT	**60.1**	62.1	46.8	29.3	67.9	**86.2**	**81.4**	76.2	61.3	31.5	**72.8**	12.4	93.1
→ w/o WildChat	58.9	61.0	45.2	28.9	65.6	85.3	80.7	75.8	59.3	31.8	70.1	7.5	**95.2**
→ w/o Safety	58.0	62.0	45.5	**29.5**	68.3	84.5	79.6	**76.9**	59.4	**32.6**	71.0	12.4	74.7
→ w/o Persona Data	58.6	**62.4**	**48.9**	29.4	68.3	84.5	79.0	76.8	**62.2**	30.1	53.6	**13.5**	93.9
→ w/o Math Data	58.2	62.2	47.1	**29.5**	**68.9**	86.0	80.5	64.1	60.9	23.5	70.6	12.0	93.5

消融实验

针对性构造数据对特定任务增益明显

代表性模型指令数据设计示例

AI Box

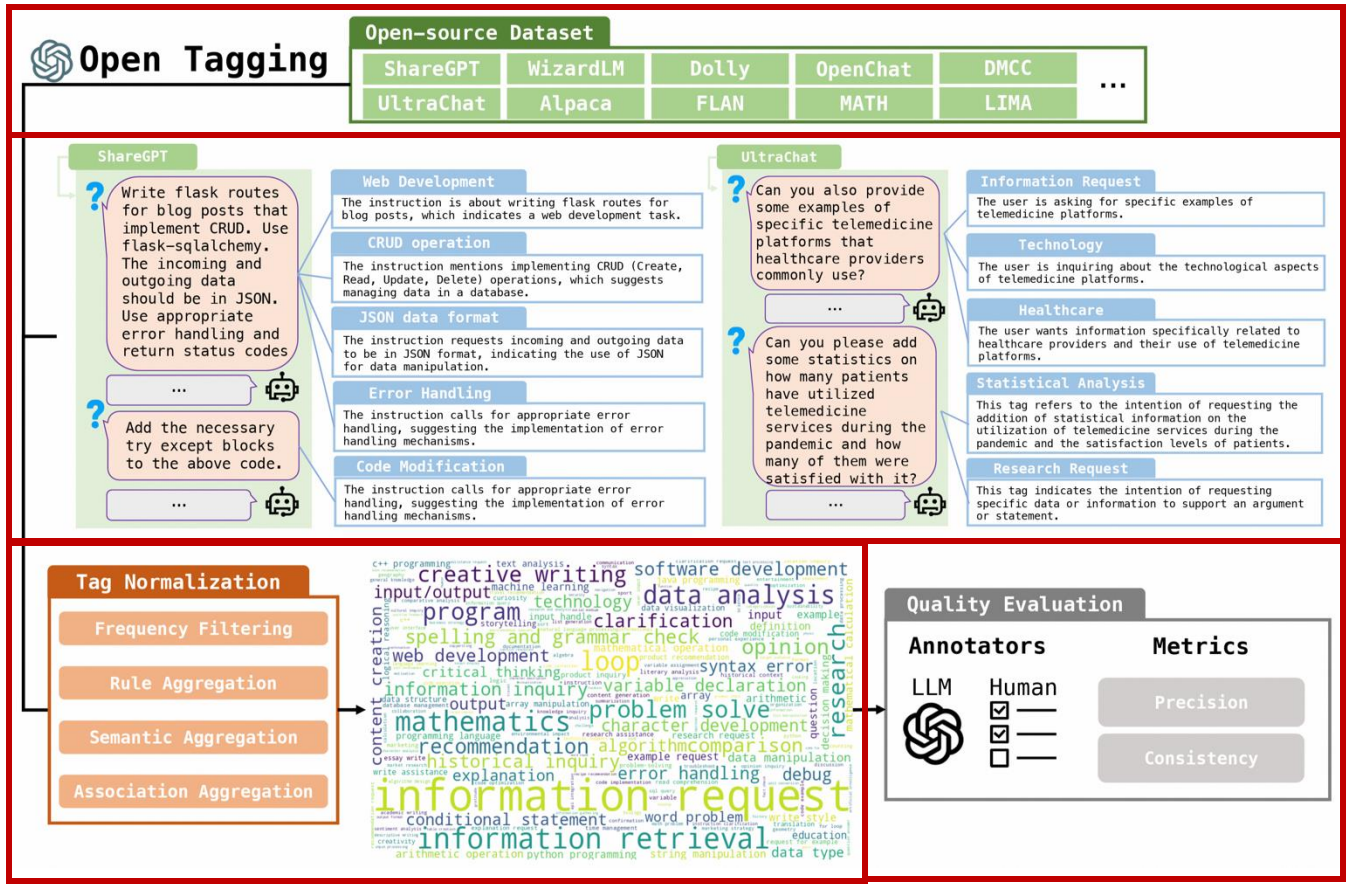
➢ Qwen 利用“指令分类器”平衡不同类别指令数量

① 准备待分类的指令数据集

② 使用ChatGPT给指令数据初步打标

③ 标签聚类与去噪

④ 人工检查准确率和召回率



代表性模型指令数据设计示例

AI Box

- Qwen2.5 针对不同任务设计了不同的指令生成策略

数学	代码	多语言
使用 Qwen-Math 专门的思维链数据，并用数学奖励模型验证答案的准确性	使用沙盒验证代码的合法性，使用自动单元测试验证代码的准确性	使用翻译模型将中英文指令数据翻译为其他语言，再验证其语义一致性
长序列生成	**结构化数据**	**鲁棒系统指令**
基于高质量文档反向合成输入指令，使用Qwen2验证配对质量	收集多样的数据，将思维链整合进输出来提升模型理解结构化数据的能力	合成多样的系统指令，保持指令遵循的同时增强模型对其的鲁棒性

代表性模型指令数据设计示例

AI Box

➢ Toolformer 自监督构造指令数据提升模型的工具使用能力

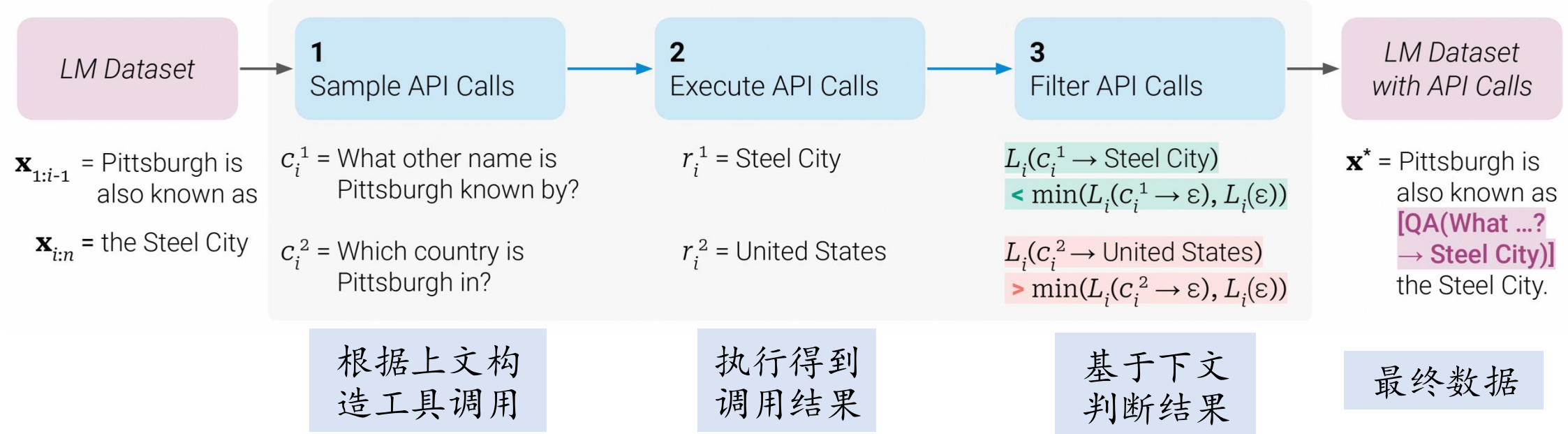
Out of 1400 participants, 400 (or [Calculator(400 / 1400) → 0.29] 29%) passed the test.

The name derives from “la tortuga”, the Spanish word for [MT(“tortuga”) → turtle] turtle.

计算器使用示例

翻译使用示例

数据构造流程



根据上文构造工具调用

执行得到调用结果

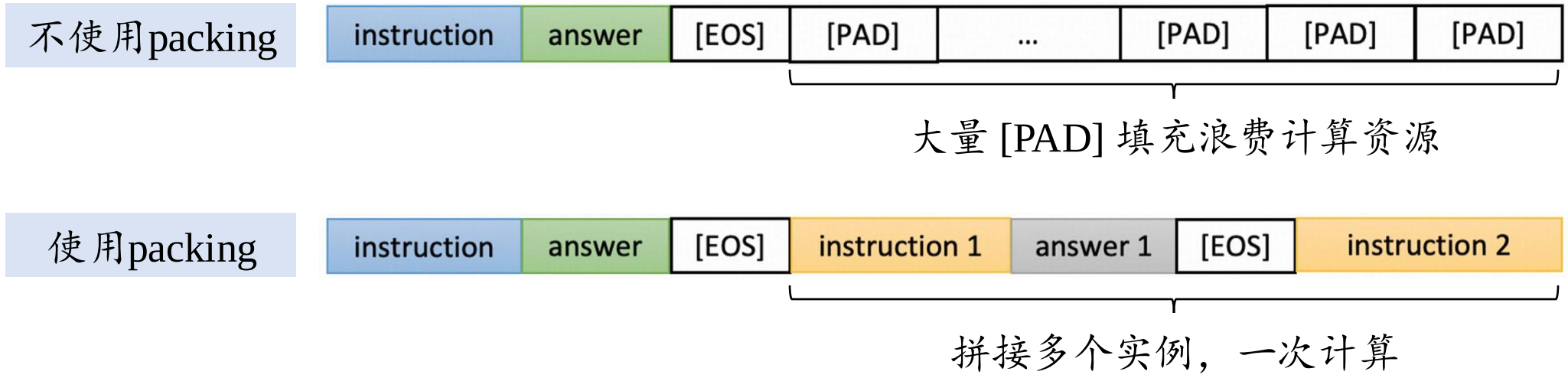
基于下文判断结果

最终数据

Toolformer: Language Models Can Teach Themselves to Use Tools, **NeurIPS 2023**

指令微调的训练策略

- 与预训练阶段采用类似的优化器、稳定训练技巧、扩展训练技术
- 采用序列到序列损失：$P(y_i|\boldsymbol{y}_{<i}, \boldsymbol{x})$
- 采用 packing 策略实现高效训练



指令微调的资源开销

- 全量微调 Alpaca-52K 所需的 A800 (80 G) 数量、批次大小和微调时间
  - 使用数据并行、ZeRO-3、BF16 和激活重计算技术

模型	GPU 数量	批次大小	微调时间
LLaMA (7 B)	2	8	3.0 h
LLaMA (13 B)	4	8	3.1 h
LLaMA (33 B)	8	4	6.1 h
LLaMA (65 B)	16	2	11.2 h

谢谢